Genome Sequence of Talaromyces atroroseus, Which Produces Red Colorants for the Food Industry

Thrane, Ulf; Rasmussen, Kasper Bøwig; Petersen, Bent; Rasmussen, Simon; Sicheritz-Pontén, Thomas; Mortensen, Uffe Hasbro

Published in:
Genome Announcements

Link to article, DOI:
10.1128/genomeA.01736-16

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

EUKARYOTES

Genome Sequence of *Talaromyces atroroseus*, Which Produces Red Colorants for the Food Industry

Ulf Thrane,[a] Kasper Bøwig Rasmussen,[a]* Bent Petersen,[b] Simon Rasmussen,[b] Thomas Sicheritz-Pontén,[b] Uffe Hasbro Mortensen[a]

Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark[a]; Department of Bio and Health Informatics, Technical University of Denmark, Kongens Lyngby, Denmark[b]

ABSTRACT *Talaromyces atroroseus* is a known producer of *Monascus* colorants suitable for the food industry. Furthermore, genetic tools have been established that facilitate elucidation and engineering of its biosynthetic pathways. Here, we report the draft genome of a potential fungal cell factory, *T. atroroseus* IBT 11181 (CBS 123796).

Received 22 December 2016 **Accepted** 26 December 2016 **Published** 2 March 2017

Citation Thrane U, Rasmussen KB, Petersen B, Rasmussen S, Sicheritz-Pontén T, Mortensen UH. 2017. Genome sequence of *Talaromyces atroroseus*, which produces red colorants for the food industry. Genome Announc 5:e01736-16. https://doi.org/10.1128/genomeA.01736-16.

Address correspondence to Ulf Thrane, uthrane@gmail.com.

* Present address: Kasper Bøwig Rasmussen, Novo Nordisk A/S, Gentofte, Denmark.

The genus *Talaromyces* primarily contains saprophytic fungi and encompasses medically and industrially relevant species such as the opportunistic human pathogen *T. marneffei* (formerly *Penicillium marneffei*), species with high production of cellulolytic enzymes, i.e., *T. cellulolyticus* (1), as well as the interesting pigment-producing species *T. atroroseus* (2). Several strains of *T. atroroseus* and closely related species are recognized as potential cell factories for *Monascus* pigment production, as they may serve as mycotoxin-free alternatives to *Monascus* spp. (2–4).

T. atroroseus IBT 11181 was originally isolated from red sweet bell pepper bought in a Danish supermarket and is deposited in the CBS collection at CBS-KNAW, Utrecht, the Netherlands, as CBS 123796 and CBS 238.95. We intend to implement this isolate as a model for *T. atroroseus* by investigating its growth physiology (5), by establishing genetic tools (6), and by reporting here the full-genome sequence of *T. atroroseus* IBT 11181.

Genomic DNA was extracted from the mycelium with a slightly modified protocol of the cetyltrimethylammonium bromide method used by Fulton et al. (7). The *T. atroroseus* IBT 11181 genome was sequenced using an Illumina HiSeq 2000 platform on a 180-bp paired-end library and a 6-kb mate-paired library both with reads of 2 × 100 bp by Beijing Genome Institute (BGI), Hong Kong. Sequencing depth was 193×, and assembly of the genome was performed with the ALLPATHS-LG algorithm (8). The final assembly resulted in 48 scaffolds with a G+C content of 44.35% and a total assembly size of 30.85 Mb corresponding to 93% of the estimated genome size from *k*-mer spectral analysis. The minimum number of sequences making up 50% of the genome assembly was seven, and the N_{50} length was 1,577,401 bp. The CEGMA pipeline (9) identified 242 of the 248 core eukaryotic genes, assessing the genome assembly completeness to be 97.58%. This indicated that the draft genome assembly was good with a high completeness and was valid to use for whole-genome analysis.

Gene-calling of the genome was performed using a pipeline of first masking the genome with RepeatMasker version 4.0.5 (Institute for Systems Biology, Seattle, WA, USA; http://www.repeatmasker.org), and then gene-calling with AUGUSTUS version 3.0.3 (10, 11), FGENESH version 3.1.2 (SoftBerry) (12), and GeneMark-ES (13). The individual *ab initio* gene predictions were merged into a consensus gene prediction using EVidenceModeler (14), resulting in a total of 9,519 protein-encoding genes serving as the final gene prediction. The genome sequence reported here represents a

useful resource for further research into the metabolism of *T. atroroseus* and its potential as a cell factory for colorant production.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number LFMY00000000. The version described in this paper is the first version, LFMY01000000.

ACKNOWLEDGMENT

This work was supported by grant 09-064967 from the Danish Council for Independent Research, Technology, and Production Sciences.

REFERENCES

1. Inoue H, Decker SR, Taylor LE 2nd, Yano S, Sawayama S. 2014. Identification and characterization of core cellulolytic enzymes from *Talaromyces cellulolyticus* (formerly *Acremonium cellulolyticus*) critical for hydrolysis of lignocellulosic biomass. Biotechnol Biofuels 7:151. https://doi.org/10.1186/s13068-014-0151-5.
2. Frisvad JC, Yilmaz N, Thrane U, Rasmussen KB, Houbraken J, Samson RA. 2013. *Talaromyces atroroseus*, a new species efficiently producing industrially relevant red pigments. PLoS One 8:e84102. https://doi.org/10.1371/journal.pone.0084102.
3. Mapari SAS, Nielsen KF, Larsen TO, Frisvad JC, Meyer AS, Thrane U. 2005. Exploring fungal biodiversity for the production of water-soluble pigments as potential natural food colorants. Curr Opin Biotechnol 16:231–238. https://doi.org/10.1016/j.copbio.2005.03.004.
4. Mapari SAS, Meyer AS, Thrane U, Frisvad JC. 2009. Identification of potentially safe promising fungal cell factories for the production of polyketide natural food colorants using chemotaxonomic rationale. Microb Cell Fact 8:24. https://doi.org/10.1186/1475-2859-8-24.
5. Mapari SAS, Thrane U, Meyer AS. 2010. Fungal polyketide azaphilone pigments as future natural food colorants? Trends Biotechnol 28:300–307. https://doi.org/10.1016/j.tibtech.2010.03.004.
6. Nielsen ML, Isbrandt T, Rasmussen KB, Thrane U, Hoof JB, Larsen TO, Mortensen UH. 2017. Genes linked to production of secondary metabolites in *Talaromyces atroroseus* revealed using CRISPR-Cas9. PLoS One 12:e0169712. https://doi.org/10.1371/journal.pone.0169712.
7. Fulton TM, Chunwongse J, Tanksley SD. 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Rep 13:207–209. https://doi.org/10.1007/BF02670897.
8. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A 108:1513–1518. https://doi.org/10.1073/pnas.1017351108.
9. Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061–1067. https://doi.org/10.1093/bioinformatics/btm071.
10. Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7:62. https://doi.org/10.1186/1471-2105-7-62.
11. Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics 24:637–644. https://doi.org/10.1093/bioinformatics/btn013.
12. Salamov AA, Solovyev VV. 2000. Ab initio gene finding in *drosophila* genomic DNA. Genome Res 10:516–522. https://doi.org/10.1101/gr.10.4.516.
13. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res 18:1979–1990. https://doi.org/10.1101/gr.081612.108.
14. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol 9:R7. https://doi.org/10.1186/gb-2008-9-1-r7.