Genomics Data 8 (2016) 115–116

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in brief

Draft genome sequence of the docosahexaenoic acid producing thraustochytrid *Aurantiochytrium* sp. T66

Bin Liu [a], Helga Ertesvåg [a], Inga Marie Aasen [b], Olav Vadstein [a], Trygve Brautaset [a,b], Tonje Marita Bjerkan Heggeset [b,*]

[a] Department of Biotechnology, NTNU Norwegian University of Science and Technology, Trondheim, Norway
[b] Department of Biotechnology and Nanomedicine, SINTEF Materials and Chemistry, Trondheim, Norway

ARTICLE INFO

Article history:
Received 21 April 2016
Accepted 23 April 2016
Available online 29 April 2016

Keywords:
Thraustochytrid
Aurantiochytrium
Whole genome sequencing
Docosahexaenoic acid

ABSTRACT

Thraustochytrids are unicellular, marine protists, and there is a growing industrial interest in these organisms, particularly because some species, including strains belonging to the genus *Aurantiochytrium*, accumulate high levels of docosahexaenoic acid (DHA). Here, we report the draft genome sequence of *Aurantiochytrium* sp. T66 (ATCC PRA-276), with a size of 43 Mbp, and 11,683 predicted protein-coding sequences. The data has been deposited at DDBJ/EMBL/Genbank under the accession LNGJ00000000. The genome sequence will contribute new insight into DHA biosynthesis and regulation, providing a basis for metabolic engineering of thraustochytrids.

Specifications

Organism/cell line/tissue	*Aurantiochytrium* sp. T66
Sex	N/A
Sequencer or array type	Illumina HiSeq 2 × 100 bp paired-end and Roche 454 FLX++
Data format	Analyzed
Experimental factors	DNA extracted from a pure strain, no treatment
Experimental features	Draft genome sequencing
Consent	N/A
Sample source location	The coast of Madeira, Portugal

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/bioproject/?term=LNGJ00000000

2. Introduction

Eicosapentaenoic acid (EPA; 20:5n−3), docosapentaenoic acid (DPA; 22:5n−3), and docosahexaenoic acid (DHA; 22:6n−3) are the main long chain polyunsaturated omega-3 fatty acids (ω-3 LC-PUFA). Over the past decades, the importance of ω-3 LC-PUFA in many aspects of human health, including brain and neural development, cardiovascular function and immune system regulation has been uncovered [1–4]. At present, marine fish and fish oils are the main sources of EPA and DHA. However, the need for new sustainable ω-3 LC-PUFA sources has attracted increasing attention in recent years [5,6].

Aurantiochytrium spp. belong to the thraustochytrids, which are unicellular, heterotrophic, marine protists, abundant in marine water, and able to grow on various carbon sources [7]. Recent studies have demonstrated that *Aurantiochytrium* spp. can be cultivated to high cell densities and produce biomass with up to 70% lipids of the dry cell weight, and up to 70% of the lipids may be DHA [8]. In a previous study of *Aurantiochytrium* sp. strain T66 (ATCC PRA-276), we demonstrated that its lipid content and fatty acid profile can be manipulated by changing the growth conditions [9]. The genetic, regulatory, and biochemical basis of DHA biosynthesis in thraustochytrids are largely unknown due to the scarcity of genome sequences. The genome sequences of the two thraustochytrids *Schizochytrium* sp. CCTCC M209059 [10] and Quahog Parasite Unknown (QPX) [11] as well as transcriptomes of QPX [12] and *Aurantiochytrium* sp. SD116 [13] were recently published. Here, we report the draft genome sequence of *Aurantiochytrium* sp. strain T66.

