Complete genome analysis of bacteriochlorophyll \(a\)-containing \(R.\) antarcticum ZS2-28\(^T\) reveals its adaptation to Antarctic intertidal environment

ZENG Yinxin\(^*\), YU Yong, LI Huirong, LUO Wei & DING Haitao

MNR Key Laboratory for Polar Science, Polar Research Institute of China, Shanghai 200136, China

Received 12 November 2020; accepted 8 March 2021; published online 18 March 2021

Abstract Aerobic anoxygenic phototrophic bacteria (AAPB) are photoheterotrophic prokaryotes able to use both light and dissolved organic matter as energy sources. \(R.\) antarcticum ZS2-28\(^T\) was isolated from intertidal sediment in the Larsemann Hills, Princess Elizabeth Land, Antarctica, and was able to produce bacteriochlorophyll \(a\). It is the type strain of the sole species within the genus \(R.\) antarcticum. The complete genome sequence of the bacterium was determined using Illumina HiSeq X and PacBio RSII systems. The genome of \(R.\) antarcticum ZS2-28\(^T\) was 4253095 bp and consisted of one chromosome and four plasmids. A number of genes related to the bacteriochlorophyll \(a\) production, photosynthetic reaction, cold adaptation, salt adaptation, ultra-violet resistance and DNA damage repairing were found in the genome. In addition to genomic islands and type IV secretion systems, genes related to gene transfer agents were detected in the genome of \(R.\) antarcticum ZS2-28\(^T\), suggesting that this bacterium can adapt to its environment by acquiring exogenous nucleic acids. The annotated complete genome sequence provides genetic insights into the environmental adaptation and ecological function of \(R.\) antarcticum ZS2-28\(^T\) in Antarctic coastal area.

Keywords \(R.\) antarcticum, complete genome, adaptation, gene transfer, strain, intertidal sediment, Antarctica


1 Introduction

Aerobic anoxygenic phototrophic bacteria (AAPB) are bacteriochlorophyll \(a\)-containing bacteria with the capability of photoheterotrophy; they appear to play a unique role in the ocean's carbon cycle (Swingley et al., 2007; Tang et al., 2010; Graham et al., 2018). Heterotrophy usually is the main system of energy gain of AAPB (Beatty, 2002), and phototrophy is minimal (Ferrera et al., 2017). However, light can enhance the growth rates of AAPB (Ferrera et al., 2017; Piwosz et al., 2018). AAPB are widely distributed in open and coastal oceans (Jiao et al., 2007), including Arctic and Antarctic marine environments (Boeuf et al., 2013; Zeng et al., 2016). AAPB are phylogenetically diverse and include members of the Alpha-, Beta- and Gammaproteobacteria (Imhoff et al., 2017; Lehours et al., 2018; Auladell et al., 2019). Physiological constraints play an important role in structuring AAPB assemblages at a global scale, and salinity seems to favor lineage-specific adaptations of AAPB (Lehours et al., 2018). AAPB belonging to Alphaproteobacteria and Betaproteobacteria dominate the offshore and river-influenced surface waters in the western Arctic Ocean, respectively (Boeuf et al., 2013). Represented by the orders \(Rhodospirillales,\) \(Rhizobiales\)
Genome analysis of *Roseicitreum antarcticum* ZS2-28T

2 Materials and methods

*R. antarcticum* ZS2-28T was isolated from sandy intertidal sediment samples in the Larsemann Hills, Princess Elizabeth Land, Antarctica (Yu et al., 2011). The bacterium was deposited into the China General Microbiological Culture Collection Center (CGMCC) with the accession number CGMCC 1.8894T and the Belgian Coordinated Collections of Micro-organisms (BCCM) with the accession number LMG 24863T. The complete chromosome sequence and four plasmid sequences of *R. antarcticum* ZS2-28T have been deposited in GenBank under the accession numbers CP061498, CP061499, CP061500, CP061501 and CP061502, respectively.

The general features of *R. antarcticum* ZS2-28T and MIGS mandatory information are listed in Table 1. Genomic DNA was extracted from overnight cultures using a MagAttract HMW DNA Kit (Qiagen, Germany) according to the manufacturer’s instructions. The harvested DNA was visualized on 1% (w/v) agarose gels, and DNA concentration and purity were measured with a Qubit 2.0 Fluorometer (Life Technologies, USA). Purified DNA was used to construct an Illumina standard shotgun library with an insert size of 300–400 bp followed the NEB Next Ultra DNA Library Prep Kit for Illumina (New England Biolabs, USA), and then was sequenced using the Illumina HiSeq X platform using the PE150 model. A 10-kb DNA library was constructed by the PacBio SMRTbell 10 kb Library preparation Kit (Pacific Biosciences, USA) according to the manufacturer’s instructions. Library construction and sequencing were performed at Sangon Biotech Co. Ltd (Shanghai, China). The whole genome sequencing was performed using Illumina HiSeq X (Illumina, USA) and PacBio RSII (Pacific Biosciences, USA) systems.

