



HAL
open science

Sequencing, De Novo Assembly, and Annotation of the Complete Genome of a New Thraustochytrid Species, Strain CCAP_4062/3

Khawla Seddiki, François Godart, Riccardo Aiese Cigliano, Walter Sanseverino, mohamed Barakat, Philippe Ortet, Fabrice Rébeillé, Eric Maréchal, Olivier Cagnac, Alberto Amato

► To cite this version:

Khawla Seddiki, François Godart, Riccardo Aiese Cigliano, Walter Sanseverino, mohamed Barakat, et al.. Sequencing, De Novo Assembly, and Annotation of the Complete Genome of a New Thraustochytrid Species, Strain CCAP_4062/3. *Genome Announcements*, 2018, 6 (11), pp.e01335-01317. 10.1128/genomeA.01335-17 . hal-01743654

HAL Id: hal-01743654

<https://hal.science/hal-01743654v1>

Submitted on 28 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Sequencing, *De Novo* Assembly, and Annotation of the Complete Genome of a New Thraustochytrid Species, Strain CCAP_4062/3

Khawla Seddiki,^a François Godart,^b Riccardo Aiese Cigliano,^c Walter Sanseverino,^c Mohamed Barakat,^d Philippe Ortet,^d Fabrice Rébeillé,^a Eric Maréchal,^a Olivier Cagnac,^b  Alberto Amato^a

^aLaboratoire de Physiologie Cellulaire et Végétale, UMR 5168 CNRS, CEA, INRA, Université Grenoble Alpes, Biosciences and Biotechnology Institute of Grenoble, CEA-Grenoble, Grenoble, France

^bFermentalg, Libourne, France

^cSequentia Biotech SL, Barcelona, Spain

^dAix Marseille Université, CEA, CNRS, UMR7265, LEMIRE, Laboratoire d'Ecologie Microbienne de la Rhizosphère et Environnement Extrêmes, ECCOREV FR3098, St. Paul Les Durance, France

ABSTRACT Thraustochytrids are ecologically and biotechnologically relevant marine species. We report here the *de novo* assembly and annotation of the whole-genome sequence of a new thraustochytrid strain, CCAP_4062/3. The genome size was estimated at 38.7 Mb with 11,853 predicted coding sequences, and the GC content was scored at 57%.

Thraustochytrids are marine unicellular heterokont eukaryotes. Although their distribution in aquatic ecosystems can be broad, up to a depth of 4,000 m, they colonize mainly mangrove habitats (1). They are characterized by (i) monocentric thallus, (ii) ectoplasmic net systems, (iii) scaly cell walls, and (iv) intricate life cycles (2). The ability of some species to produce long-chain polyunsaturated fatty acids has attracted industrial interest. Four thraustochytrid strains have been sequenced to date (3, 4). Here, we present the genome of a new thraustochytrid species, which is highly similar to *Aurantiochytrium limacinum/mangrovei* and *Schizochytrium aggregatum*. The taxonomic position of this new species will be determined later, after more thorough analyses.

The pollen grain bait isolation method (5) was used to obtain monospecific axenic cultures from coastal seawater samples gathered in Mayotte (Indian Ocean, 12°48'51.8"S, 45°14'21.7"E). Once the culture was proved to be monospecific and axenic, it was deposited at the Culture Collection of Algae and Protozoa (CCAP) under the accession number CCAP_4062/3.

Cells were propagated in M1 medium (6) at 30°C under agitation (140 rpm). Genomic DNA was extracted and purified using a Quick-DNA Miniprep Plus kit (Zymo Research) according to the manufacturer's instructions. Quality and concentration were estimated using the Quant-iT PicoGreen dsDNA assay kit. Shotgun sequencing libraries were prepared and sequenced as paired-end reads (100 bp) on an Illumina HiSeq platform (Illumina, Inc.). The quality of the raw reads was assessed with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>).

About 137,997,490 paired-end reads were *de novo* assembled using SPAdes version 3.10.1 (7) with a multi-*k*-mer approach ($k = 27, 31, 35, 41, 45, 49, \text{ and } 53$). Contigs were scaffolded with Satsuma using the genome of *Schizochytrium* sp. CCTCC M209059 (GenBank accession number JTFK00000000) as the reference. The final assembly consisted of 4,503 contigs ≥ 300 bp assembled in 2,232 scaffolds. The N_{50} value was 236,568 bp, and the overall GC content was 57%. AUGUSTUS version 3.3 was used for

Received 26 October 2017 **Accepted** 8 February 2018 **Published** 15 March 2018

Citation Seddiki K, Godart F, Aiese Cigliano R, Sanseverino W, Barakat M, Ortet P, Rébeillé F, Maréchal E, Cagnac O, Amato A. 2018. Sequencing, *de novo* assembly, and annotation of the complete genome of a new thraustochytrid species, strain CCAP_4062/3. *Genome Announc* 6:e01335-17. <https://doi.org/10.1128/genomeA.01335-17>.