3. DNA extraction, library construction and sequencing

Aurantiochytrium sp. strain T66 was isolated from a mixture of marine sediment and seawater sampled from the coast of Madeira, Portugal [14]. Total genomic DNA was isolated with the Blood & Cell culture DNA kits (Qiagen, Hilden, Germany). DNA quality was assessed by gel electrophoresis, and the purity and quantity were determined by the

* Corresponding author at: SINTEF Materials and Chemistry, Department of Biotechnology and Nanomedicine, Postboks 4760 Sluppen, N-7465 Trondheim, Norway.
E-mail address: tonje.heggeset@sintef.no (T.M.B. Heggeset).

http://dx.doi.org/10.1016/j.gdata.2016.04.013

NanoDrop 1000 UV–Vis spectrophotometer (Thermo Scientific) and Qubit 2.0 fluorometer using the Qubit® dsDNA BR Assay Kit (ThermoFisher Scientific). The genome of *Aurantiochytrium* sp. T66 was sequenced using the Illumina HiSeq 2 × 100 bp paired-end and Roche 454 FLX++ platforms at Eurofins Genomics GmbH, Ebersberg, Germany. Three libraries of Illumina HiSeq were prepared (shotgun, 8 kbp and 20 kbp long jumping distance libraries). Following quality clipping and adapter trimming by Trimmomatic version 0.30 [15], genome sequence assembly was achieved by the combination of Velvet version 1.2.10 [16], Newbler v2.9 (454 Life Sciences), and Convey GraphConstructor (Cnykmer, Cnygc v2.2.5526, http://www.conveycomputer.com/). Gene prediction was done by a combination of homology-based approaches and *de novo* predictions [17–21].

4. Data analysis and results

The cleaned data output of Illumina HiSeq were 4.8 Gbp, 5.0 Gbp, and 3.6 Gbp, which represent estimated genome coverages of 102-fold, 115-fold and 82-fold, respectively. Roche 454 FLX++ sequencing resulted in 179 Mbp of cleaned data corresponding to an estimated genome coverage of 4.1-fold. The draft genome of *Aurantiochytrium* sp. T66 is 43 Mbp, with a G + C content of 62.8%, distributed on 495 large scaffolds (≥1000 bp) with an N_{50} length of 1,342,793 bp, L_{50} count of 3, N_{75} length of 594,063 bp, and a N_{90} length of 115,579 bp. A total of 11,683 putative protein-coding genes, 112 tRNA genes, 20 rRNA genes and 4 snRNA genes were predicted. Repetitive regions were estimated to comprise 7.1% of the genome. The *Aurantiochytrium* sp. T66 draft genome sequence generated in this study represents a new source of knowledge which can be used as a reference to study thraustochytrids and it will help to further understand the genetic mechanisms of DHA biosynthesis and regulation. It will also be valuable in comparative genomic studies of other *Aurantiochytrium* sp. strains as well as for metabolic engineering of thraustochytrids.

5. Nucleotide accession number

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession LNGJ00000000.

Conflict of interest

The authors declare that there is no conflict of interests on the work published in this paper.

Acknowledgments

This work was supported by a grant from the Research Council of Norway.

