Genome Sequencing and Analysis of the Psychrophilic Anoxygenic Phototrophic Bacterium *Rhodoferax antarcticus* sp. ANT.BR

by

Tingting Zhao

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved July 2011 by the
Graduate Supervisory Committee:

Jeffrey Touchman, Chair
Michael Rosenberg
Kevin Redding
Valerie Stout

ARIZONA STATE UNIVERSITY

August 2011

ABSTRACT

Rhodoferax antarcticus strain ANT.BR, a purple nonsulfur bacterium isolated from a microbial mat in Ross Island, Antarctica, is the first described anoxygenic phototrophic bacterium that is adapted to cold habitats and is the first beta-proteobacterium to undergo complete genome sequencing. *R. antarcticus* has unique absorption spectra and there are no obvious intracytoplasmic membranes in cells grown phototrophically, even under low light intensity. Analysis of the finished genome sequence reveals a single chromosome (3,809,266 bp) and a large plasmid (198,615 bp) that together harbor 4,262 putative genes. The genome contains two types of Rubiscos, Form IAq and Form II, which are known to exhibit quite different kinetic properties in other bacteria. The presence of multiple Rubisco forms could give *R. antarcticus* high metabolic flexibility in diverse environments. Annotation of the complete genome sequence along with previous experimental results predict the presence of structural genes for three types of light-harvesting (LH) complexes, LH I (B875), LH II (B800/850), and LH III (B800/820). There is evidence that expression of genes for the LH II complex might be inhibited when *R. antarcticus* is under low temperature and/or low light intensity. These interesting condition-dependent light-harvesting apparatuses and the control of their expression are very valuable for the further understanding of photosynthesis in cold environments. Finally, *R. antarcticus*

exhibits a highly motile lifestyle. The genome content and organization of all putative polar flagella genes are characterized and discussed.

DEDICATION

For the whole time I am in the U.S., my parents are always there supporting me. I have to say it is not easy to study abroad, but whenever I feel alone or upset, a phone call from them in China really makes me feel at home again and all the physical distance and time difference do not matter at all. There is no word to describe how grateful I am to have such wonderful parents, to love me, to understand me, to guide me, and to be friends with me. I never would have made it this far without them. I am sorry for not being around my parents, taking care of them, but I hope what I did here at Arizona State University make them really proud of me.

There is another person that I could never miss more. Jing, your love and support means the world to me. I really appreciate your patience and willing to grow up with me. And thank you for putting up with me all the time.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. Jeffrey Touchman. I could still remember the first time I saw him after I came to the U.S., a very charming, kind and considerate person who made my transition to the U.S. and his lab so easily and comfortable. And without his constant support and guidance, this thesis would never be possible and I would never become the person who I am. His help and faith in me has made my stay in his lab very enjoyable and my journey in science a wonderful time which I will remember for my whole life.

I would like to thank everyone in Dr. Touchman's lab, Renxia Huang who did some of the experiments with me and helped me deal with various problems all the time and Yih-Kuang Lu who gave me some suggestions about my thesis. I have spent a wonderful time with them.

I would also like to express my appreciation to my committee members, Dr. Kevin Redding, Dr. Michael Rosenberg, and Dr. Valerie Stout who gave me a lot of valuable advice on my thesis. And also the members in Dr. Ferran Garcia-Pichel's lab, Dr. Garcia-Pichel, Ankita Kothari, Qunjie Gao, Doerte Hoffmann, Hugo Beraldi, Natalie Myers and Ruth Potrafka. They helped me a lot when I did the rotation and I spent a good time there. And thank my mentor in the Molecular and Cellular Biology program, Karla Moeller, for her continued support, help and providing a sympathetic ear!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Extremophilic Anoxygenic Phototrophic Bacteria

Anoxygenic phototrophic bacteria are very good model organisms to investigate the origin and evolution of life, photosynthesis, and nitrogen fixation mechanisms (Yang et al., 2008), and more and more research efforts have been focused on this group of bacteria. Just in the year 2007, 19 new species of anoxygenic phototrophic bacteria were reported in the scientific literature. As more species of this type are described, many interesting findings as well as ongoing research projects in that area have shed light on the evolution of bacteria and photosynthesis. For instance, there is a newly emerging group of anoxygenic phototrophic bacteria that produce bacteriochlorophyll (Bchl) *a* and carotenoid pigments, but instead of displaying phototrophic growth under anaerobic conditions, which is a typical characteristic for the previously defined anoxygenic phototrophic bacteria, these newly discovered bacteria are grown aerobically. In this case, these bacteria are referred to as aerobic anoxygenic phototrophic bacteria (V. V. Yurkov & Beatty, 1998). Although similarities have been found in their electron transfer carriers, composition of their photosynthetic apparatuses (Garcia, Richaud, Breton, & Vermeglio, 1994; Okamura, Mitsumori, Ito, Takamiya, & Nishimura, 1986; Okamura, Takamiya, & Nishimura, 1985; Vladimir Yurkov, Gad'on, & Drews, 1993; V. Yurkov, Gad'on, N., Angerhofer,

A. and Drews, G., 1994; V. Yurkov, Schoepp, & Vermeglio, 1995), as well as the amino acid sequences of light-harvesting I polypeptides (Liebetanz, Hornberger, & Drews, 1991) and the reaction center between the aerobic phototrophic bacteria and anaerobic purple phototrophic bacteria, "efficient photoinduced electron transfer is operative only under aerobic conditions in the aerobic phototrophic bacteria" (Garcia, et al., 1994; Okamura, et al., 1985; V. Yurkov, et al., 1995) which is unusual and interesting. An excellent review (V. V. Yurkov & Beatty, 1998) discusses how aerobic anoxygenic phototrophic bacteria might enhance our understanding of the evolution of photosynthesis and the photometabolism of various phototrophic bacteria.

In particular, anoxygenic phototrophic bacteria living in extreme environments have historically attracted the attention of researchers seeking to expand our knowledge of how anoxygenic photosynthesis works in extreme environments, to discover photosynthetic diversity, evolution, and ecology, as well as to define the physiochemical limits of photosynthesis. Rapid advances in DNA sequencing technologies have made it possible for scientists to obtain complete microbial genomes of interest efficiently, so the new genetic resources represented by extremophilic phototrophs would be of great value to understand photosynthetic mechanisms under extreme environment. For example, data mining of these genetic resources may be useful in genetically engineering crop plants to thrive in some unusual climates (Madigan, 2003), like very hot, cold or

even dry environments. One crucial factor to define an extremophile is that "extreme environments are typically constant in their physiochemical properties, and thus the extreme is not simply a transient condition but rather an inherent part of the ecosystem" (Madigan & Marrs, 1997), which means "extremophiles do not just tolerate their extreme condition, but actually require it for optimal growth and metabolism" (Madigan & Marrs, 1997). Therefore, researchers have been exploring places where constant extreme conditions exist to isolate extremophilic anoxygenic phototrophs optimally adapted to their habitats.

For example, the first thermophilic anoxygenic phototroph *Chloroflexus aurantiacus*, a green nonsulfur bacterium whose optimum growth temperature is 55℃, was isolated from a hot spring (B. K. Pierson & Castenho.Rw, 1974; B.K. Pierson & Castenholz, 1974). For some strains, the growth can even occur up to 70℃ (B. K. Pierson & Castenho.Rw, 1974; B.K. Pierson & Castenholz, 1974), which is the upper temperature limit for photosynthesis by anoxygenic phototrophs so far, and is close to the limit for oxygenic photosynthesis (74℃, (Ward & Castenholz, 2000)). *Chloroflexus aurantiacus* has changed tremendously the way researchers think about green bacteria. Its unique pathway for autotrophic CO_2 fixation, the hydroxypropionate pathway, is absent from all other anoxygenic phototrophic bacteria. It is possible that the hydroxypropionate pathway was the first autotrophic pathway to have evolved in anoxygenic phototrophs (Madigan, 2003). Furthermore, chemotrophically grown cells of *C. aurantiacus* lack

chlorosomes and when switched to phototrophic conditions, the steps in chlorosome synthesis can be tracked (B. Pierson & Castenholz, 1995).

