De novo whole genome sequencing data of two mangrove-isolated microalgae from Terengganu coastal waters

Kit Yinn Teh, C.L.Wan Affudeen, Ahmad Aziz, Li Lian Wong, Saw Hong Loh, Thye San Cha

A R T I C L E   I N F O

Article history:
Received 15 August 2019
Received in revised form 23 September 2019
Accepted 11 October 2019
Available online 18 October 2019

Keywords:
Next generation sequencing
Oleaginous microalgae
Salinity
IDBA-UD
Chlorophyta

A B S T R A C T

Interest in harvesting potential benefits from microalgae renders it necessary to have the many ecological niches of a single species to be investigated. This dataset comprises de novo whole genome assembly of two mangrove-isolated microalgae (from division Chlorophyta); *Chlorella vulgaris* UMT-M1 and *Messastrum gracile* SE-MC4 from Universiti Malaysia Terengganu, Malaysia. Library runs were carried out with 2× 150 base paired-ends reads, whereas sequencing was conducted using Illumina Novaseq 2500 platform. Sequencing yielded raw reads amounting to ~11 Gb in total bases for both species and was further assembled de novo. Genome assembly resulted in a 50.15 Mbp and 60.83 Mbp genome size for UMT-M1 and SE-MC4, respectively. All filtered and assembled genomic data sequences have been submitted to National Centre for Biotechnology Information (NCBI) and can be located at DDBJ/ENA/GenBank under the accession of VJNP00000000 (UMT-M1) and VIYE00000000 (SE-MC4).

© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Response of microalgae to environmental stimuli is species-specific and may even vary from strain to strain [1,2]. Moreover, mangrove dwelling microalgae are often exposed to impending high and low tides making them unique assemblages in a marginal ecosystem niche with possibly unique responses. Being able to regulate and exert control over the outcome of those responses remain as the most difficult conundrums in phycology research. Both UMT-M1 and SE-MC4 used in this research are oleaginous native species isolated from the mangrove areas in Terengganu, Malaysia. UMT-M1 has been intensively studied in our previous research for oil and fatty acid productions under various culture conditions, such as nitrogen starvation [3], phytohormones treatments [4–6], as well as strain improvement through genetic modifications [7,8]. On the other hand, SE-MC4 is a non-model species which has been observed to produce more than 50% (of dry weight) of total oil content in our laboratory. The exploration on novel genome in a non-model microalga is imperative in order to enrich the available genome data for further biodiesel development applications.

Efforts to improve microalgae feedstock from a molecular aspect is often curtailed by the limited number of available microalgae genomes [9]. Moreover, available C. vulgaris genome only constitutes a freshwater species [10]. Following in that prospect, the de novo WGS of C. vulgaris UMT-M1 featured in this report represents a mangrove dwelling microalga that is able to adapt and survive in a wide range of salinity. Besides that, exploration of potentially high-oil producing non-model species such as M. gracile SE-MC4 is pertinent for adding genetic variety to the presently available genetic databank [11].

In UMT-M1, subsequent sequencing generated 73,495,318 raw reads, amounting to 11,097,793,018 (11.09 Gb) in total bases (Table 1). Overall, 89.58% of total bases achieved a Phred score of Q30 with GC content of 62.29%. High quality raw reads from Table 1 were then filtered, normalized and assembled de novo using IDBA-UD assembler [12]. The IDBA-UD assembler internally pipes contigs into scaffolds to form assembled scaffolds. Scaffolds with less than 200 bases were removed. Assembly produced
2547 scaffolds amounting to a total of 50,153,796 bases (50 Mbp). The scaffold positioned at the N50 and N90 were 56,390 and 14,886 bases, respectively (Table 2).

In SE-MC4, total bases generated from sequencing amounted to 10,984,065,858 bp (10.98 Gb) with 68.27% GC content and a Phred score of 90.52%. Sequencing data statistics are summarised in Table 1. De novo assembly in SE-MC4 obtained 32,473 scaffolds and a total length of 60,830,643 bp (60.83 Mb) with maximum length of 52,109 bp and minimum length of 201 bp. Mean length (N50) of scaffolds is 2915 bp, while N90 is 802 bp. Statistics of the genome assembly are as shown in Table 2.

