Genomes of Diverse Isolates of Prochlorococcus High-Light-Adapted Clade II in the Western Pacific Ocean

Wei Yan 1,*, Xuejin Feng 2, Wei Zhang 1, Muhammad Zohaib Nawaz 2, Tingwei Luo 2, Rui Zhang 2,3,* and Nianzhi Jiao 2,*

1 College of Marine Science and Technology, China University of Geosciences, Wuhan, China, 2 State Key Laboratory of Marine Environmental Science, Fujian Key Laboratory of Marine Carbon Sequestration, College of Ocean and Earth Sciences, Xiamen University, Xiamen, China, 3 Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai, China

Keywords: Prochlorococcus, high-light-adapted clade II, genomes, the western Pacific Ocean, co-cultured heterotrophic bacteria

INTRODUCTION

Members of the cyanobacterium genus Prochlorococcus are the most abundant photosynthetic organisms in global oceans (Chisholm et al., 1992; Partensky et al., 1999), and contributes ∼10% of ocean primary productivity (Flombaum et al., 2013). Prochlorococcus clades (ecotypes) are generally divided into viz. high-light-adapted (HL) clades and low-light-adapted clades based on the physiological characteristics, ecological distribution, and phylogeny (Moore and Chisholm, 1999; Rocap et al., 2002; Johnson et al., 2006). With diversified ecotypes, Prochlorococcus maintains a high genomic diversity, and has evolved continuously during the process of adapting to the marine environment (Bill et al., 2015). At least Fifty-two genomes of the Prochlorococcus genus were published (Bill et al., 2014; Yan et al., 2018a,b; Becker et al., 2019).

To date the Prochlorococcus HL clade II (HLII) is regarded as the dominant ecotype in global oceans and accounts for more than 90% of all Prochlorococcus in the upper layer of tropical waters (Johnson et al., 2006), exhibiting a fairly large repertoire of genomic and functional diversity (Partensky and Garczarek, 2010; Kashtan et al., 2014; Biller et al., 2015). HLII has gained significant research interests due to its streamlined genome, making it a model to study genome reduction (Kettler et al., 2007; Partensky and Garczarek, 2010; Biller et al., 2015). The western Pacific Ocean, having both a local and global climate impact (McPhaden et al., 2006), is well-known for low nutrient levels, low primary production, and strong light radiation (Schneider and Zhu, 1998). In addition, as part of the warmest ocean waters, the western Pacific Ocean represents an ideal site to study the effect of rising temperatures on the marine ecosystem (Rowe et al., 2012). However, only five Prochlorococcus genomes are reported in this region to date (Bill et al., 2014; Yan et al., 2018a,b; Becker et al., 2019).

The present study reports 15 HLII Prochlorococcus and 101 co-cultured heterotrophic bacterial genomes in the western Pacific Ocean and the South China Sea. The genomes discussed here have been deposited in the National Center for Biotechnology Information, and require further analysis to explore the fine-scale diversity of Prochlorococcus and their future applications in marine microbiology and ecology.

MATERIALS AND METHODS

Isolation of the Prochlorococcus HLII Strains

The Prochlorococcus HLII strains discussed in the present study were isolated from a depth of 50–150 m at seven different stations in the western Pacific Ocean and the South China Sea in 2014.
The isolation process was performed as previously described (Yan et al., 2021). Briefly, seawater collected by a Niskin bottle was subjected to gravity filtration through double polycarbonate filters (Millipore, USA) with a pore size of 0.6 µm (Chisholm et al., 1992). Then, a Pro2 medium nutrient stock solution was added to the filtrate (Moore et al., 2007). The filtrate was placed in an incubator onboard for initial enrichment for 4–8 weeks. After confirmation by a flow cytometer, the Prochlorococcus strains were maintained at a constant temperature of 22°C and a continuous light intensity of 10–20 µmol photons m⁻² s⁻¹.