<table>
<thead>
<tr>
<th>Table 1 General features of <em>Roseicitreum antarcticum</em> ZS2-28T and MIGS mandatory information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Items</strong></td>
</tr>
<tr>
<td>Classification</td>
</tr>
<tr>
<td>Gram-staining</td>
</tr>
<tr>
<td>Cell shape</td>
</tr>
<tr>
<td>Motility</td>
</tr>
<tr>
<td>Temperature</td>
</tr>
<tr>
<td>Salinity</td>
</tr>
<tr>
<td>pH</td>
</tr>
<tr>
<td>Submitted to INSDC</td>
</tr>
<tr>
<td>Investigation type</td>
</tr>
<tr>
<td>Project name</td>
</tr>
<tr>
<td>Collection date a</td>
</tr>
<tr>
<td>Geographic location a</td>
</tr>
<tr>
<td>Environment (biome)</td>
</tr>
<tr>
<td>Environment (feature)</td>
</tr>
</tbody>
</table>
De novo genome assembly was performed using continuous long reads following the Canu workflow v1.3 (Koren et al., 2017), and then Pilon v1.23 was engaged to correct assembled contigs with Illumina reads (Walker et al., 2014). Annotation of the genome was generated by using Prokka v1.10 (Seemann, 2014) to predict coding sequences, ribosomal RNA genes, and transfer RNA genes. A whole genome Blast (v2.2.28) search was performed against the databases CDD, PFAM, COG, NR, Swiss-Prot, and TrEMBL. KEGG ontology was identified by submitting predicted peptides to the KAAS server (http://www.genome.jp/tools/kaas/). GO was detected from Swiss-Prot and TrEMBL annotation results. Four additional databases PHI, VFDB, CARD, and CAZy, were used to annotate peptides. Signal peptides were detected on the genome assembly by SignalP v4.1 (Petersen et al., 2011). Transmembrane proteins were detected by TMHMM v2.0 (Möller et al., 2001). Lipoproteins were detected with LipoP v1.0 (Juncker et al., 2003). Repeat regions within the genome were detected with RepeatModeler (http://www.repeatmasker.org/RepeatModeler/) and RepeatModeler (http://repeatmasker.org), and CRISPRs were detected with CRISPRCasFinder (https://crisprcas.i2bc.paris-saclay.fr). Genomic islands were predicted using IslandPath-DIOMB (Bertelli and Brinkman, 2018). The prophage regions were detected with PhiSpy (Akhter et al., 2012). Architecture of the photosynthesis gene cluster was produced using gggenes v0.4.1 (https://wilcox.org/gggenes/). Multiple sequence alignments (MSAs) were produced with ClustalW algorithm implemented in the MEGA 5.05 software package (https://www.megasoftwa re.net). DNA sequences of pufL and pufM genes were obtained from GenBank. pufL and pufM MSAs were concatenated to form a single pufLM MSA. A neighbor-joining phylogenetic tree was constructed based on pufLM gene sequences using MEGA 5.05.

3 Results and discussion

3.1 General description of the genome

The genome sequence of *R. antarcticum* ZS2-28T was obtained from a total of 3.0 Gb of high-quality data, which comprised 1.653 Gb Illumina HiSeq X data and 1.396 Gb PacBio RSII data. These data respectively represented 391- and 331-fold coverage of the genome. The genome of *R. antarcticum* ZS2-28T consists of one chromosome (3537072 bp, 63.45 mol% G + C), and four plasmids, named as pRA01 (425208 bp, 62.70 mol% G + C), pRA02 (131922 bp, 59.56 mol% G + C), pRA03 (103727 bp, 60.17 mol% G + C) and pRA04 (55166 bp, 58.51 mol% G + C), respectively (Table 2). Graphic circular maps of *R. antarcticum* ZS2-28T are shown in Figure 1.

