

DATA NOTE

Open Access



Draft genome sequence and annotation of the polyextremotolerant polyol lipid-producing fungus *Aureobasidium pullulans* NRRL 62042

Marie R. E. Dielentheis-Frenken¹, Daniel Wibberg^{2,3}, Lars M. Blank¹ and Till Tiso^{1*}

Abstract

Objectives The ascomycotic yeast-like fungus *Aureobasidium* exhibits the natural ability to synthesize several secondary metabolites, like polymalic acid, pullulan, or polyol lipids, with potential biotechnological applications. Combined with its polyextremotolerance, these properties make *Aureobasidium* a promising production host candidate. Hence, plenty of genomes of Aureobasidia have been sequenced recently. Here, we provide the annotated draft genome sequence of the polyol lipid-producing strain *A. pullulans* NRRL 62042.

Data description The genome of *A. pullulans* NRRL 62042 was sequenced using Illumina NovaSeq 6000. Genome assembly revealed a genome size of 24.2 Mb divided into 39 scaffolds with a GC content of 50.1%. Genome annotation using Genemark v4.68 and GenDBE yielded 9,596 genes.

Keywords *Aureobasidium pullulans*, Polyextremotolerant, Polyol lipid, Pullulan, Polymalic acid

Objective

Strains belonging to the genus *Aureobasidium* are polyextremotolerant ascomycetes [1, 2]. This genus encompasses a diverse array of black yeasts prevalent in terrestrial and aquatic habitats across the globe. Boasting remarkable ecological versatility, *Aureobasidium* species have been isolated from soil, water, air, and even extreme environments like cold deserts and limestone caves [3–6]. While other species exist, *Aureobasidium pullulans*

stands out as the most ubiquitous and well-studied member [7]. Among many beneficial traits, strains belonging to the genus *Aureobasidium* exhibit remarkable synthesis capabilities to produce a spectrum of biotechnologically interesting products, like polymalic acid, pullulan, extracellular enzymes, and polyol lipids (a.k.a. liamocins) [3, 8–10]. Additionally, *Aureobasidium* features polyextremotolerance, which is manifest by resistance towards high salt concentrations, acidic and basic conditions, and a temperature spectrum from polar to tropical conditions [11, 12]. These properties make it a potential industrial production organism. As a result, a large number of genomes have been sequenced in the past years [1, 7, 13–16].

Aureobasidium species are known for the production of a plethora of secondary metabolites, and two species were even named after their mainly produced secondary

*Correspondence:

Till Tiso

till.tiso@rwth-aachen.de

¹Institute of Applied Microbiology, RWTH Aachen University, Aachen, Germany

²Center for Biotechnology, Universität Bielefeld, Bielefeld, Germany

³Computational Metagenomics, Institute of Bio- and Geosciences IBG-5, Forschungszentrum Jülich GmbH, Jülich, Germany



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data set 1	AP_62042.custom.alignment	Binary Alignment Map (.bam)	NCBI Sequence Read Archive [http://identifiers.org/ncbiprotein:SRR27493674] [42]
Data set 2	AP62042sub	Fasta (.fa)	NCBI GenBank Database [http://identifiers.org/ncbi-protein:JASGXD010000000] [43]

metabolites. While representatives of the *A. melanogenum* species often synthesize the pigment melanin [17], *A. pullulans* strains are used in industrial-scale processes to produce pullulan [18, 19]. *A. pullulans* NRRL 62042 was isolated from a leaf in Patalung, Thailand, in 2010 [20]. This strain was reported as a producer of various secondary metabolites, but is mainly known for polyol lipid production [20–22]. The polyol lipids produced by *Aureobasidium* are amphiphilic molecules and, as such, might be applicable as biosurfactants [23, 24]. Since not many strains feature the ability to produce these biosurfactants, it is of particular interest to unravel the genetic foundation and metabolic pathways underlying polyol lipid synthesis.