**DNA Isolation, Library Preparation, and DNA Sequencing**

DNA isolation, library preparation, and DNA sequencing were performed as previously described (Yan et al., 2021). Briefly, genomic DNA was collected from 25 ml laboratory cultures by centrifugation (10,000×g for 30 min) and extracted using a QIAamp DNA mini kit (Qiagen, Germany). One µg of extracted DNA was fragmented by a Covaris ME220 Focused-ultrasonicator (Covaris, USA). DNA library was constructed using a NEBNext Ultra™ DNA Library Prep Kit for Illumina in accordance with the manufacturer's instructions (NEB, USA). Ten ng of library DNA was taken and subjected to bidirectional sequencing using an Illumina NovaSeq 6000 instrument with a read length of 150 bp. All library construction and sequencing were performed at Shanghai Majorbio Bio-pharm Technology Co., Ltd (Shanghai, China).

**Assembly and Annotation**

To recover Prochlorococcus and heterotrophic bacterial genomes from non-axenic cultures, genome assembly was performed using the MetaWRAP v1.2.1 pipeline on a Linux cluster with 96 cores and 512 GB of RAM (Uritskiy et al., 2018). Briefly, the reads from all samples were trimmed using the metaWRAP-Read_qc module and then individually assembled with the metaWRAP assembly module using MEGAHIT as a metagenomic assembler (Li et al., 2015). Bins were calculated using three binning modules, including CONCOCT (Alneberg et al., 2014), MetaBat (Kang et al., 2015), and MaxBin2 (Wu et al., 2016). Then, the bin refinement module was used to combine and improve the results generated by the three binners. Finally, the reassemble_bins module was used to attain better bins. The quality cutoffs of all steps were set as: completeness > 90% and contamination < 5% using CheckM (Parks et al., 2015). The assembled genomes were annotated using the Rapid Annotation using Subsystem Technology online server (FIGfam version Release 70) (Aziz et al., 2008).

**Phylogenomic Analysis**

Phylogenetic relationships of HLII Prochlorococcus strains were reconstructed by using concatenated protein sequences. Briefly, protein sequences of 1102 core genes defined at 70% similarity level were aligned with MUSCLE (Edgar, 2004) and concatenated using Bacterial Pangenome Analysis Pipeline (https://bpaap.berkeley.edu/).
v1.3.0 (Chaudhari et al., 2016). Gblocks software v0.91b was used for degaping in the divergent regions (Castresana, 2000). The neighbor joining phylogenetic tree with 1000 bootstraps was constructed using MEGA X (Kumar et al., 2018), and visualized using iTOL version 4 (Letunic and Bork, 2019).
Figure 2 | Phylogenetic tree of 101 co-cultured heterotrophic bacteria genomes, with the Synechococcus WH5701 strain as an outgroup. A maximum likelihood phylogenetic tree was reconstructed using the concatenated amino acid sequences of 120 bacterial ortholog genes. Bootstrap values for 1,000 resamplings are indicated by numbers at the nodes (at least 50% support).
Evolutionary relationships and taxonomic classification of co-cultured heterotrophic bacteria were reconstructed based on concatenated amino acid sequences of 120 bacterial ortholog genes using GTDB-tk v1.3.0 (Chaumeil et al., 2019). Briefly, the 120 ortholog genes were identified and aligned with HMMER (Finn et al., 2011), concatenated into a single multiple sequence alignment, and trimmed with the 5,000-column bacterial mask (Chaumeil et al., 2019). The maximum likelihood phylogenetic tree was constructed using FastTree v2.1.10 (Price et al., 2009). The phylogenetic tree presented in the present study was visualized using iTOL version 4 (Letunic and Bork, 2019).

Pan-Genome Analysis
Gene-based pan-genome analysis of the HLII Prochlorococcus strains was conducted using the Bacterial Pangenome Analysis Pipeline v1.3.0 (Chaudhari et al., 2016), which uses the USEARCH algorithms to identify core genes. The similarity cutoff for amino acid sequences of the core gene was set as 50%.