Besides thermophiles, some other extremophilic anoxygenic phototrophic bacteria were found to optimally grow at high pH (Imhoff, 1992), low pH (Pfennig, 1969, 1974), or high salt (Imhoff, 1992) (Table 1). Nevertheless, despite quite a few extremophilic anoxygenic phototrophic bacteria already discovered, species growing best at low temperatures, have only recently been described (Brenchley, 1996; Gounot, 1991), and no genome sequences are yet available to represent this group of bacteria. Therefore, my current study focuses on the first described example of psychrophilic anoxygenic phototrophic bacteria, *Rhodoferax antarcticus*, trying to obtain its complete genome sequence so as to fill the gap in the available genomic data for these photosynthetic prokaryotes. I also carry out comparative genome analysis with other psychrophiles or other related anoxygenic phototrophic bacteria to further explore some unique features of *Rhodoferax antarcticus* in order to understand further its physiology and ecology.

Table 1

Extremophilic anoxygenic phototrophic bacteria (Madigan, 2003).

Class	Organism	Habitat	Special properties
Thermophiles	*Chloroflexus aurantiacus*	Neutral to alkaline hot springs;	Optimum 55 ºC; growth up to 70 ºC;
	Chloroflexus aggregans	Japanese hot spring;	Optimum 55 ºC; growth up to at least 60 ºC;
	Thermochromatium tepidum	Neutral hot springs containing sulfide;	Optimum 49 ºC; growth up to 56 ºC;
	Heliobacterium modesticaldum	Neutral to alkaline hot springs;	Optimum 52 ºC; growth up to 57 ºC;
Halophiles	*Halochromatium salexigens*	Saltern Salin de Giraud, France	Optimum 8-11% NaCl; growth up to 20%;
	Halochromatium glycolicum	Solar Lake, Sinai	Optimum 4-6% NaCl; growth up to 20%;
	Marichromatium purpuratum	Marine sponges; seawater;	Optimum 5% NaCl; growth up to 7%;
	Halohodospira halophila	Summer Lake; Oregon;	Optimum 15% NaCl; growth up to 30%;
Alkaliphiles	*Ectothiorhodospira haloalkaliphila*	Soda lakes, Wadi Natroun, Egypt	Optimum pH9; growth up to pH10.5;
	Halorhodospira abdelmalekii	Soda lakes, Wadi Natroun, Egypt	Optimum pH8.5-9; growth up to pH10.5;
	Heliorestis daurensis	Lake Barun Torey, Siberia, Russia	Optimum pH9; growth up to pH10.5;
	Heliorestis baculatum	Lake Ostozhe, Siberia, Russia	Optimum pH9; growth up to pH10.5;
Acidophiles	*Rhodoblastus acidophila*	Acidic marshwaters, soils	Optimum pH 5.8; growth down to pH4.8;
	Rhodophila globiformis	Acidic warm sulfur springs,Yellowstone National Park	Optimum pH5; growth down to pH4.2;
Psychrophiles	*Rhodoferax antarcticus*	Microbial mats and lakewater, McMurdo Dry Valleys, Antarctica	Optimum growth temperature, 18ºC; growth down to 0 ºC; no growth at 25 ºC;

Rhodoferax antarcticus

Habitat and Morphology

Rhodoferax antarcticus is a purple nonsulfur bacterium (Madigan, Jung, Woese, & Achenbach, 2000). The strain used in my study, *R. antarcticus* ANT.BR was first isolated from an algal-bacterial mat sample collected by Dr. R. W. Castenholz from ponds in and around Cape Royds, Ross island (77-78°S) (Vincent, Downes, Castenholz, & Howard-Williams, 1993). They collected the sample in mid-austral summer when the maximum water temperature for some of the ponds was at 8 °C (Vincent, Downes, et al., 1993). This organism was first fully characterized by Madigan (2000) who designated it as strain ANT.BR and proposed to recognize strain ANT.BR as a novel species, *Rhodoferax antarcticus* sp. nov., named for its known habitat, the Antarctic (Madigan, et al., 2000). The cells of strain ANT.BR grown phototrophically are highly motile and exist as small spirals and curved rods, 0.7×2-3μm (Figure 1 (A)). The scanning electron micrograph (SEM) shows that one or more polar flagella are attached to one end of the cells (Figure 1 (B)). The thin sections of cells indicate a gram-negative bacterial cell wall. Although prokaryotes do not contain complex membranous organelles such as mitochondria, chloroplast, or Golgi apparatus, for some prokaryotes, especially photosynthetic bacteria (like cyanobacteria, and purple bacteria), internal membranous structures exist to accomplish various cellular functions. For instance, a plasma membrane is invaginated to form an internal

light-harvesting membrane in photosynthetic bacteria and nitrifying bacteria with very high respiratory activity (Willey et al., 2008). However, even when the cells were grown at very low light intensities, no obvious intracytoplasmic photosynthetic membranes were observed (Figure 1(C)), which makes this species very interesting. And as a purple bacterium, mass cultures of phototrophically grown cells were peach-colored (Madigan, et al., 2000).

G+C Content and Phylogeny

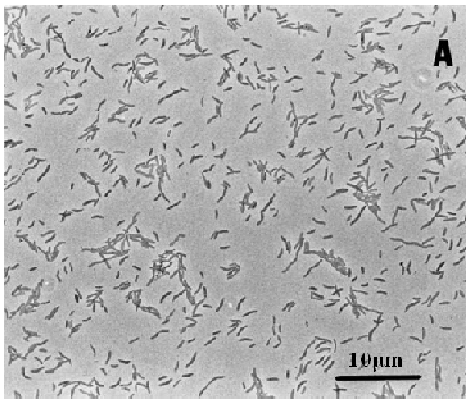
The G+C content of strain ANT.BR was estimated to be 61.5mol% by thermal denaturation. And 16S rRNA gene analysis showed strain ANT.BR belonged to beta-proteobacteria (Jung, Achenbach, Karr, Takaichi, & Madigan, 2004; Madigan, et al., 2000). Phylogenetic reconstruction indicated very close relationships between strain ANT.BR and another purple nonsulfur bacterium *Rhodoferax fermentans* FR2^T (Hiraishi, Hoshino, & Satoh, 1991) with only 1.5% distance difference (Figure 2) and between strain ANT.BR and strain *Rhodoferax antarcticus* Fryx1(Jung, et al., 2004) with >99.8% identity in rRNA gene sequence (Figure 3). Of note, strain ANT.BR is also close to a chemoorganotroph *Polaromonas vacuolata*, a psychrophilic Antarctic sea-ice bacterium with a maximum growth temperature of 12℃ (Irgens, Gosink, & Staley, 1996) as well as Arctic sea ice bacterium ARK10281, a beta-proteobacterium isolated from the

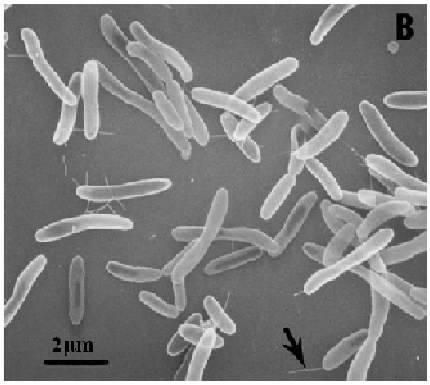
melt pools on Arctic pack ice floes (Brinkmeyer, Glockner, Helmke, & Amann, 2004).