Data description

Prior to DNA isolation, the strain *A. pullulans* NRRL 62042 (ARS culture collection, Peoria, Illinois, USA) was grown in YPD medium for 20 h at 30 °C and 200 rpm. Genomic DNA was isolated using the Monarch Genomic DNA Purification Kit (New England Biolabs, Frankfurt am Main, Germany). Cells were lysed mechanically according to the protocol provided by NEB using a bead-beater at 6 m s⁻¹ for 40 s (FastPrep-24TM, MP Biomedicals, Santa Ana, Kalifornien, USA). A standard genomic library was created using unique dual indexing, and paired-end 150 bp whole genome sequencing was carried out on an Illumina NovaSeq 6000 (Eurofins Genomics, Ebersberg, Germany). This resulted in 2,611,482,000 sequenced bases and a genome coverage of 104x. A total of 17,409,870 raw reads were cleaned and filtered to 17,254,440 high-quality reads using fastp software for quality control processing [25]. For error correction and normalization, bbnorm was used [26]. The cleaned and normalized reads were assembled using SPAdes (version 3.15.0) [27], resulting in 53 contigs (largest contig: 1,570,265 bp; N50=1,070,283 bp; average coverage depth: 97) and 39 scaffolds (largest scaffold: 1,934,914 bp; N50=1,169,448 bp; average coverage depth: 97). The genome size was determined to be 24.2 Mb with a GC content of 50.1%. The quality of the assembly was evaluated using QUAST [28] by mapping the reads back to the assembly, resulting in a genome mapping of 99.98%. Additionally, a BUSCO analysis of the assembly was performed [29, 30], yielding a score of 97.1%. Genome annotation for *A. pullulans* 62042 was performed as described

recently [31–33] using Genemark v4.68 [34] and GenDBE [35]. For automatic annotation within the platform, similarity searches against different databases, including COG [36], KEGG [37], and SWISS-PROT [38] were performed. In addition to genes, putative tRNA genes were identified with tRNAscan-SE [39]. In total, 9,596 genes were annotated. A BUSCO analysis of the predicted genes [29, 30] resulted in a score of 97.8%. Data sets 1 and 2 (Table 1) were deposited in the NCBI Bioproject PRJNA972899 [40], Biosample SAMN35102111 [41].

Limitations

The data presented in this genome note is limited to a single draft genome sequence. Additionally, the genome sequence was generated by Illumina sequencing and is thus relatively fragmented. To overcome these limitations, the data provided here should be set in the context of other sequenced *Aureobasidium* genomes. Furthermore, long-read sequencing methods like Nanopore or Pacbio sequencing could be used to generate a completely assembled genome.

Abbreviations

bp Base pairs

Acknowledgements

The authors thank Team Aureo (Frederick Haala, Philipp Kohl, Marielle Driller, Karla Stein, Difan Xiao) for their support. Bioinformatics support by the BMBF-funded project “Bielefeld-Gießen Center for Microbial Bioinformatics—BiGi (Grant Number 031A533)” within the German Network for Bioinformatics Infrastructure (de.NBI) is gratefully acknowledged.

Author contributions

Conceptualization: MDF, TT. Funding acquisition: TT. Data acquisition: MDF, DW. Data analysis: MDF, DW. Draft – writing: MDF, TT, DW. Draft – review and editing: MDF, DW, LMB, TT. All authors read and approved the manuscript.

Funding

This work was funded by the Fachagentur Nachwachsende Rohstoffe (FNR) of the German Federal Ministry of Food and Agriculture (BMEL) within the project ViRiDi Aurum (grant 2220NR262X) and the Federal Ministry for Education and Research (BMBF) in the frame of de.NBI/ELIXIR-DE (grant number 031A532B). TT acknowledges funding by the Ministry of Culture and Science of the State of North Rhine-Westphalia within the framework of the NRW Strategieprojekt Bioeconomy Science Center (BioSC) (No. 313/323-400-002 13) in the project *Surflin* and by the German Federal Ministry of Education and Research (BMBF) within the project AureoFACTory (grant 031B1203). The laboratory of LMB is partially funded by the Deutsche Forschungsgemeinschaft (DFG) (German Research Foundation) under Germany’s Excellence Strategy within the Cluster of Excellence FSC 2186 The Fuel Science Center. Open Access funding enabled and organized by Projekt DEAL.