INTERPRETATION OF THE DATA
Genomic Data of the Prochlorococcus HLII Strains
In the present study, The DNA of 15 Prochlorococcus HLII cultures were sequenced, and the raw data were quilty controlled and assembled. Genome sizes of the 15 HLII isolates ranged from 1,631,569 bp to 1,721,994 bp, with an average GC content of 31.26% (s.d. = 0.11) and an average of 1,976 coding sequences per genome (s.d. = 28) (Table 1). Their phylogenetic position was confirmed through phylogenomic tree construction using core genome amino acid sequences at a similarity cutoff of 70% (Figure 1). In this study, 101 co-cultured heterotrophic bacterial genomes were obtained from the non-axenic HLII cultures using the binning method. The completeness of these genomes ranged from 90.2 to 100%, with an average of 97.7%. The completeness, contamination, genome size, GC content, and number of coding sequences of each genome are shown in Supplementary Table 1. Their phylogenetic positions were confirmed through phylogenetic trees using 120 bacterial orthologous genes (Figure 2). The phylogenetic tree comprised four bacterial clades, including Alphaproteobacteria, Rhodothermia, Gammaproteobacteria, and Bacteroidia, and was further separated into six classes, 22 families, and 37 genera (Figure 2; Supplementary Table 1).

REUSE POTENTIAL
The genomes of the Prochlorococcus HLII strains and their co-cultured heterotrophic bacteria discussed here warrant further analysis to explore the fine-scale diversity of Prochlorococcus and co-cultured heterotrophic bacteria, and their future applications in marine microbiology and ecology.

DATA AVAILABILITY STATEMENT
The raw sequence data and assembled genome data of the 15 HLII Prochlorococcus strains and 101 co-cultured heterotrophic bacteria reported in the present study have been deposited in the National Center for Biotechnology Information GenBank under the BioProject number PRJNA664924. Accession numbers for the individual genomes used in this study are listed in Table 1 and Supplementary Tables 1, 2.

AUTHOR CONTRIBUTIONS
Wy, Rz, and nj designed the study. Wy conducted the isolation of Prochlorococcus strains. Xf, Wz, and tl participated in genome analysis. Wy, mn, and rz wrote the manuscript. All authors contributed to the final version of the manuscript.

FUNDING
This work was supported by National Natural Science Foundation of China (41861144018 to wy, 91751207 to tl), China Ocean Mineral Resources R&D Association (DY135-E2–1–4 to nj), the Fundamental Research Funds for the Central Universities (20720190141 to wy, 20720170107 to nj). The Senior User Project of RV KEXUE (KEXUE2020G10 to RZ).

SUPPLEMENTARY MATERIAL
The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2020.619826/full#supplementary-material

REFERENCES
Alneberg, J., Bjarnason, B. S., de Bruijn, L., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. Nat. Methods 11, 1144–1146. doi: 10.1038/nmeth.3103
Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75. doi: 10.1186/1471-2164-9-75
Becker, J. W., Hogle, S. L., Rosendo, K., and Chisholm, S. W. (2019). Co-culture and biogeography of Prochlorococcus and SAR11. ISME J.13, 1–14. doi: 10.1038/s41396-019-0365-4
Biller, S. J., Berube, P. M., Berta-Thompson, J. W., Kelly, L., Roggensack, S. E., Awad, L., et al. (2014). Genomes of diverse isolates of the marine cyanobacterium Prochlorococcus. Sci. Data 1:140034. doi: 10.1038/sdata.2014.34
Biller, S. J., Berube, P. M., Lindell, D., and Chisholm, S. W. (2015). Prochlorococcus: the structure and function of collective diversity. Nat. Rev. Micro. 13, 13–27. doi: 10.1038/nrmicro3378
Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Frontiers in Marine Science | www.frontiersin.org 5 February 2021 | Volume 7 | Article 619826