### Table 2 Genomic features of *Rosenicicrium antarcticum* ZS2-28T

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Chromosome</th>
<th>pRA01</th>
<th>pRA02</th>
<th>pRA03</th>
<th>pRA04</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size/bp</td>
<td>3537072</td>
<td>425208</td>
<td>131922</td>
<td>103727</td>
<td>55166</td>
</tr>
<tr>
<td>(G + C) content/mol%</td>
<td>63.45</td>
<td>62.7</td>
<td>59.56</td>
<td>60.17</td>
<td>58.51</td>
</tr>
<tr>
<td>Total genes</td>
<td>3457</td>
<td>412</td>
<td>145</td>
<td>113</td>
<td>54</td>
</tr>
<tr>
<td>Protein coding genes</td>
<td>3407</td>
<td>411</td>
<td>145</td>
<td>113</td>
<td>54</td>
</tr>
<tr>
<td>rRNA genes</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>tRNA genes</td>
<td>44</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Genes assigned to COG</td>
<td>2471</td>
<td>311</td>
<td>75</td>
<td>52</td>
<td>20</td>
</tr>
<tr>
<td>Genes assigned to PFAM</td>
<td>2582</td>
<td>305</td>
<td>85</td>
<td>63</td>
<td>28</td>
</tr>
<tr>
<td>Genes assigned to KEGG</td>
<td>1439</td>
<td>203</td>
<td>26</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Transmembrane proteins</td>
<td>739</td>
<td>94</td>
<td>28</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Signal peptides</td>
<td>241</td>
<td>32</td>
<td>10</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
Genome analysis of *Roseicitreum antarcticum* ZS2-28<sup>T</sup>

The genome contains 45 tRNA genes and 2 rRNA operons. The numbers of 5S, 16S and 23S rRNA genes were all two. Among the 4181 predicted protein-coding genes, 3251 (78.93%), 2929 (71.11%), 3918 (95.12%), 3063 (74.36%), 2773 (67.32%), 2692 (65.36%), and 1856 (45.06%) were annotated by querying the CDD, COG, NR, PFAM, Swiss-Prot, TrEMBL, GO, and KEGG databases, respectively. There were 257 (6.14%) genes that failed to annotate in at least one database. Based on KEGG pathway classification (Figure 2), 74.77% of the annotated genes were found to be involved in metabolisms, including amino acid metabolism (17.75%), carbohydrate metabolism (14.11%), overview (11.61%), energy metabolism (6.76%), metabolism of cofactors and vitamins (6.75%), nucleotide metabolism (5.14%), xenobiotics biodegradation and metabolism (4.26%), lipid metabolism (4.26%), biosynthesis of other secondary metabolites (1.68%), metabolism of terpenoids and polyketides (1.58%), and glycan biosynthesis and metabolism (0.99%).

