



Data in Brief

Draft genome sequence of *Mameliella alba* strain UMTAT08 isolated from clonal culture of toxic dinoflagellate *Alexandrium tamiyavanichii*



Muhd Danish-Daniel^{a,b,*}, Gan Han Ming^{c,d}, Mohd Ezhar Mohd Noor^a, YeongYik Sung^b, Gires Usup^e

^a School of Fisheries and Aquaculture Sciences, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia

^b Institute of Marine Biotechnology, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia

^c School of Science, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway, 47500 Petaling Jaya, Selangor, Malaysia

^d Monash University Malaysia Genomics Facility, Jalan Lagoon Selatan, Bandar Sunway, 47500 Petaling Jaya, Selangor, Malaysia

^e School of Environment and Natural Resource Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

ARTICLE INFO

Article history:

Received 22 August 2016

Accepted 24 August 2016

Available online 26 August 2016

ABSTRACT

Mameliella alba strain UMTAT08 was isolated from clonal culture of paralytic shellfish toxin producing dinoflagellate, *Alexandrium tamiyavanichii*. Genome of the strain UMTAT08 was sequenced in order to gain insights into the dinoflagellate-bacteria interactions. The draft genome sequence of strain UMTAT08 contains 5.84Mbp with an estimated G + C content of 65%, 5717 open reading frames, 5 rRNAs and 49 tRNAs. It contains genes related to nutrients uptake, quorum sensing and environmental tolerance related genes. Gene clusters for the biosynthesis of type 1 polyketide synthase, bacteriocin, microcin, terpene and ectoine were also identified. This is suggesting that the bacterium possesses diverse adaptation strategy to survive within the dinoflagellate phycosphere. The draft genome sequence and annotation have been deposited at DDBJ/EMBL/GenBank under the accession number JSUQ00000000.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Specifications

Organism	<i>Mameliella alba</i>
Strain	UMTAT08
Sequencer	Illumina HiSeq 2000
Data format	Assembled
Experimental factors	Bacterial strain isolated from paralytic shellfish toxins producing dinoflagellate
Experimental features	<i>De novo</i> genome sequencing followed by assembly and annotation
Consent	N/A
Sample source location	Clonal culture of dinoflagellate <i>Alexandrium tamiyavanichii</i> from Straits of Malacca, Malaysia

1. Direct link to deposited data

<http://www.ncbi.nlm.nih.gov/nucleotide/JSUQ00000000>

* Corresponding author at: School of Fisheries and Aquaculture Sciences, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia.
E-mail address: mdda@umt.edu.my (M. Danish-Daniel).

2. Experimental design, materials and methods

Dinoflagellates are phytoplanktons that play an important role in the primary productivity of the world's oceans. However, several species of this microalgae are known to produce a wide variety of toxins that can cause seafood poisoning *via* transfer through the food chain. Moreover, many countries are affected by harmful algal blooms (HABs) and related shellfish toxicity and fish mortality events. A clonal culture of toxic dinoflagellate *Alexandrium tamiyavanichii* was established during an investigation of HABs and shellfish poisoning reported after consumption of green mussels originated from a mussels culture farm in Malacca, Malaysia. Toxins analysis confirmed the ability of this dinoflagellate produces PSP toxins [1]. A specific community of bacteria is believed to be associated with marine dinoflagellates. The bacteria can be free living in the phycosphere, can be attached to the surface of the algal cells or can occur as intracellular algal symbionts [2]. They may play an important role in growth, physiology and modulating the biosynthesis of PSP toxins in dinoflagellates [3]. In order to understand the dinoflagellate-bacteria interactions, a study was carried out to investigate the culturable bacterial diversity associated with toxic dinoflagellate *A. tamiyavanichii*. During this study, *Mameliella alba* strain UMTAT08 was isolated.

