

Short Research Communication

Draft Genome Sequence of *Providencia sneebia* Strain ST1, a Quorum Sensing Bacterium Associated with Marine Microalgae

Jin Zhou, Yong-Min Lao, Zhong-Hua Cai ✉

The Division of Ocean Science and Technology, Graduate School at Shenzhen, Tsinghua University, Shenzhen, 518055, Guangdong Province, P. R. China.

✉ Corresponding author: ZH Cai, Address: Room 304, L-Building, Graduate School at Shenzhen, Tsinghua University, Shenzhen University Town, Xili Town, Shenzhen City, 518055, Guangdong Province, P R. China. Tel: +86-755-26036108; Fax: +86-755-26036108. E-mail address: caizh@sz.tsinghua.edu.cn

© Ivyspring International Publisher. Reproduction is permitted for personal, noncommercial use, provided that the article is in whole, unmodified, and properly cited. See <http://ivyspring.com/terms> for terms and conditions.

Published: 2016.03.10

Abstract

Providencia sneebia strain ST1 is a symbiotic bacterium (belonging to phylum gammaproteobacteria) with marine microalgae. This bacterium exhibits the ability to produce N-Acyl homoserine lactone signal molecule. To date, no genome that originates from marine *Providencia* spp. has been reported. In this study, we present the genome sequence of this strain. It has a genome size of 4.89 M, with 19 contigs and an average G+C of 51.97%. The function of 4,631 proteins was predicted, and 3,652 proteins were assigned to COG functional categories. Among them, 407 genes are involved in carbohydrate metabolism, 306 genes participate in nitrogen utilization and energy conversion, and 185 genes related to signal transduction process. Thus, this strain plays an active role in the biogeochemical cycle in algal life history. The whole-genome of this isolate and annotation will help enhance understanding of bacterial ecological behavior in the phycosphere.

Key words: *Providencia sneebia*, Marine bacteria, Genome sequence, N-Acyl homoserine lactone

Introduction

Cell-cell communication in bacteria is accomplished through the exchange of chemical signaling molecules called autoinducers (also called cell-density regulates factors). This process, termed quorum sensing (QS; especially the N-Acyl homoserine lactone, AHL), allows bacterial populations to coordinate gene expression (1). Marine environments, such as those in the phycosphere area, are abundant in nutrients and rich with diverse populations of microorganisms. Interactions of microalgae and bacteria affect the physiology of both partners, alter the chemistry of their environment, and shape ecosystem diversity (2). Previous studies confirmed that algae are greatly affected by symbiotic bacteria, and these interactions are mediated through

the production and exchange of infochemicals (2-4). Understanding signal language may shed light on the interaction between algae-associated microbial communities in the native host.

Providencia sneebia strain ST1 was isolated from the dinoflagellate (*Scrippsiella trochoidea*) in Shenzhen seacoast, Guangdong Province, China, using a seawater LB medium. *P. sneebia* ST1 belongs to γ -proteobacteria based on its 16S-rRNA sequences (5). This strain is a Gram-negative, aerobic, motile, and long-rod shape bacterium. Its optimal growth temperature was 30 °C (5). The *P. sneebia* ST1 featuring high efficiency in utilize nitrogen and competitiveness in algae-bacteria symbiosis, and can be applied as a potential algae-inhibitor (5). The ability to inhibit

algae probably is owing to its cell-density, which is modulating by quorum sensing substance (e.g. AHL molecules). The screening experiment in our laboratory using AHL biosensor *Chromobacterium violaceum* CV026, showed that this isolate possesses AHL activities (5). The QS property of ST1 strain has been observed, but the gene responsible for its AHL production remains unknown. Thus, we performed whole genome sequencing of this bacterium with the ultimate goal of searching for its AHL synthase gene.

Data description

The genomic DNA of *P. sneebia* ST1 bacterium was extracted and purified using genomic DNA extraction and a clean kit, respectively (Mo Bio, CA, USA). This process was conducted following the protocol of the manufacturer. The concentration and quality of the genomic DNA was detected using a nano-drop spectrophotometer (Thermo Scientific) and gel electrophoresis (Bio-Rad, USA), respectively. Whole genome sequencing of the normalized DNA was performed using Illumina Hiseq 2000 (Illumina, San Diego, CA) by generating paired-end libraries with an insert size of 476 bp and mate-paired libraries with an insert size of 6,020 bp. The read length was set to 125 bp. The raw sequence data were trimmed and subsequently assembled. De novo assembly was performed using CLC Genomics Workbench Version 5.1 (CLC Bio, Denmark) and trimmed using a minimum Phred quality scores of 20, a minimum length of 50 bp, allowing no ambiguous nucleotides and trimming off some low quality nucleotides. The trimmed sequences were assembled using CLC's de novo assembly algorithm, using a minimum scaffold length of 1 K bp. The predicted coding sequences (CDSs) were translated and used to search the NCBI nonredundant and other databases. The gene prediction was performed with a prokaryote gene prediction algorithm using Prodigal (version 2.6) (6), while rRNA and tRNA were predicted with RNAmmer (7) and tRNAscan-SE version 1.21 (8), respectively. Subsequently, the strain sequence was annotated with RAST (9). Additional gene prediction analysis manual functional annotations were performed within the Integrated Microbial Genomes (IMG) platform (<http://img.jgi.doe.gov>).