Pigment Analysis

Despite no obvious intracytoplasmic membranes in the cells of strain ANT.BR even at very low light intensities, the existence of carotenoids and bacteriochlorophyll a in the phototrophic cells was very clear. Major peaks at 799nm, 819nm and 866nm in the spectrum of intact cells and 770nm of methanol extracts suggested the presence of bacteriochlorophyll a ((Ke, 2001),Figure 4).

Both the color of mass cultures of phototrophically grown strain ANT.BR as mentioned above and the absorption spectra suggested that carotenoids spheroidenone or spheroidene and spirilloxanthin, or derivatives of these pigments predominate in this organism (Jung, et al., 2004; Madigan, et al., 2000). Composition analyses of carotenoids of several *Rhodoferax* species further confirmed this hypothesis (Table 2). The hydroxy-spheroidene (OH-spheroidene) level in strain ANT.BR was fairly high. That agreed with the preliminary carotenoid analyses result by Dr. Shinichi Takaichi who reported that OH-spheroidene is a major carotenoid of this strain (Madigan, et al., 2000). Another supporting point is when the phototrophic culture of strain ANT.BR was under exposure to air, its color changed from peach to rose-pink, which is a common observation in species of purple bacteria using spheroidene pathway to produce carotenoids (Takaichi, 1999).





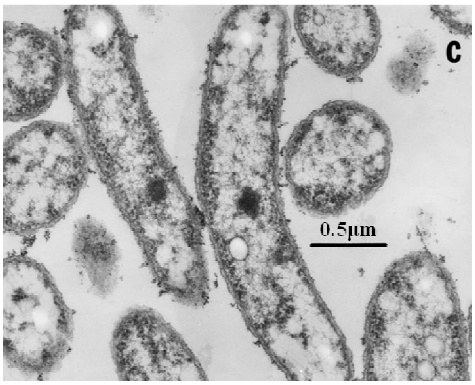


Figure 1 Morphology of cells of *R. antarcticus* ANT.BR. (A) Phase-contrast photomicrograph; (B) Scanning electron micrograph; (C) Transmission electron micrograph (Madigan, et al., 2000).

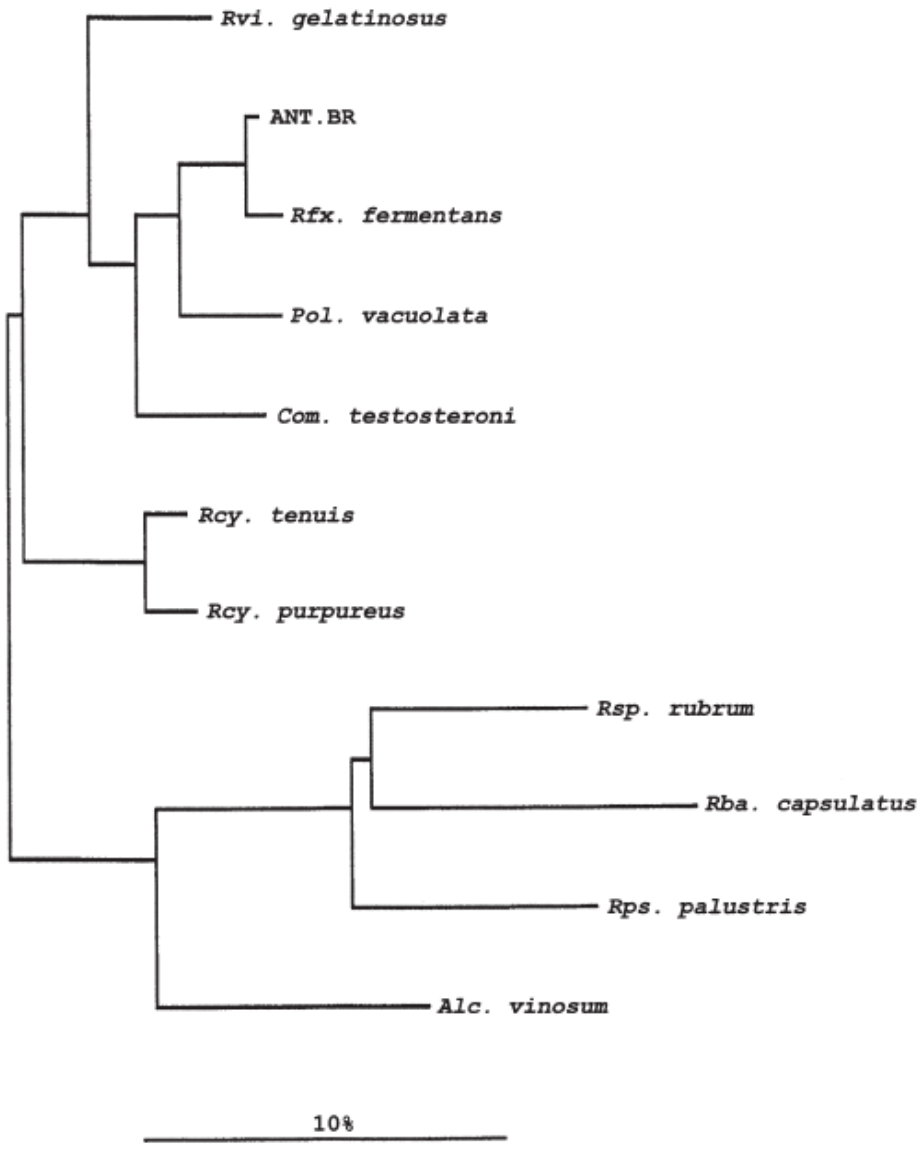


Figure 2 Evolutionary distance tree of *R. antarcticus* ANT.BR and related bacteria based on comparisons of 16S rRNA gene sequences (Madigan, et al., 2000).