Mameliella alba belongs to the *Roseobacter* clade in the order *Rhodobacterales* [4]. Members of the *Roseobacter* clade are abundant in the marine environments suggesting that they play important roles in

Name	Count	Percent
Amino acid transport and metabolism	499	11.66%
Carbohydrate transport and metabolism	310	7.24%
Cell cycle control, cell division, chromosome partitioning	38	0.89%
Cell motility	57	1.33%
Cell wall/membrane/envelope biogenesis	201	4.7%
Chromatin structure and dynamics	3	0.07%
Coenzyme transport and metabolism	260	6.08%
Cytoskeleton	1	0.02%
Defense mechanisms	69	1.61%
Energy production and conversion	316	7.38%
Extracellular structures	13	0.3%
Function unknown	258	6.03%
General function prediction only	427	9.98%
Inorganic ion transport and metabolism	280	6.54%
Intracellular trafficking, secretion, and vesicular transport	53	1.24%
Lipid transport and metabolism	226	5.28%
Mobilome: prophages, transposons	44	1.03%
Nucleotide transport and metabolism	109	2.55%
Posttranslational modification, protein turnover, chaperones	187	4.37%
Replication, recombination and repair	112	2.62%
Secondary metabolites biosynthesis, transport and catabolism	169	3.95%
Signal transduction mechanisms	138	3.23%
Transcription	292	6.82%
Translation, ribosomal structure and biogenesis	217	5.07%
Not in COG	2037	34.98%

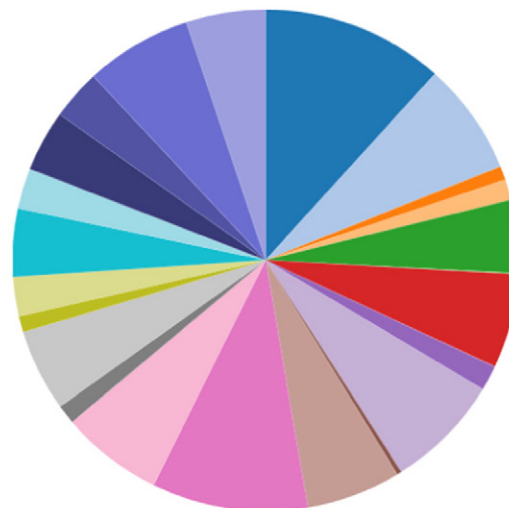


Fig. 1. Distribution and counts of genes in COG categories for genome of strain UMTAT08.

marine ecosystems, such as the degradation of aromatic compounds and the biogeochemical cycles of carbon and sulfur [5]. The genome of strain UMTAT08 was sequenced in order to gain genetic insights into the role of *Mameliella alba* in the phycosphere of dinoflagellate. Strain UMTAT08 was cultured in Marine Broth 2216 (MB; Difco). Genomic DNA was then extracted using the GF-1 nucleic acid extraction kit (Vivantis, Malaysia). Sequencing was performed on the Illumina HiSeq2000 generating 31,195,648 raw FASTQ paired-end reads. Two million reads were sub-sampled for error correction and *de novo* assembly using SPAdes v.3.1.0 [6]. Resulting contigs were used for scaffolding then gap-closing using SSPACE 2.0 and GapFiller v1.11 [7,8]. Sixty-two gap filled contigs with an N50 of 10 sequences longer than 226, 111 bp were produced and the total sequence length was 5,837,382 bp with a 62× coverage. Prokka v1.8 annotation pipeline comprising of Prodigal (v2.60), RNAmmer (v1.2) and Aragorn (v1.2.36) was used to annotate the genome predicting 5717 open reading frames, 5 rRNAs and 49 tRNAs [9–12]. The predicted 16S rRNA was analyzed to identify its phylogenetic affiliation using the EzTaxon server (<http://www.ezbiocloud.net/eztaxon>) [13]. Results showed that strain UMTAT08 was most closely related to *Ponticoccus litoralis* (96.73% similarity) and *Mameliella alba* (95.62% similarity). However, average nucleotide identity (ANI) was then calculated and strain UMTAT08 was confirmed as *Mameliella alba* [14]. InterProScan5 was used to provide additional annotation to the predicted protein sequences [15]. In addition, distribution of genes in COG categories for strain UMTAT08 (Fig. 1) was obtained through the IMG system [16]. Furthermore, antiSMASH was used to identify the presence of secondary metabolite biosynthesis gene clusters in the genome [17].