Data availability

The data described in this data note can be freely and openly accessed on NCBI Bioproject PRJNA972899 and Biosample SAMN35102111. The Whole Genome Shotgun project has been deposited at GenBank under the accession JASGXD000000000. The version described in this paper is version JASGXD010000000. Please see Table 1 for details and links to the data.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 28 May 2024 / Accepted: 26 July 2024

Published online: 20 August 2024

References

- Gostinčar C, Turk M, Zajc J, Gunde-Cimerman N. Fifty *Aureobasidium pullulans* genomes reveal a recombining polyextremotolerant generalist. *Environ Microbiol*. 2019;21(10):3638–52.
- de Hoog GS, Yurlova NA. Conidiogenesis, nutritional physiology and taxonomy of *Aureobasidium* and *Hormonema*. *Antonie Van Leeuwenhoek*. 1994;65(1):41–54.
- Zalar P, Gostinčar C, de Hoog GS, Uršič V, Sudhaham M, Gunde-Cimerman N. Redefinition of *Aureobasidium pullulans* and its varieties. *Stud Mycol*. 2008;61:21–38.
- Urzi C, De Leo F, Lo Passo C, Criseo G. Intra-specific diversity of *Aureobasidium pullulans* strains isolated from rocks and other habitats assessed by physiological methods and by random amplified polymorphic DNA (RAPD). *J Microbiol Methods*. 1999;36(1–2):95–105.
- Gunde-Cimerman N, Zalar P, De Hoog S, Plemenitis A. Hypersaline waters in salterns - natural ecological niches for halophilic black yeasts. *FEMS Microbiol Ecol*. 2000;32:235–40.
- Mazina SE, Gasanova TV, Kozlova EV, Popkova AV, Fedorov AS, Bukharina IL et al. Biodiversity of phototrophs and culturable fungi in gobiustan caves. *Life*. 2023;13(1).
- Gostinčar C, Ohm RA, Kogej T, Sonjak S, Turk M, Zajc J et al. Genome sequencing of four *Aureobasidium pullulans* varieties: biotechnological potential, stress tolerance, and description of new species. *BMC Genomics*. 2014;15(549).
- Chi Z, Wang F, Chi Z, Yue L, Liu G, Zhang T. Bioproducts from *Aureobasidium pullulans*, a biotechnologically important yeast. *Appl Microbiol Biotechnol*. 2009;82(5):793–804.
- Prasongsuk S, Lotrakul P, Ali I, Bankeeree W, Punnapayak H. The current status of *Aureobasidium pullulans* in biotechnology. *Folia Microbiol (Praha)*. 2018;63(2):129–40.
- Tiso T, Welsing G, Blank LM. Proposal for a systematic naming convention for liamocins. *J Surfactants Deterg*. 2024;(January):2–4.
- Gostinčar C, Stajich JE, Zupan J, Zalar P. Genomic evidence for intraspecific hybridization in a clonal and extremely halotolerant yeast. *BMC Genomics*. 2018;19(364):1–12.
- Gostinčar C, Muggia L, Grube M. Polyextremotolerant black fungi: Oligotrophism, adaptive potential, and a link to lichen symbioses. *Front Microbiol*. 2012;3(NOV):1–6.
- Xiao D, Blank LM, Tiso T. Draft whole-genome sequence of the black yeast *Aureobasidium pullulans* NRRL 62031. *Microbiol Resour Announc*. 2023;12(5).
- Rueda-Mejía MP, Nägeli L, Lutz S, Hayes RD, Varadarajan AR, Grigoriev IV, et al. Genome, transcriptome and secretome analyses of the antagonistic, yeast-like fungus *aureobasidium pullulans* to identify potential biocontrol genes. *Microb Cell*. 2021;8(8):184–202.
- Vignolle GA, Mach RL, Mach-Aigner AR, Derntl C. Genome sequence of the black yeast-like strain *aureobasidium pullulans* var. *Aubasidani* CBS 100524. *Microbiol Resour Announc*. 2021;10(12):21–2.
- Zhao SF, Jiang H, Chi Z, Liu GL, Chi ZM, Chen TJ et al. Genome sequencing of *Aureobasidium pullulans* P25 and overexpression of a glucose oxidase gene for hyper-production of Ca²⁺-gluconic acid. *Antonie van Leeuwenhoek, Int J Gen Mol Microbiol* [Internet]. 2019;112(5):669–78. <https://doi.org/10.1007/s10482-018-1197-3>
- Jiang H, Chi Z, Liu GL, Hu Z, Zhao SZ, Chi ZM. Melanin biosynthesis in the desert-derived *Aureobasidium melanogenum* XJ5-1 is controlled mainly by the CWI signal pathway via a transcriptional activator Cmr1. *Curr Genet* [Internet]. 2020;66(1):173–85. <https://doi.org/10.1007/s00294-019-01010-9>