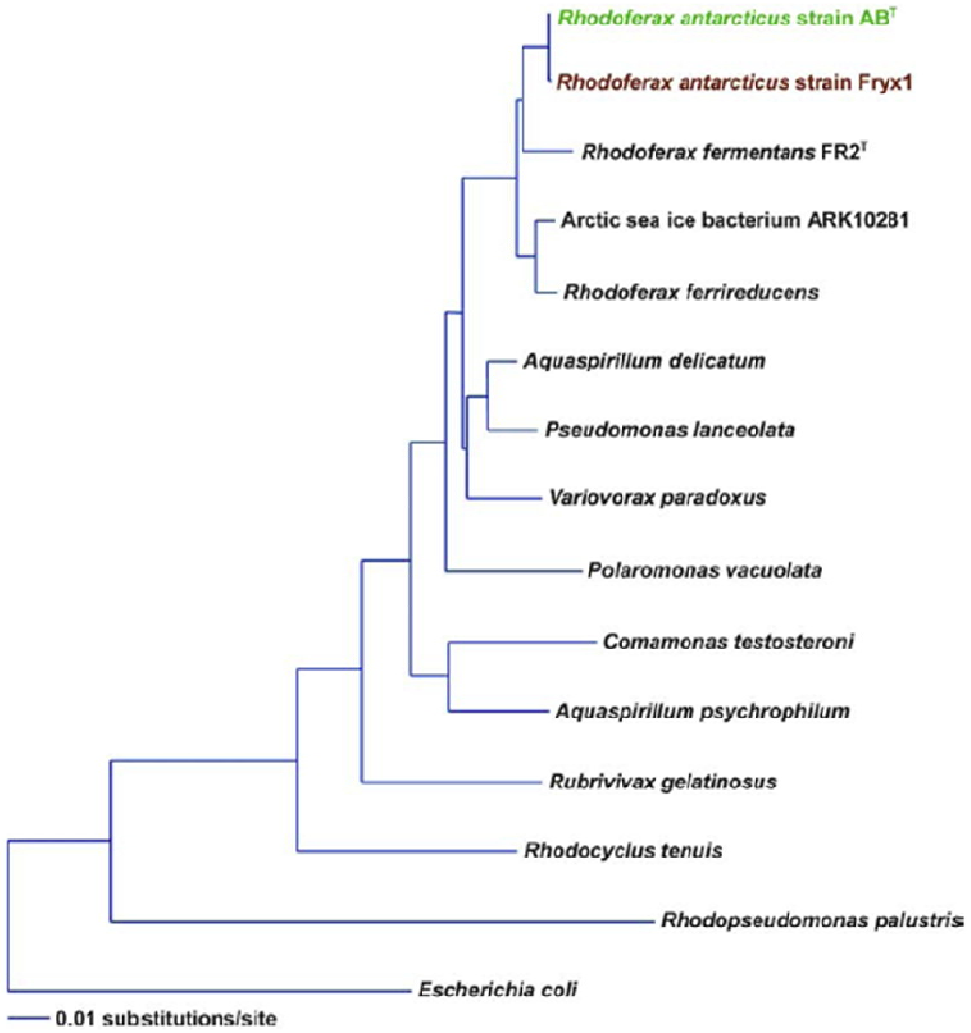


Figure 3 Phylogenetic tree of *R. antarcticus* ANT.BR and other related bacteria based on comparisons of 16S rRNA gene sequences (Jung, et al., 2004).

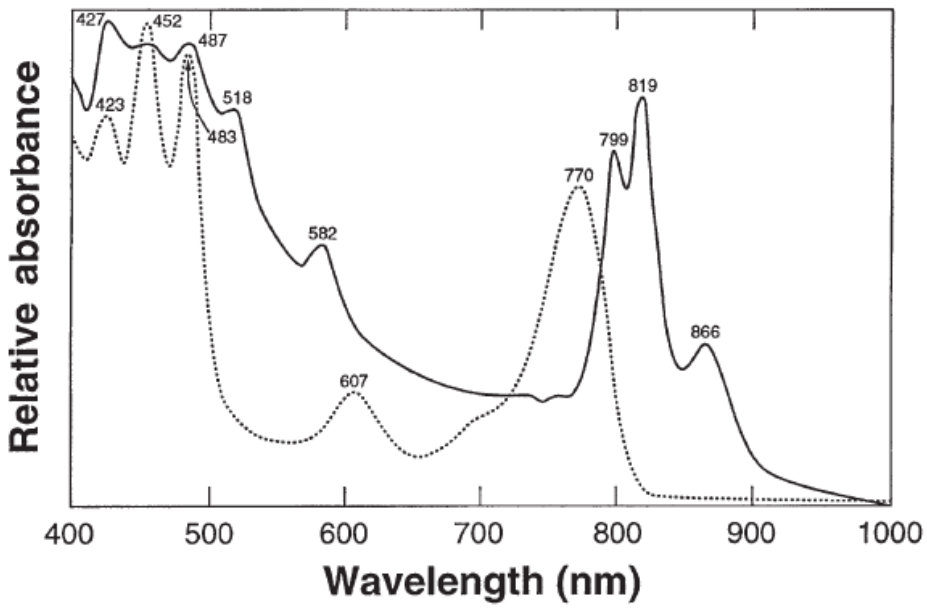


Figure 4 Absorption spectra of *Rhodoferax antarcticus* ANT.BR. (smooth line: intact cells; dotted line: methanol extracts of cells) (Madigan, et al., 2000).

Table 2

Composition (mol% of total) of carotenoids in phototrophic cells of strain ANT.BR. (ND: not detected; SE: spheroidene; SO: spheroidenone) (Jung, et al., 2004).

Carotenoid	***Rfx. antarcticus* strain ANT.BR**
Spheroidene	8
OH-Spheroidene	72
Methoxy-spheroidene	2
Spheroidenone	1
OH-Spheroidenone	15
Methoxy-spheroidenone	ND
Spirilloxanthin	2
Keto-spirilloxanthin	<1
Diketo-spirilloxanthin	<1
(OH-SE + OH-SO)/(SE + SO)	9
(SO + OH-SO)/(SE + OH-SE)	0.19

Physiology

Since strain ANT.BR was isolated from a microbial mat in Antarctica (Vincent, Castenholz, Downes, & Howard-Williams, 1993), the effect of temperature on its phototrophic growth was of great interest. The best growth of strain ANT.BR was achieved between 12ºC and 18ºC (Figure 5) with a broad growing temperature range of 0–25ºC. ANT.BR cannot grow above 25ºC indicating a significant degree of cold-adaptation (Madigan, et al., 2000).

Strain ANT.BR can grow both chemoorganotrophically (dark/oxic) and phototrophically (light/anoxic) and When growing photoheterotrophically, strain ANT.BR can use many compounds including the fatty acid acetate, organic acids and some sugars (Table 3). Grown on N_2 at 5°C, cells can reduce acetylene to ethylene at temperatures as low as 2°C, the lowest temperature tested, suggesting the possible presence of a nitrogenase system in this phototroph (Madigan, et al., 2000).

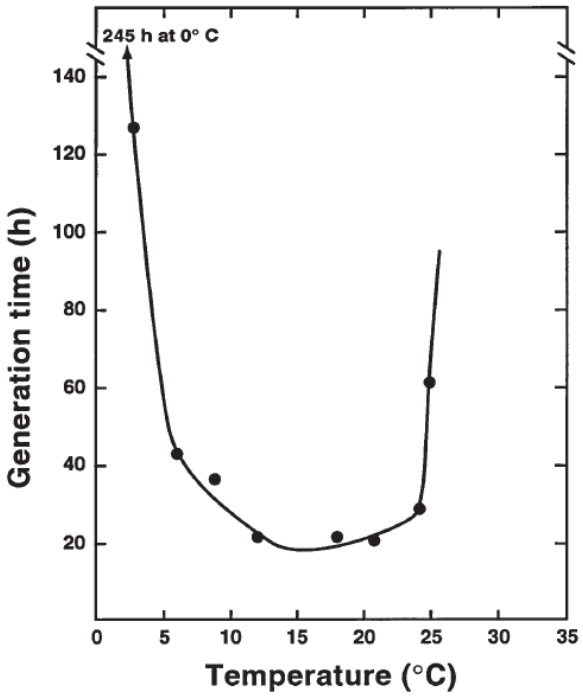


Figure 5 Growth of strain ANT.BR as a function of temperature. Cells were grown photoheterotrophically in medium AB containing 0.1% yeast extract (Madigan, et al., 2000).

Table 3

Summary of major properties of phototrophically grown cells of strain ANT.BR (Madigan, et al., 2000).