### 3.2 Genes related to photoheterotrophic lifestyle

Strain ZS2-28<sup>T</sup> is obligately heterotrophic (Yu et al., 2011), utilizing L-arabinose, cellobiose, D-galactose, gentiobiose, D-glucose, maltose, D-mannose, L-rhamnose, D-ribose, sucrose, trehalose, turanose, D-xylene, D-mannitol, adipic acid, gluconate, malic acid, glycerin, amyladal, pyruvate, casein hydrolysate and yeast extract as sole carbon and energy sources. This strain is also positive for arginine dihydrolase, lysine decarboxylase, ornithine decarboxylase, urease, β-galactosidase and β-glucosidase activities (Yu et al., 2011). The major COG categories of the genome were amino acid transport and metabolism (10.17%), carbohydrate transport and metabolism (8.71%), transcription (7.17%), replication, recombination and repair (7.10%), energy production and conversion (6.21%), inorganic ion transport and metabolism (5.84%), and translation, ribosomal structure and biogenesis (5.71%). Annotation based on the CAZy database indicates that strain ZS2-28<sup>T</sup> genome contains a large number of carbohydrate-active genes, including 18 auxiliary activities (AA), 6 carbohydrate-binding modules (CBM), 36 carbohydrate esterases (CE), 50 glycoside hydrolases (GH), 36 glycosyl transferases (GT), and 2 polysaccharide lyase (PL). In addition, annotation based on the PFAM database reveals a total of 93 peptidases and 23 proteases in the genome.
The strain ZS2-28<sup>T</sup> can produce Bchl <sub>a</sub> (Yu et al., 2011). In the biosynthesis of BChl <sub>a</sub>, conversion of chlorin to bacteriochlorin ring is known to be catalyzed by chlorophyllide <sub>a</sub> oxidoreductase (COR), which is a nitrogenase-like enzyme and a three-subunit complex consisting of BchX, BchY and BchZ (Tsukatani et al., 2013). Product of this reaction is further catalyzed by BchF and BchC (Tsukatani et al., 2013). According to genome information, the <i>bchXYZ</i> gene set, and <i>bchC</i> and <i>bchF</i> genes were present in strain ZS2-28<sup>T</sup> (Figure 3a), and showed more than 72% sequence similarities to those of Rhodobacteriaceae based on BlastX searching in NCBI database. There were 23 genes related to the formation of Bchl found in the chromosome based on GO annotation. Two proteins (Pfam05398 and Pfam07284) involved in Bchl biosynthesis pathway were observed. At the same time, similar to the oxygen regulated puf operon of purple photosynthetic bacterium <i>Rhodobacter sphaeroides</i> (Chidgey et al., 2017), <i>pufQ</i>, <i>pufB</i> and <i>pufA</i> genes encoding the light-harvesting (LH)1 α and β polypeptides, and <i>pufL</i>, <i>pufM</i> and <i>pufH</i> genes encoding the type-II photosynthetic reaction center (RC) L, M and H subunits were observed in the chromosome based on GO annotation. Two proteins (Pfam05398 and Pfam07284) involved in Bchl biosynthesis pathway were observed. At the same time, similar to the oxygen regulated puf operon of purple photosynthetic bacterium <i>Rhodobacter sphaeroides</i> (Chidgey et al., 2017), <i>pufQ</i>, <i>pufB</i> and <i>pufA</i> genes encoding the light-harvesting (LH)1 α and β polypeptides, and <i>pufL</i>, <i>pufM</i> and <i>pufH</i> genes encoding the type-II photosynthetic reaction center (RC) L, M and H subunits were observed in the genome of strain ZS2-28<sup>T</sup> (Lee et al., 1989; Hunter et al., 1991; Imhoff et al., 2019). In addition, a photosynthetic reaction center cytochrome C encoding gene <i>pufC</i> was found in the genome. The cytochrome associated with the photosynthetic reaction center is an important component in many of the PS-II type photosynthetic bacteria (Imhoff et al., 2019). However, different from <i>Rhodobacter sphaeroides</i>, <i>pufX</i> gene encoding the PufX polypeptide was absent from the genome of strain ZS2-28<sup>T</sup>. Three genes encoding proteins (Pfam02276 for photosynthetic reaction center cytochrome C subunit, Pfam03073 for sensory protein and Pfam04940 for blue light sensor protein) involved in photosynthesis were found in the chromosome. Phylogenetic analysis (Figure 3b) based on <i>pufL</i> and <i>pufM</i> genes indicated that strain ZS2-28<sup>T</sup> fell into the Roseobacter clade containing the genera <i>Jannaschia</i>, <i>Mameliella</i>, <i>Roseovarius</i>, <i>Tateyamaria</i> and <i>Thalassococcus</i>. The <i>pufL</i> and <i>pufM</i> genes of strain ZS2-28T showed close relationship (77.6% sequence similarity) to <i>Jannaschia</i> sp. CCS1.

The <i>acsF</i> gene encoding Mg-protoporphyrin IX monomethyl ester cyclase was observed in strain ZS2-28<sup>T</sup>, showing more than 80% sequence similarities to those of Rhodobacteriaceae based on BlastX searching in NCBI database. AcsF activity exists only under aerobic growth conditions in a Betaproteobacterium (Pinta et al., 2002). As an analogue of the oxygen-dependant cyclase encoded by <i>acsF</i> gene, <i>bchE</i> gene encoding the oxygen-independent Mg-protoporphyrin monomethylster cyclase was not detected in the genome of strain ZS2-28<sup>T</sup>. The <i>bchE</i> gene seems to be unessential for phototrophy in Roseobacter species (Koblížek et al., 2013). Results support that strain ZS2-28<sup>T</sup> is an aerobic photosynthetic purple bacterium, and are consistent with previous finding that strain ZS2-28<sup>T</sup> is photoheterotrophic (Yu et al., 2011).

### 3.3 Genes related to genetic exchange

Horizontal gene transfer (HGT) has been regarded to play an important role in the adaptation of the microbes to environment by providing the tools that are necessary to face the adversity and survive in a harsh environment (Springael and Top, 2003; Paquola et al., 2018). Consistent with the finding of a gene transfer agent (GTA) capsid
Figure 3  Photosynthesis genes in the genome of Roseicitreum antarcticum ZS2-28T. a, Genes in the potential photosynthesis gene cluster of R. antarcticum ZS2-28T; b, Neighbor-joining phylogenetic tree based on pufL and pufM gene sequences showing the relationship between strain ZS2-28T and representatives of some other related taxa. Numbers at nodes indicate percentages of 1000 bootstrap re-samplings, only values above 50% are shown. Bar, 0.1 substitutions per site.

protein gene (g5) in strain ZS2-28T (Zeng, 2019), a total of six genes related to GTA were observed in the chromosome based on NR annotation. GTAs have evolved from prophages that have lost the ability to target their own DNA for packaging (Lang et al., 2012). However, no prophage region was found in the genome of strain ZS2-28T. Neither repeated regions nor the CRISPR-Cas system were detected in the genome.