- Wei X, Liu GL, Jia SL, Chi Z, Hu Z, Chi ZM. Pullulan biosynthesis and its regulation in *Aureobasidium* spp. *Carbohydr Polym* [Internet]. 2021;251(August 2020):117076. <https://doi.org/10.1016/j.carbpol.2020.117076>
- Leathers TD. Biotechnological production and applications of pullulan. *Appl Microbiol Biotechnol*. 2003;62(5–6):468–73.
- Manitchotpitit P, Skory CD, Peterson SW, Price NPJ, Vermillion KE, Leathers TD. Poly(β-L-malic acid) production by diverse phylogenetic clades of *Aureobasidium pullulans*. *J Ind Microbiol Biotechnol*. 2012;39(1):125–32.
- Bischoff KM, Leathers TD, Price NP, Manitchotpitit P. Liamicin oil from *Aureobasidium pullulans* has antibacterial activity with specificity for species of *Streptococcus*. *J Antibiot (Tokyo)* [Internet]. 2015;68(10):642–5. <https://doi.org/10.1038/ja.2015.39>
- Rich JO, Manitchotpitit P, Peterson SW, Liu S, Leathers TD, Anderson AM. Phylogenetic classification of *Aureobasidium pullulans* strains for production of feruloyl esterase. *Biotechnol Lett*. 2016;38(5):863–70.
- Kim JS, Lee IK, Yun BS. A novel biosurfactant produced by *Aureobasidium pullulans* L3-GPY from a tiger lily wild flower, *Lilium lancifolium* thunb. *PLoS ONE*. 2015;10(4):1–12.
- Manitchotpitit P, Price NPJ, Leathers TD, Punnapayak H. Heavy oils produced by *Aureobasidium pullulans*. *Biotechnol Lett*. 2011;33(6):1151–7.
- Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:884–90.
- Bushnell B, Rood J, Singer E. BBMerge – accurate paired shotgun read merging via overlap. *PLoS ONE*. 2017;12(10):1–15.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455–77.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. Genome analysis QAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072–5.
- Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with broader and deeper phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38(10):4647–54.
- Manni M, Berkeley MR, Seppely M, Zdobnov EM. BUSCO: assessing genomic data Quality and Beyond. *Curr Protoc*. 2021;1(12):1–41.
- Wibberg D, Stadler M, Lambert C, Bunk B, Spröer C, Rückert C et al. High quality genome sequences of thirteen Hypoxylaceae (Ascomycota) strengthen the phylogenetic family backbone and enable the discovery of new taxa. *Fungal Divers* [Internet]. 2021;106(1):7–28. <https://doi.org/10.1007/s13225-020-00447-5>
- Schaffhauser T, Wibberg D, Binder A, Rückert C, Busche T, Wohlleben W et al. Genome assembly and genetic traits of the pleuromutilin-producer *clitopilus passeckerianus* DSM1602. *J Fungi*. 2022;8(8).
- Ullmann L, Wibberg D, Busche T, Rückert C, Müsgens A, Kalinowski J et al. Seventeen *Ustilaginaceae* high-quality genome sequences allow phylogenomic analysis and provide insights into secondary metabolite synthesis. *J Fungi*. 2022;8(3).
- Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res*. 2008;18(12):1979–90.
- Rupp O, Becker J, Brinkrolf K, Timmermann C, Borth N, Pühler A et al. Construction of a public CHO cell line transcript database using versatile bioinformatics analysis pipelines. *PLoS ONE*. 2014;9(1).
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*. 2003;4:1–14.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res*. 2004;32(DATABASE ISS.).
- Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*. 2003;31(1):365–70.
- Lowe TM, Eddy SR. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1996;25(5):955–64.

40. Dielentheis-Frenken MRE, Blank LM, Tiso T. Bioproject. [Internet]. NCBI. 2023. <http://identifiers.org/ncbiprotein:PRJNA972899>
41. Dielentheis-Frenken MRE, Blank LM, Tiso T. Biosample. [Internet]. NCBI. 2023. <http://identifiers.org/ncbiprotein:SAMN35102111>
42. Dielentheis-Frenken MRE, Blank LM, Tiso T. Data set 1. SRA data. [Internet]. NCBI SRA. 2024. <http://identifiers.org/ncbiprotein:SRR27493674>
43. Dielentheis-Frenken MRE, Wibberg D, Blank LM, Tiso T. Data set 2. Genome. [Internet]. NCBI. 2023. <http://identifiers.org/ncbiprotein:JASGXD010000000>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.