Property	ANT.BR
Morphology	Cells vibrio- to spirillum-shaped, 0.7×2–3 μm
Absorption maxima (nm) in vivo	866, 819, 799, 582, 518, 487, 452, 427
Carbon sources utilized[a]	Acetate, pyruvate, lactate, succinate, malate, fumarate, glucose, fructose (2); sucrose, citrate, aspartate (1)
Vitamins required	Biotin
Optimum temperature (°C)	12–18
DNA G+C content (mol%)	61.5 (T_m)
Habitat	Antarctic microbial mats

Genome Sequencing Rationale

Recent study suggests that oxygenic photosynthesis was preceded by anoxygenic photosynthesis by approximately a billion years (Blankenship, 2001). Because of that, anoxygenic phototrophs have long been an excellent target for analysis and speculation of the origin and early evolution of photosynthesis. Certain current extreme environments exhibit remarkable similarities to the habitats present on the early Earth (Madigan & Marrs, 1997), so extremophilic anoxygenic phototrophic bacteria are of particular interest to research investigators. Furthermore, according to the recent molecular evolutionary analysis of the chlorophyll biosynthetic pathway, the most ancient known photosystem is thought to exist in the anoxygenic purple photosynthetic bacteria (Blankenship, 2001). From what has been discussed above, there is little doubt that as a member of extremophilic anoxygenic purple phototrophic bacteria, *R. antarcticus* makes a great model for photosynthesis analysis.

R. antarcticus is the first anoxygenic phototrophic β-proteobacterium that has been chosen for complete genome sequencing. Therefore, its complete genome sequence will without question expand the available genomic database for photosynthetic prokaryotes and thus assist us to better understand this group of bacteria at the genetic level.

Furthermore, as described above, psychrophiles have only recently been investigated and remain one of the least explored, studied and understood

microbes, in part because of the difficulties associated with sample collection. Strain *R. antarcticus* ANT.BR is the first described example of an anoxygenic phototrophic bacterium that is known to have cold-adaptation ability. In fact, it was the only known example until 2003 (Madigan, 2003), after which a few other psychrophilic anoxygenic phototrophic bacteria were discovered. For instance, an additional psychrophilic anoxygenic phototrophic bacterium is strain *Rhodoferax antarcticus* Fryx1 that was isolated from the permanently ice-covered Lake Fryxell, Antarctica. Its optimal growth temperature is between 15°C and 18°C (Jung, et al., 2004). Therefore, we predict that complete genome sequences can allow us to unravel the underlying mechanisms for cold-adaptation. In particular, what are the cold-active photosynthetic proteins, how do cold-active photosynthetic proteins work and what kind of special metabolic pathways are involved in the cold-adaption process.

Another interesting point, as mentioned before, is that no obvious intracytoplasmic photosynthetic membranes are observed in *R. antarcticus*, even in cells grown at very low light intensities (Madigan, et al., 2000). From this perspective, strain ANT.BR could be an interesting model phototroph to study how pigment-protein complexes carry out their functions to accomplish optimal photosynthesis at low temperatures.

In addition, properties of the cold-active nitrogenase of strain ANT.BR may be of interest to biochemists because of the catalytic constraints that low

temperatures are known to place on nitrogenases from mesophilic diazotrophs (Miller & Eady, 1988).

In summary, the completion of the genome sequence of *R. antarcticus* and the subsequent silico analysis based on its complete genetic information will deepen our understanding of this purple nonsulfur bacterium with regard to photosynthesis evolution, cold adaptation ability, function of pigment-complexes, and nitrogenase properties. It will also provide an essential framework for future wet-bench experiments into these important processes.

Chapter 2

MATERIALS AND METHODS

Genomic DNA and Cells

The genomic DNA and cells of strain *R. antarcticus* ANT.BR were kindly provided by Dr. Michael Madigan, Southern Illinois University, USA. Details about the growing conditions of this strain can be found in the article (Madigan, et al., 2000). The strain has been deposited in the American Type Culture Collection as ATCC 700587 and its ribosomal RNA gene sequence accession number in GenBank is AF084947(Madigan, et al., 2000).

Roche 454 Life Science's Pyrosequencing

The genomic DNA of strain *R. antarcticus* ANT.BR was sequenced using a pyrosequencing approach on a Roche/454 Genome Sequencer FLX instrument (http://uagc.arl.arizona.edu/index.php/next-gen-sequencing-services/roche454.html). A total of 380,000 reads were generated that resulted in an average sequence coverage of 14-fold, which means that each base pair of the genome is sampled 14 times on average. The high level of coverage is able to maximize the size of assembled "contigs", increase the sequence quality, and simplify the effort devoted to the genome finishing phase.

Illumina Genome Analyzer Sequencing

The genomic DNA of strain ANT.BR was also sequenced on an Illumina Genome Analyzer IIx system (Center for Personalized Diagnostics in the Biodesign Institute at Arizona State University). A 300-bp library was constructed using the Illumina Paired-End Sample Prep Kit (P/N 1001809) following the manufacturer's instructions and the library was loaded onto two (out of eight) lanes of an Illumina flow-cell. A total of 32,561,890 high-quality ~36-bp reads were generated that consisted of 1.17 billion base pairs. This significant over-sampling represents roughly 292-fold sequence coverage of the *R. antarcticus* genome. Of note, this level of coverage was not targeted, but rather represents a proof-of-principle experiment on a newly installed Illumina instrument at Arizona State University (sequencing was free-of-charge).

Fosmid Library Construction

The EpiFOS^{TM} Fosmid Library Production Kit (Epicentre) was used to construct a fosmid library of strain *Rhodoferax antarcticus* sp. ANT.BR following the manufacturer's instructions. Briefly, a 1% agarose gel was used to check the general integrity of the genomic DNA sample. A majority of the genomic DNA migrated with a 42-kb Fosmid Control DNA also loaded on the gel, indicating most of the genomic DNA was approximately 40-kb fragments; thus, no further shearing of the genomic DNA was needed. Then, the ends of genomic DNA fragments were repaired to form 5'-phosphorylated blunt ends and then purified

by ethanol precipitation. Blunt-ended genomic DNA was ligated to the Cloning-Ready CopyControl pCC2FOS Vector at room temperature for 4 hours and then transferred to 70℃ for 10 minutes to inactivate the Fast-Link DNA Ligase. The ligated genomic DNA was then packaged into lambda phage, EPI300-T1R *E. coli* were infected, and infected EPI300-T1R cells were spread on LB plates + 12.5 μg/ml chloramphenicol, and incubated overnight at 37℃ to select for the CopyControl Fosmid clones. Individual CopyControl Fosmid clones were transferred to 96-well plates (each well contains 1 ml LB medium + 2 μl 500× CopyControl Fosmid Autoinduction Solution + 12.5 μg/ml chloramphenicol) and grown overnight at 37℃ with shaking at 250 rpm. Fosmid DNA was purified for sequencing. Twenty 96-well plates were sent out to the Arizona Genomics Center at the University of Arizona using their existing high-throughput pipeline for Sanger sequencing (Sequencing service information is available at http://www2.genome.arizona.edu/welcome). Each Fosmid DNA template was sequenced from both ends of the insert using dye-terminator chemistry and pCC2^{TM} sequencing primers (forward and reverse).

OpGen Optical Mapping

Cells of strain *Rhodoferax antarcticus* ANT.BR was sent to OpGen, Inc, which has developed a unique and powerful optical mapping technology for comparative genomics, strain typing, and whole genome sequence assembly. The

detailed process of how OpGen generates an in silico whole genome Optical Map and how a restriction enzyme is chosen can be found at http://www.opgen.com/.

Gap Closure

The sequence data from 454 pyrosequencing, the Illumina Genome Analyzer, along with the contig ordering and orientation information provided by fosmid paired-end sequences and OpGen optical map, were used in the genome finishing stage of the *R. antarcticus* genome project. Sequence reads were manually checked in the Consed software to find contigs that can be joined together. For those contigs which cannot be joined together, PCR implementation and mini-shotgun library construction were employed for the gap closure. For the gaps less than about 6-kb, PCR primers were designed and the corresponding fosmid DNA clone was used as the PCR template to generate PCR products for sequencing and further gap closure. For gaps greater than that, "mini" shotgun libraries of the corresponding fosmid DNA were constructed and sequenced entirely by Sanger sequencing. As for PCR, Platinum® *Taq* DNA Polymerase, 10×PCR buffer (-$MgCl_2$), 50mM $MgCl_2$, 2.5mM dNTPs were all from invitrogen™. PCR primers were ordered from Integrated DNA Technologies (IDT). With regards to the PCR steps, the initialization step was 94 ºC for 5 min, and then 30 cycles of 94 ºC for 5 seconds, 55 ºC (or sometimes 52 ºC) for 20 seconds, 70 ºC for 2 min, and last is the elongation step which was 72 ºC for 7 min.