A total of 16 genomic islands were predicted in the chromosome. At the same time, five type IV secretion system genes were observed in the genome based on Swiss-Prot annotation. Type IV secretion systems and genomic islands-mediated horizontal gene transfer have been reported in Pseudomonas and Haemophilus (Juhas, 2015). Transposases and integrases can mediate the movement of DNA sequences within or between genomes (Rice and Baker, 2001). Strain ZS2-28T contains 74 transposase and 42 integrase genes, and annotation based on the Pfam database indicated that diverse transposases (Pfam02371, Pfam03050, Pfam03400, Pfam04986, Pfam05598, Pfam13005, Pfam13340, Pfam13586, Pfam13610, Pfam13737, and Pfam13751) and integrases (Pfam00589, Pfam00665, and Pfam13683) are present in the genome. These results suggest that R. antarcticum ZS2-28T may adapt to the environment by acquiring exogenous nucleic acids.

3.4 Genes related to adaptation to Antarctic intertidal environment

The combination of seasonal scouring and encasement in ice, high UV irradiation, and high levels of salinity and temperature fluctuations make the Antarctic intertidal zone possibly the world’s most physically disturbed environment (Peck et al., 2006). Cells of strain ZS2-28T were surrounded with slime (Yu et al., 2011). Slime is actually a kind of exopolysaccharides (EPS) produced by bacteria, and is helpful for bacteria surviving in extreme environments (Flemming, 2016). Four capsule polysaccharide biosynthesis proteins (Pfam05159) and 21 glycosyl
transfersases (Pfam00534, Pfam00535, Pfam00591, Pfam00953, Pfam00982, Pfam03808, Pfam04101, Pfam13632, Pfam13641 and Pfam13704) were detected in the genome.

One ultra-violet resistance protein (Pfam12344), showing 85% similarity to excinuclease ABC subunit UvrB of Rhodobium loti strain MAFF303099 based on Swiss-Prot database, was detected in the chromosome. At the same time, KEGG pathway annotation reveals that 8 genes encoding single-strand DNA-binding protein Ssb and ATP-dependent DNA helicase RecG are present in the genome. Those proteins may play role in repairing of DNA damage caused by ultra-violet irradiation (Trgovečević et al., 1989; Xu et al., 2020). It is interesting to find that all three AAPB in the Antarctic marine environment.

The genome sequence will improve our understanding of adaptive strategies required for survival in extreme environments. These strategies include ultra-violet resistance, cold adaptation, salt adaptation, photosynthesis, and horizontal gene transfer events. The genome sequence may provide insights into the evolution of these adaptations and ecological functions of Antarctic bacteria under low temperatures (Médigue et al., 2005).

Annotation based on the Pfam database indicates that diverse RNA helicases, including 4 DEAD/DEAH box helicases (Pfam00270), 1 DEAD/H associated (Pfam08494), 1 SecA DEAD-like domain (Pfam07517), and 3 helicase conserved C-terminal domains (Pfam00271) are present in the genome of strain ZS2-28T. Four genes related to cold-shock DNA-binding domain (Pfam00313) were also observed in the genome. At the same time, five salt adaptation-related genes, including ectABC, trkA and trkH (Kraegeloh et al., 2005; Sadeghi et al., 2014), were detected in the genome based on Pfam annotation. These results suggest that strain ZS2-28T is adapted to cold intertidal environment.

In conclusion, the genomic analysis of Roseicitreum antarcticum ZS2-28T isolated from Antarctic intertidal sediment has revealed that its genome contains various genes involved in the bacterium’s Bchl a production, photosynthetic reaction, cold adaptation, salt adaptation, ultra-violet resistance, and horizontal gene transfer events. The genome sequence will improve our understanding of the environmental adaptations and ecological functions of AAPB in the Antarctic marine environment.

Acknowledgments This work was supported by the National Key R & D Program of China (Grant no. 2018YFC1406903) and the National Natural Science Foundation of China (Grant nos. 91851201 and 41476171). We would like to thank two anonymous reviewers and Associate Editor, Dr Steve Coulson for their valuable suggestions and comments on further improvement of this article.

References