2. Experimental design, materials, and methods

2.1. Sample preparation

Inoculum stock was obtained from microalgae culture collection at the Universiti Malaysia Terengganu. Stock cultures were maintained under axenic and sterile culture conditions in modified Guillard’s F2 medium [3] prepared with artificial seawater (30 ppt). Microalgae cells were harvested at mid-stationary phase. Microalgal cells were harvested from 50 mL of culture by centrifugation at 7000 rpm for 5 min. DNA was extracted from fresh pellet using Wizard® Genomic DNA Puriﬁcation Kit (Promega, USA). All extraction steps were carried out as per manufacturer’s protocol. Prior to sequencing, DNA purity was evaluated via absorbance values of (260/280, 260/230) ratio, gel electrophoresis pattern and double-strand DNA concentration measurements.

2.2. De novo WGS sequencing

Library preparation and sequencing were conducted by Theragen Bio Itex, South Korea. Library preparation was carried out using TruSeq Nano DNA Library Prep Kit (Illumina, USA). Library construction was made by DNA size selection attached with adaptors to produce an insert size of 350 bp [13]. Runs were conducted with 2 × 150 base paired-end reads. Sequencing was then performed on Illumina Novaseq 2500 platform. Cluster generation on flow cells was performed by using constructed libraries on cBot equipment (Illumina, USA). Following sequencing of raw reads, adapter sequences were trimmed via cutadapt v1.10 [14] and quality filtering was performed to remove contaminants. Reads that scored above Q30 were selected for assembly. De novo assembly of high quality reads was then carried out using IDBA-UD assembler to form scaffolds [12]. Scaffolds that were <200 bp in length were removed manually.

2.3. Deposition of genome data

Raw data sequence and assembled genome were deposited in NCBI depository portal. Steps by steps guidelines on submission was followed as in NCBI author guide via https://www.ncbi.nlm.nih.gov/genbank/genomesubmit/. Breakdown of the project accession is shown in Table 3.
Table 3
Sequence accession numbers and directory links.

| Species            | Directory/Data | Accession number | Links                                |
|--------------------|----------------|------------------|--------------------------------------|
| C. vulgaris UMT-M1 | BioProject     | PRJNA550188      | https://www.ncbi.nlm.nih.gov/bioproject/PRJNA550188 |
|                    | BioSample      | SAMN121111214    | https://www.ncbi.nlm.nih.gov/biosample/SAMN121111214 |
|                    | Raw sequence (SRA) | SRR9478717     | https://www.ncbi.nlm.nih.gov/sra/SRR9478717/ |
|                    | Assembled genome | VJNP000000000   | https://www.ncbi.nlm.nih.gov/nuccore/VJNP000000000 |
| M. gracile SE-MC4  | BioProject     | PRJNA550185      | https://www.ncbi.nlm.nih.gov/bioproject/PRJNA550185 |
|                    | BioSample      | SAMN121111213    | https://www.ncbi.nlm.nih.gov/biosample/SAMN121111213 |
|                    | Raw sequence (SRA) | SRR9587833     | https://www.ncbi.nlm.nih.gov/sra/SRRX6353668/ |
|                    | Assembled genome | VIYE000000000   | https://www.ncbi.nlm.nih.gov/nuccore/VIYE000000000 |

Acknowledgments

This research was supported by Japan Science and Technology Agency/Japan International Cooperation Agency, Science and Technology Research Partnership for Sustainable Development through the project for Continuous Operation System for Microalgae Production Optimized for Sustainable Tropical Aquaculture (COSMOS), and the SATREPS-COSMOS Matching Fund from the Ministry of Education Malaysia (MOE) (Vot No: 53222).