Chapter 3

RESULTS AND DISCUSSION

Results

In order to determine the complete genome sequence of strain *Rhodoferax antarcticus* sp. ANT.BR, two major next-generation sequencing strategies were applied in this study, Roche/454 pyrosequencing and Illumina Genome Analyzer sequencing. Additionally, two auxiliary sequencing strategies, fosmid paired-end sequences and OpGen optical map, were used to provide the ordering and orientation details of the sequence assemblies.

454 Pyrosequencing Data

More than 370,000 random sequence reads were collected across the *R. antarcticus* genome by 454 pyrosequencing followed by assembly using the program Newbler to generate non-redundant consensus contigs. A summary of the basic statistics about raw 454 sequencing reads and assembled contigs are shown in Table 4. The publicly available finishing package Consed (Gordon et al., 1998) was used to further analyze the draft assemblies for low quality regions and to join contigs by helping to close gaps remaining in the assembly. Figure 6 shows an example of what one contig looks like in the Consed software. To resolve gaps in the assembly, paired primers were designed at two adjacent

contigs and then the Polymerase Chain Reaction (PCR) was performed to generate a putative PCR product for sequencing to "close" the gap.

Table 4

Summary of basic statistics about raw 454 sequencing reads and assembled contigs for strain ANT.BR.

Read Status	Num of Aligned Reads	372896
	Num of Aligned Bases	135799035
	Number Assembled	370419
Large Contig Metrics	Number of Contigs	156
	Number of Bases	4082109
	Average Contig Size	26167
	N50 Contig Size	74281
	Largest Contig Size	211345

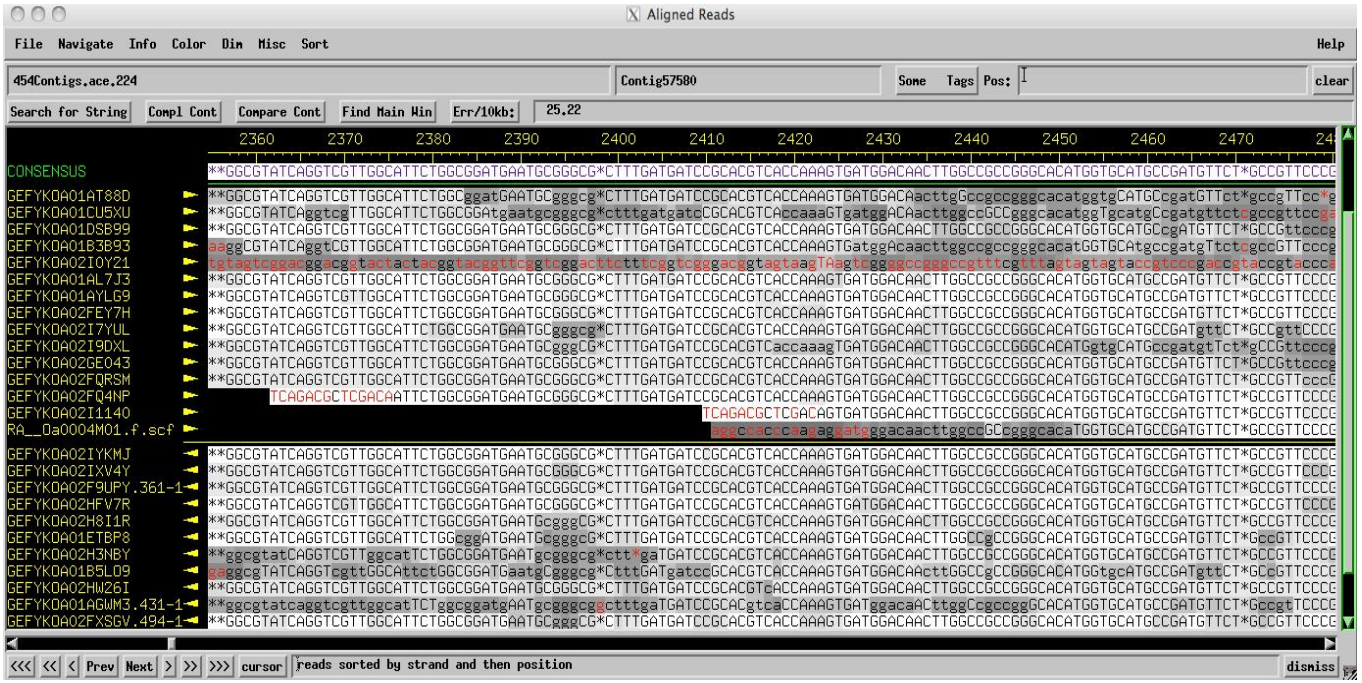


Figure 6 One of the two major windows in Consed for displaying the supporting sequence reads for each contig.

Illumina Genome Analyzer Reads

With regards to Illumina Solexa sequencing, more than 37 million 36-bp Illumina Genome Analyzer reads were generated and then assembled by the publicly available assembler, Velvet (Zerbino and Birney, 2008). The draft assemblies were subsequently combined with the Roche/454 assembly and analyzed further with Consed for genome finishing. The summary of the Illumina Genome Analyzer reads are shown in Table 5.

Fosmid Paired-end Sequences

The insert ends of large (~40-kb) fosmid clones were sequenced using a traditional Sanger approach to provide long range positional information (or scaffolding) of sequence contigs generated by the previous high-throughput approaches. A total of 1,920 random fosmid subclones were isolated for end-sequencing. These sequences represent an approximate coverage of the genome of 0.5-fold and were performed at the Arizona Genomics Institute at the University of Arizona. Sequence reads were processed using Phred, which is able to read the DNA sequencing trace files, calls bases, and assign a quality value to each called base (Ewing & Green, 1998; Ewing, Hillier, Wendl, & Green, 1998) and then the sequence reads were incorporated into the Consed program to help order and orientate a majority of sequence contigs and link these contigs into scaffolds so as to facilitate the finishing process. Figure 7 illustrates how paired-end fosmid DNA

reads serve to organize the contigs in the Consed software and provide the orientation in the Assembly View window. Notice the purple connecting lines that effectively orient adjacent contigs—these lines represent the association of two end-read pairs generated from a single fosmid clone.

Table 5

Summary of basic statistics about Illumina sequencing reads for strain ANT.BR.