The genome features of *P. sneebia* ST1 are summarized in Table 1. There were 4631 encoding gene sequences with an average size of 933 bp predicted using Glimmer version 3.02 (<http://www.cbcb.umd.edu/software/glimmer>), giving a coding intensity of 88.31%. The assembly of

the genome yielded 19 contigs with a total contig size of 4,881,765 bp. The largest contig size was 2,612,975 bp. The final draft genome of *P. sneebia* strain ST1 contained 4,891,646 bases and a G+C content of 51.97%. The analyses of the draft genomes identified 4,631 open reading frames, and 152 non-coding RNAs were found in the genome, including 81 tRNA, 19 rRNA, and 52 sRNA. Homologous comparison by BLAST found 3,652 CDS involving 21 functional COG groups and a part of the CDS involving 34 metabolic pathway KEGG groups.

Genome annotation on predicted genes was carried out by BLAST searches against a nonredundant protein sequence database and other databases available online, such as Clusters of Orthologous Groups (COG) (10) and Kyoto Encyclopedia of Genes and Genomes (KEGG) groups (11). Based on the functional categories of COG (<http://www.ncbi.nlm.nih.gov/COG/>), 407 genes are involved in carbohydrate metabolism and 306 genes participate in nitrogen utilization and energy conversion, which allows the microorganism to adapt to and compete in algae-bacteria symbiosis. After gene annotation analysis, 185 genes were related to signal transport and interaction processing. Potentially, these genes are a key feature of *P. sneebia* ST1 that enable it to release or receive all kinds of signals, including biological and chemical information. To the AHL signals, the AHL encoding gene (*LuxR*) was predicted to be located at contig 2, and the gene length is 599 bp. This gene has a relatively high identity of the *LuxR* gene of *Citrobacter freundii* (GenBank: WP_003025485.1). In addition, a putative AI-2 (autoinducer-2) production protein *LuxS* gene was also found. This protein (158 amino acids) had a 75% identity of the *LuxS* gene of *Vibrio harveyi* (GenBank: AF120098.1). The density regulate factors (AHL signal) allows *P. sneebia* ST1 to grow fast and outcompete its host (algae). Hence, it can be as a potential candidate to control algae, especially the harmful algae. This whole-genome sequence provides deeper understanding of the interactions between bacteria (*P. sneebia* ST1) and phytoplankton under AHLs' regulation, and may facilitate to develop new microecological methods to control harmful algal bloom.

Nucleotide sequence accession number

This whole-genome shotgun project has been deposited at GenBank under the accession number LJOF00000000. The version described in this paper is the first version LJOF00000000.1.

Table 1: Genome features of *Providencia sneebia* strain ST1.

Attributes	Values
Genome size	4.89 Mb
GC content %	51.97%
Number of Contigs	19
Total contig size	4,881,765
Largest contig	2,612,975
Scaffolds	8
Total scaffold size	4,891,646
Largest scaffold	3,448,390
Protein encoding genes	4631
tRNAs	81
rRNAs	19
Minisatellite DNA	54
Microsatellite DNA	3
Class of COG predicted function	34
Predicted AHL syntheses gene (<i>LuxR</i>) site	Contig 2
Potential encoding- <i>LuxR</i> gene length	599 bp

- Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007; 35: 3100-3108.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. The RAST server: rapid annotations using subsystems technology. *BMC Genomics.* 2008; 9: 75.
- Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 2000; 28: 33-36.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 2004; 32: 277-280.

Acknowledgements

This study was supported by S&T Projects of Shenzhen Science and Technology Innovation Committee of Shenzhen Municipality (JCYJ20130402145002375, JCYJ20140417115840286, JSGG20140519113458237, and CXZZ20150529165045063), and the Project of Marine Fishery Science and Technology & Industry Development of Administration of Ocean and Fisheries of Guangdong Province (A201503D07).

Competing Interests

The authors have declared that no competing interest exists.

References

- Chen X, Schauder S, Potier N, Dorsselaer AV, Pelczer I, Bassler BL, Hughson FM. Structural identification of a bacterial quorum-sensing signal containing boron. *Nature.* 2008; 415: 545-548.
- Amin SA, Hmelo LR, van Tol HM, Durham BP, Carlson LT, Heal KR, Morales RL, Berthiaume CT, Parker MS, Djunaedi B, Ingalls AE, Parsek MR, Moran MA, Armbrust EV. Interaction and signaling between a cosmopolitan phytoplankton and associated bacteria. *Nature.* 2015; 522(7554): 98-101.
- Amin SA, Parker MS, Armbrust EV. Interactions between diatoms and bacteria. *Microbiol Mol Biol Rev.* 2012; 76: 667-684.
- Jonsson PR, Pavia H, Toth G. Formation of harmful algal blooms cannot be explained by allelopathic interactions. *Proc Natl Acad Sci USA.* 2009; 106: 11177-11182.
- Lv H, Zhou J, Cai ZH. The dynamic variation process of quorum sensing bacterium in a *Scrippsiella Trochoidea* bloom. *Ecol Sci.* (in Chinese, English abstract). 2015; in press.
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010; 11: 119.
- Lowe TM, Eddy SR. tRNA scan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997; 25: 955-964.