Analysis of the genome revealed genes associated with nutrients uptake and environmental adaptation, including osmotic and oxidative stress as well as quorum sensing. The genome is predicted to have genes involved in the degradation of aromatic compounds and metabolism of nitrogenous and sulfur compounds, potentially enabling strain UMTAT08 to utilize the dinoflagellate exudates as an immediate source of energy. Interestingly, several gene clusters for the biosynthesis of type 1 polyketide synthase, bacteriocin, microcin, terpene and ectoine have been identified (contig 1, 2, 7, 12, 14 & 23). This suggests that

strain UMTAT08 possess better ability to compete with other bacteria to colonize the host dinoflagellate as survival mechanism.

Nucleotide sequence accession number

The draft genome sequence of *Mameliella alba* strain UMTAT08 has been deposited at DDBJ/EMBL/GenBank under the BioProject number PRJNA265019, BioSample number SAMN03145050 and Accession number JSUQ00000000.

Acknowledgments

This study was funded by Malaysian government through grant FRGS Vot No. 59230. Bioinformatics infrastructure for genome assembly and annotation was kindly provided by the Monash University Malaysia Tropical Medicine Biology Multidisciplinary Platform.

References

- [1] G. Usup, L.C. Pin, A. Ahmad, L.P. Teen, *Alexandrium* (Dinophyceae) species in Malaysian waters. *Harmful Algae* 1 (3) (2002) 265–275.
- [2] S. Jasti, M.E. Sieracki, N.J. Poulton, M.W. Giewat, J.N. Rooney-Varga, Phylogenetic diversity and specificity of bacteria closely associated with *Alexandrium* spp. and other phytoplankton. *Appl. Environ. Microbiol.* 71 (7) (2005) 3483–3494.
- [3] P. Uribe, R.T. Espejo, Effect of associated bacteria on the growth and toxicity of *Alexandrium catenella*. *Appl. Environ. Microbiol.* 69 (1) (2003) 659–662.
- [4] Q. Zheng, C. Chen, X.-J. Yan, Y.-N. Wang, Y.-H. Zeng, L.-K. Hao, W.-H. He, N.-Z. Jiao, *Mameliella alba* gen. nov., sp. nov., a marine bacterium of the Roseobacter clade in the order Rhodobacterales. *Int. J. Syst. Evol. Microbiol.* 60 (4) (2010) 953–957.
- [5] I. Wagner-Döbler, H. Biebl, Environmental biology of the marine Roseobacter lineage. *Annu. Rev. Microbiol.* 60 (2006) 255–280.
- [6] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, ... P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19 (2012) 455–477.
- [7] M. Boetzer, C.V. Henkel, H.J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27 (2011) 578–579.
- [8] M. Boetzer, W. Pirovano, Toward almost closed genomes with GapFiller. *Genome Biol.* 13 (2012) R56.
- [9] T. Seemann, Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30 (2014) 2068–2069.

- [10] D. Hyatt, G.L. Chen, P.F. Locascio, M.L. Land, F.W. Larimer, L.J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.* 11 (2010) 119.
- [11] K. Lagesen, P. Hallin, E.A. Rodland, H.H. Staerfeldt, T. Rognes, D.W. Ussery, RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35 (2007) 3100–3108.
- [12] D. Laslett, B. Canback, ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32 (2004) 11–16.
- [13] O.S. Kim, Y.J. Cho, K. Lee, S.H. Yoon, M. Kim, H. Na, ... J. Chun, Introducing EzTaxon: a prokaryotic 16S rRNA Gene sequence database with phylotypes that represent uncultured species. *Int. J. Syst. Evol. Microbiol.* 62 (2012) 716–721.
- [14] J. Goris, K.T. Konstantinidis, J.A. Klappenbach, T. Coenye, P. Vandamme, J.M. Tiedje, DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57 (1) (2007) 81–91.
- [15] P. Jones, D. Binns, H.Y. Chang, M. Fraser, W. Li, C. Mcanulla, H. McWilliam, ... S. Hunter, InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30 (2014) 1236–1240.
- [16] V.M. Markowitz, I.-M.A. Chen, K. Palaniappan, K. Chu, E. Szeto, Y. Grechkin, A. Ratner, et al., IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* 40 (D1) (2012) D115–D122.
- [17] M.H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M.A. Fischbach, ... R. Breitling, antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39 (no. suppl 2) (2011) W339–W346.