	Raw Reads	**HQ Reads**	**Raw Bases**	**HQ Bases**
Ra1	14,945,402	11,990,362	538,034,472	431,653,032
Ra2	22,465,210	20,571,528	808,747,560	740,575,008

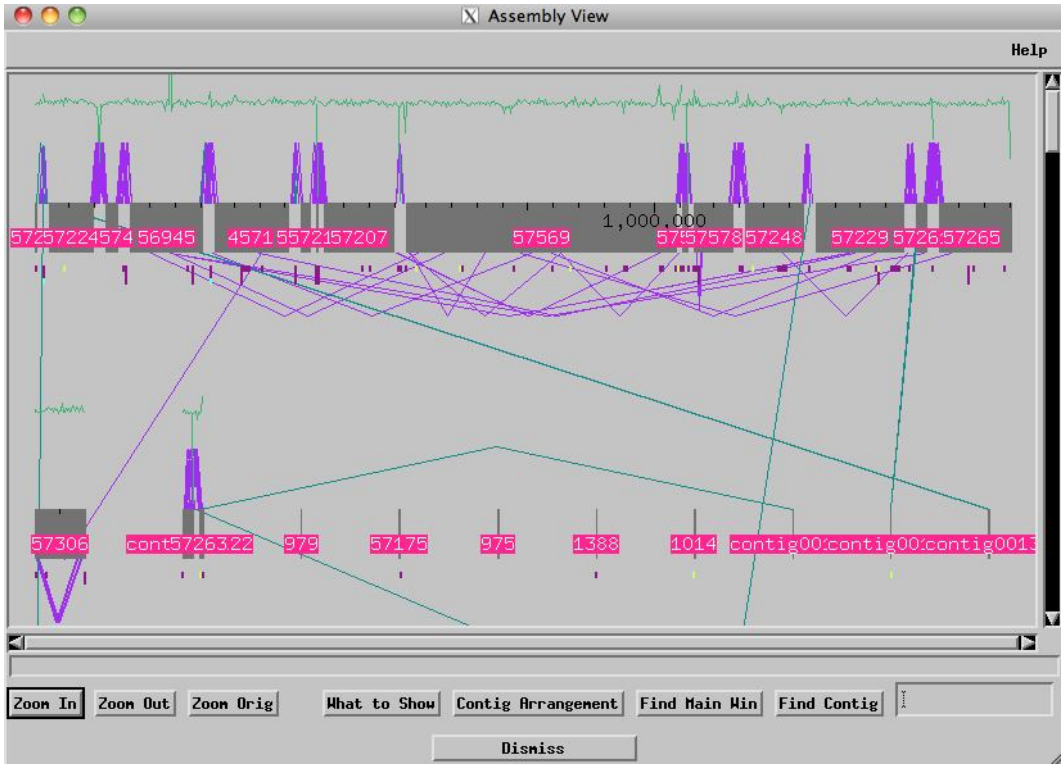
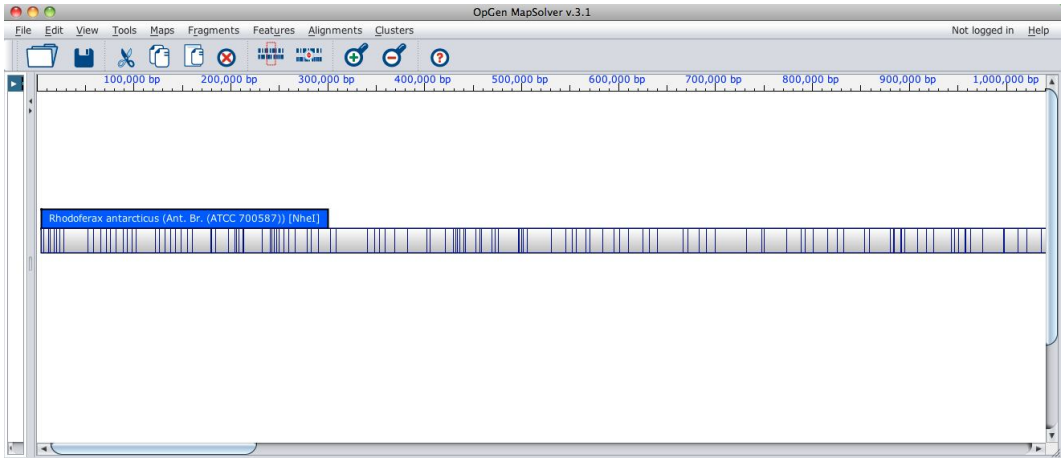


Figure 7 The other major window in Consed: Assembly View. It shows the ordering and orientation of most of the contigs organized by fosmid paired-end sequences.

OpGen Optical Mapping

A high-resolution whole genome restriction map of strain *R. antarcticus* sp. ANT.BR was generated to assist in the final assembly of the project (OpGen, Inc.). In regards to the restriction enzyme used in constructing the optical map of strain ANT.BR, NheI was applied. Figure 8 (A) shows the final restriction map. Once sequence contig data from the Consed program was loaded into the appropriate software (MapSolver), a majority of sequence contigs were aligned to the restriction map according to the restriction fragment patterns. Then, contig orientation, gap size, and gap location can be visualized and determined. Figure 8 (B) shows the alignment between the optical restriction map and the assembled sequence contigs.

(A)



(B)

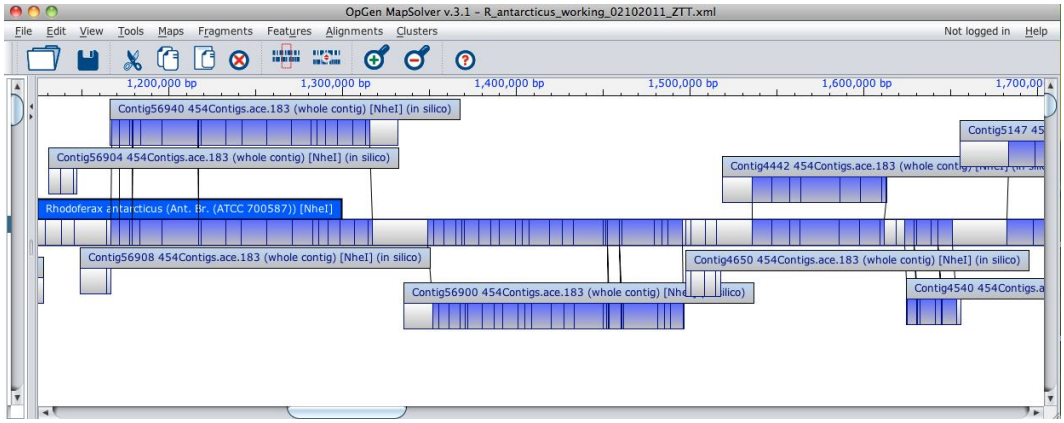


Figure 8 Optical restriction map in the MapSolver software for strain ANT.BR. Nhel restriction enzyme was used in the DNA digestion pretreatment. (A) Optical restriction map before contigs are loaded into MapSolver; (B) Optical restriction map after sequence

contigs are loaded into MapSolver. Contigs are aligned to the map according to the restriction fragment pattern.

Discussion

General Features of Draft Genome

The draft genome sequence of *R. antarcticus* sp. ANT.BR was annotated for gene features (e.g., open reading frames or ORFs) using an automated pipeline at the Institute for Genomics Sciences in the School of Medicine at the University of Maryland. Annotation Engine was first used to find the putative open reading frames by the Glimmer system and then assign functions to hypothetical proteins by search each hypothetical protein against the database like GenBank, SWISS-PROT, and PIR and JCVI's CMR. And then Manatee, a web-based genome annotation tool, was used for the manual examination and correction (if necessary) of the automated annotation results. Thus far, a chromosome and a plasmid (pMTM1, named after Dr. Michael T. Madigan who first characterized strain ANT.BR) have been annotated. More properties about this draft genome data are summarized in Table 6.

The G+C content of the chromosome (59.1%) is much higher than that of the plasmid pMTM1 (48.4%). Taking a close look at the GC profile of the chromosome, it is clear that G+C poor (A+T rich) and G+C rich regions are present alternatively as a mosaic structure. It should also be noted that a region from point 1 to point 2 (Figure 9) stands out in the GC profile, representing a

region of low G+C content. This obvious A+T rich island is highly likely due to the presence of bacteriophage lysogens or other inserted elements. Indeed, current annotation results regarding this high A+T rich region indicate a large number of bacteriophage related genes, such as the genes coding for prophage integrase proteins, phage capsid family protein, and phage terminase (data not shown). In fact, a few other bacteriophage related genes are also found scatteredly within the region from 35kb to 500kb where the G+C content is also low (Figure 9).