Copyright © 2018 Seddiki et al. 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 Olivier Cagnac, ocagnac@fermentalg.com, or Alberto Amato, amato@szi.it.

gene prediction (8). A total of 11,853 coding sequences were recorded. BUSCO (9) analysis showed that 10% of the eukaryotic conserved gene core is not represented and that 85.1% is represented by complete single-copy genes in our genome.

Functional annotation was carried out using AHRD (<https://github.com/groupschoof/AHRD>) after a BLAST search against the UniProt Eukaryote and Protist TrEMBL protein database was performed. A description was associated with 7,861 sequences (66%), while a gene ontology (GO) annotation was assigned to 7,269 transcripts (61%). Analysis with barnap software (<https://github.com/tseemann/barnap>) identified three copies of the 5.8S gene, eight partial copies of the 28S gene, and seven partial copies of the 18S gene. The 18S and 28S sequences were finally reconstructed by PCR amplification and direct Sanger sequencing.

The taxonomy and systematics of this strain are under study. This genome is considered of high interest, because metabolic pathways potentially suitable for the bioproduction of different high-added-value molecules can be identified. Other *Labyrinthulomycetes* genomes have been sequenced, and comparative analyses of closely related taxa can lead to a better understanding of the evolutionary success of thraustochytrids, particularly their ability to produce considerable amounts of lipids.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under BioProject PRJNA396017. The version described in this paper is the first version. The 18S, 5.8S, and 28S rRNAs have been deposited at GenBank under the accession numbers [MF766427](#), [MF766428](#), and [MF766429](#), respectively.

ACKNOWLEDGMENTS

Olivier Bastien (LPCV, Grenoble) is acknowledged for helping in 18S rRNA gene identification. K.S. and A.A. thank Sylvaine Roy for her substantial support and precious advice.

E.M. was supported by the Oceanomics program (Investissements d’Avenir, Biotechnology & Bioresources) from ANR. This work was supported by a Fermentalg-CEA partnership and the Trans’Alg Bpifrance project.

REFERENCES

- Bochdanky AB, Clouse MA, Herndl GJ. 2017. Eukaryotic microbes, principally fungi and labyrinthulomycetes, dominate biomass on bathypelagic marine snow. *ISME J* 11:362–373. <https://doi.org/10.1038/ismej.2016.113>.
- Alderman DJ, Harrison JL, Bremer GB, Jones EBG. 1974. Taxonomic revisions in the marine biflagellate fungi: the ultrastructural evidence. *Mar Biol* 25: 345–357. <https://doi.org/10.1007/BF00404978>.
- Liu B, Ertesvåg H, Aasen IM, Vadstein O, Brautaset T, Heggset TM. 2016. Draft genome sequence of the docosahexaenoic acid producing thraustochytrid *Aurantiochytrium* sp. T66. *Genom Data* 8:115–116. <https://doi.org/10.1016/j.gdata.2016.04.013>.
- Ji XJ, Mo KQ, Ren LJ, Li GL, Huang JZ, Huang H. 2015. Genome sequence of *Schizochytrium* sp. CCTCC M209059, an effective producer of docosahexaenoic acid-rich lipids. *Genome Announc* 3(4):e00819-15. <https://doi.org/10.1128/genomeA.00819-15>.
- Gupta A, Wilkens S, Adcock JL, Puri M, Barrow CJ. 2013. Pollen baiting facilitates the isolation of marine thraustochytrids with potential in omega-3 and biodiesel production. *J Ind Microbiol Biotechnol* 40: 1231–1240. <https://doi.org/10.1007/s10295-013-1324-0>.
- Yokochi T, Honda D, Higashihara T, Nakahara T. 1998. Optimization of docosahexaenoic acid production by *Schizochytrium limacinum* SR21. *Appl Microbiol Biotechnol* 49:72–76. <https://doi.org/10.1007/s002530051139>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- 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>.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.