References

[1] D. Li, X. Hu, Fish and its multiple human health effects in times of threat to sustainability and affordability: are there alternatives? Asia Pac. J. Clin. Nutr. 18 (2009) 553–563.
[2] E.E. Birch, S.E. Carlson, D.R. Hoffman, K.M. Fitzgerald-Gustafson, V.L. Fu, J.R. Drover, Y.S. Castaneda, L. Minns, D.K. Wheaton, D. Mundy, J. Marunycz, D.A. Diersen-Schade, The DIAMOND (DHA Intake And Measurement Of Neural Development) Study: a double-masked, randomized controlled clinical trial of the maturation of infant visual acuity as a function of the dietary level of docosahexaenoic acid. Am. J. Clin. Nutr. 91 (2010) 848–859.
[3] S.C. Cottin, T.A. Sanders, W.L. Hall, The differential effects of EPA and DHA on cardiovascular risk factors. Proc. Nutr. Soc. 70 (2011) 215–231.
[4] L.E. Robinson, A.C. Buchholz, V.C. Mazurak, Inflammation, obesity, and fatty acid metabolism: influence of n−3 polyunsaturated fatty acids on factors contributing to metabolic syndrome. Appl. Physiol. Nutr. Metab. 32 (2007) 1008–1024.
[5] D.A. Martins, L. Custodio, L. Barreira, H. Pereira, R. Ben-Hamadou, J. Varela, K.M. Abu-Salah, Alternative sources of n−3 long-chain polyunsaturated fatty acids in marine microalgae. Mar. Drugs 11 (2013) 2259–2281.
[6] N.J. Salem, M. Eggersdorfer, Is the world supply of omega-3 fatty acids adequate for optimal human nutrition? Curr. Opin. Clin. Nutr. Metab. Care 18 (2015) 147–154.
[7] X.J. Yu, Z.Q. Yu, Y.L. Liu, J. Sun, J.Y. Zheng, Z. Wang, Utilization of high-fructose corn syrup for biomass production containing high levels of docosahexaenoic acid by a newly isolated *Aurantiochytrium* sp. YLH70. Appl. Biochem. Biotechnol. 177 (2015) 1229–1240.
[8] I.M. Aasen, H. Ertesvåg, T.M. Heggeset, B. Liu, T. Brautaset, O. Vadstein, T.E. Ellingsen, Thraustochytrids as production organisms for docosahexaenoic acid (DHA), squalene, and carotenoids. Appl. Microbiol. Biotechnol. 100 (2016) 4309–4321.
[9] A.N. Jakobsen, I.M. Aasen, K.D. Josefsen, A.R. Strøm, Accumulation of docosahexaenoic acid-rich lipid in thraustochytrid *Aurantiochytrium* sp. strain T66: effects of N and P starvation and O_2 limitation. Appl. Microbiol. Biotechnol. 80 (2008) 297–306.
[10] X.J. Ji, K.Q. Mo, L.J. Ren, G.L. Li, J.Z. Huang, H. Huang, Genome sequence of *Schizochytrium* sp. CCTCC M209059, an effective producer of docosahexaenoic acid-rich lipids. Genome Announc. 3 (2015) (e00819–00815).
[11] A.E. Garcia-Vedrenne, M. Groner, A. Page-Karjian, G.F. Siegmund, S. Singhal, J. Sziklay, S. Roberts, Development of genomic resources for a thraustochytrid pathogen and investigation of temperature influences on gene expression. PLoS One 8 (2013), e74196.
[12] E. Rubin, A. Tanguy, M. Perrigault, E. Pales Espinosa, B. Allam, Characterization of the transcriptome and temperature-induced differential gene expression in QPX, the thraustochytrid parasite of hard clams. BMC Genomics 15 (2014) 245.
[13] Z. Ma, Y. Tan, G. Cui, Y. Feng, Q. Cui, X. Song, Transcriptome and gene expression analysis of DHA producer *Aurantiochytrium* under low temperature conditions. Sci. Rep. 5 (2015) 14446.
[14] A.N. Jakobsen, I.M. Aasen, A.R. Strøm, Endogenously synthesized (−)-*proto*-quercitol and glycine betaine are principal compatible solutes of *Schizochytrium* sp. strain S8 (ATCC 20889) and three new isolates of phylogenetically related thraustochytrids. Appl. Environ. Microbiol. 73 (2007) 5848–5856.
[15] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (2014) 2114–2120.
[16] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18 (2008) 821–829.
[17] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool. J. Mol. Biol. 215 (1990) 403–410.
[18] E. Birney, M. Clamp, R. Durbin, GeneWise and Genomewise. Genome Res. 14 (2004) 988–995.
[19] M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. 34 (2006) W435–W439.
[20] I. Korf, Gene finding in novel genomes. BMC Bioinf. 5 (2004) 59.
[21] W.H. Majoros, M. Pertea, S.L. Salzberg, TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. Bioinformatics 20 (2004) 2878–2879.