Equipment used in this study was obtained with financial support from the Japan Science and Technology Agency/Japan International Cooperation Agency, Science and Technology Research Partnership for Sustainable Development through the project for Continuous Operation System for Microalgae Production Optimized for Sustainable Tropical Aquaculture (COSMOS).

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] E.S. Salama, S.P. Govindwar, R.V. Khandare, H.S. Roh, B.H. Jeon, X. Li, Can omics approaches improve microalgal biofuels under abiotic stress? Trends Plant Sci. 24 (2019) 611–624.
[2] J.C. Quinn, R. Davis, The potentials and challenges of algae based biofuels: a review of the techno-economic, life cycle, and resource assessment modeling, Bioresour. Technol. 184 (2015) 444–452.
[3] T.S. Cha, J.W. Chen, E.G. Goh, A. Aziz, S.H. Loh, Differential regulation of fatty acid biosynthesis in two Chlorella species in response to nitrate treatments and the potential of binary blending microalgae oils for biodiesel application, Bioresour. Technol. 102 (2011) 10633–10640.
[4] M. Jusoh, S.H. Loh, T.S. Chuah, A. Aziz, T.S. Cha, Elucidating the role of jasmonic acid in oil accumulation, fatty acid composition and gene expression in Chlorella vulgaris (Trebouxiophyceae) during early stationary growth phase, Algal Res 9 (2015) 14–20.
[5] M. Jusoh, S.H. Loh, T.S. Chuah, A. Aziz, T.S. Cha, Indole-3-acetic acid (IAA) induced changes in oil content, fatty acid profiles and expression of four fatty acid biosynthetic genes in Chlorella vulgaris at early stationary growth phase, Phytochemistry 111 (2015) 65–71.
[6] M. Jusoh, S.H. Loh, A. Aziz, T.S. Cha, Gibberellin promotes cell growth and induces changes in fatty acid biosynthesis and upregulates fatty acid biosynthesis genes in Chlorella vulgaris UMT-M1, Appl. Biochem. Biotechnol. 188 (2018) 450–459.
[7] C.C. Lau, S.H. Loh, A. Aziz, T.S. Cha, Effects of disrupted omega-3 desaturase gene construct on fatty acid composition and expression of four fatty acid biosynthetic genes in transgenic Chlorella vulgaris, Algal Res 26 (2017) 143–152.
[8] M.N. Narashikin, S.H. Loh, A. Aziz, T.S. Cha, Metabolic engineering of fatty acid biosynthesis in Chlorella vulgaris using an endogenous omega-3 fatty acid desaturase gene with its promoter, Algal Res 31 (2018) 262–275.
[9] T.S. Cha, J.W. Chen, E.G. Goh, A. Aziz, S.H. Loh, Differential regulation of fatty acid biosynthesis genes in transgenic Chlorella vulgaris, Algal Res 26 (2017) 143–152.
[10] M.T. Guarnieri, J. Levering, C.A. Henard, J.L. Boone, M.J. Betenbaugh, K. Zengler, E.P. Knoshaug, Genome sequence of the oleaginous green alga, Chlorella vulgaris UTEX 395, Front. Bioeng. Biotechnol. 6 (2018) 37.
[11] T.G. Da Silva, C. Bock, C.J. Sant’Anna, I.L. Bagatini, S. Wodniok, A.A.H. Vieira, Selenastraceae (Sphaeropleales, Chlorophyceae): RbcL, Igs rDNA and ITS–2 secondary structure enlightens traditional taxonomy, with description of two new genera, Messaristrum gen. nov. and Curvastrum gen. nov. Fottea 17 (2017) 1–19.
[12] Y. Peng, H.C.M. Leung, S.M. Yi, F.Y.L. Chin, IDBA-UD, A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth, J. Bioinf. 28 (2012) 1420–1428.
[13] S.R. Head, H. Kiyomi Komori, S.A. LaMere, T. Whisenant, F. van Nieuwenburgh, D.R. Salomon, P. Ordoukhian, Library construction for next-generation sequencing: overviews and challenges, Biotechniques 56 (2014) 61–77.
[14] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMNet. J. 17 (2011) 10–12.