There are a total of 4,262 putative ORFs in the current genome sequence, 4,036 on the chromosome and 226 on the plasmid (Table 6). Sixty-four tRNA genes and 3 rRNA genes are also identified. More details are summarized in Table 7. Among 4,262 putative ORFs identified, about 75% of those genes begin with the ATG start codon, 14% with GTG and 11% with TTG (Table 7).

Table 6

Basic features of *R. antarcticus* draft genome.

	Chromosome				**Plasmid**			
Molecular length	3,809,266 bp				198,615 bp			
GC content	59.1%				48.4%			
Base frequencies	A	C	G	T	A	C	G	T
	20.5%	29.5%	29.6%	20.4%	26.2%	25.5%	22.9%	25.3%
ORF count	4036				226			
Average gene length	848 nt				621 nt			
Percent coding	89.9%				70.7%			

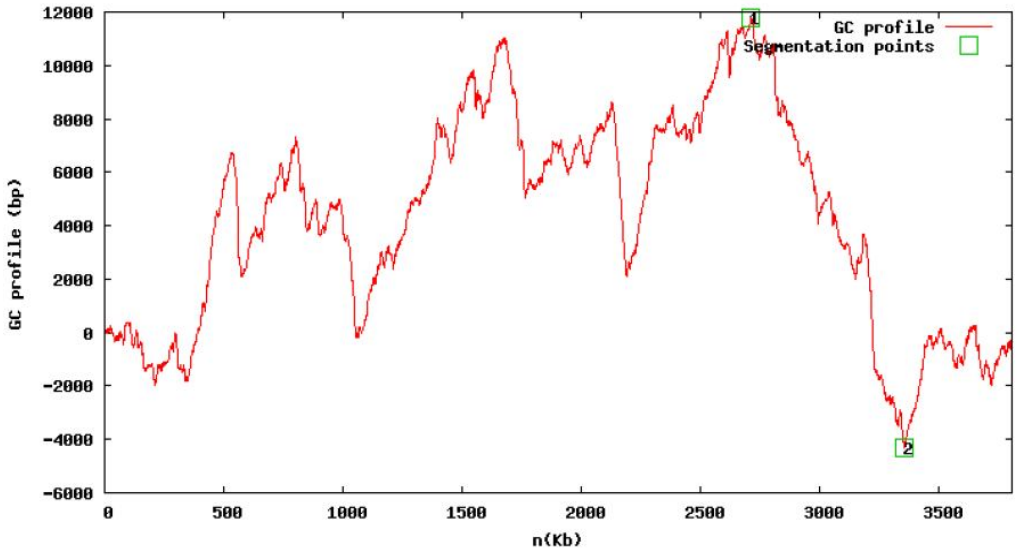


Figure 9 GC profile for the chromosome of strain ANT.BR. The negative cumulative GC profile for the chromosome of strain ANT.BR marked with the segmentation points obtained. The numbers 1 and 2 indicate the beginning and end of a A+T rich region.

Table 7

Summary of types of start codons for the 4262 ORFs.

Start codon	Number	Percentage
ATG	3189	74.8%
	(2774)	(78.5%)
GTG	603	14.1%
	(435)	(12.3%)
TTG	465	10.9%
	(322)	(9.1%)
Other	5	0.1%
	(5)	(0.1%)

Note: numbers in parentheses do not include hypothetical proteins

Nitrogen Fixation System

Rhodoferax antarcticus sp. ANT.BR is capable of nitrogen fixation (Madigan 2000) and what types of nitrogenase systems are used in this process under low temperatures is of particular interest. Possibilities include a molybdenum nitrogenase, vanadium nitrogenase, iron-only nitrogenase, or a certain combination of these nitrogenase systems. A previous study compared the effect of temperature on dinitrogen reduction by purified vanadium nitrogenase and molybdenum nitrogenase of *Azotobacter chroococcum* and the results indicated that as the assay temperature decreased from 30ºC to 5ºC, N_2 remained an effective substrate for vanadium nitrogenase, but not molybdenum nitrogenase. More specifically, the activity cross-reactions between the components of two types of nitrogenases showed the enhanced temperature activity to be associated with the Fe protein of vanadium nitrogenase, implying that low temperature favors nitrogen fixation by vanadium nitrogenase and that Fe protein might be crucial in nitrogen fixation at different temperatures (Miller & Eady, 1988). Therefore, as a psychrophile, what is special about the nitrogenase system working in strain ANT.BR? Does the bacterium have a vanadium nitrogenase system or another system to adapt to the habitats of low temperature?

According to our annotation results, only a molybdenum nitrogenase system is found in the strain ANT.BR. Genes for molybdenum nitrogenase

complex, molybdenum-iron (MoFe) protein alpha chain (NifD), MoFe protein beta chain (NifK), and nitrogenase iron protein (NifH), and genes for some other proteins involved in nitrogen fixation like transcription regulator (NifA) and MoFe cofactor biosynthesis protein (NifE) have been identified in this study thus far. Further study into the gene sequences for molybdenum nitrogenase complex and its structure is needed to help us elucidate the mechanism of nitrogen fixation process of strain ANT.BR under low temperature.

Photosynthesis Gene Cluster

The location and properties of the photosynthetic units in *R. antarcticus* sp. ANT.BR is of particular interest because of its low temperature habitat and the lack of obvious intracytoplasmic membranes, even under low light intensity. This observation is in contrast to the general phenomena for purple photosynthetic bacteria that light-reactions take place in and on highly invaginated intracytoplasmic membranes. The following discussion focuses on what the genome sequence along with the previous experimental results (Madigan, et al., 2000) tell us about the photosynthetic units in cells of *R. antarcticus* sp. ANT.BR.

The photosynthetic unit in bacteria consists of two parts, the photochemical reaction center where light-energy can be used to drive a transmembrane charge separation (Deisenhofer, Epp, Miki, Huber, & Michel, 1985; Feher, Allen, Okamura, & Rees, 1989), and the light–harvesting (LH, or antenna) complexes

where the light is absorbed and then transferred to the reaction center. Typically, the reaction center consists of L, M, H, and C subunits. Most purple photosynthetic bacteria contain two types of LH complexes. The first is the LH I complex (B875 type; the number refers to the near infrared absorption maxima), which is also called the core antenna complex because of its close association with the reaction center. The LH I complex exits in all wild-type purple bacteria (R. J. Cogdell et al., 1999). The second type is the LH II (B800/850 type) complex, which is also referred to as the peripheral antenna complex. Each of these two types of light-harvesting complexes consists of two types of polypeptides designated as α and β.

In *R. antarcticus*, the large photosynthesis gene cluster identified by the genome annotation (Figure 10) reveals the genes coding for L subunit (PufL), M subunit (PufM), c-type cytochrome subunit (PufC), and H subunit (PufH, or PuhA because of its gene position in *puh* operon).

The hydropathy plots, generated by the ProtScale program (original principle described by (Kyte & Doolittle, 1982)) at the ExPASy Proteomics Server (http://expasy.org/cgi-bin/protscale.pl), show the relative hydrophilicity and hydrophobility of the amino acids of the LMH subunits. The plots for both L and M subunits (Figure 11) indicate five putative hydrophobic segments, each of which represents a putative membrane-spanning α-helix. The hydropathy plot for the H subunit shows only one hydrophobic region (Figure 11), indicating the

presence of a single α-helix. All these features agree well with the typical hydrophobility pattern of LMH subunits in other bacteria. In this regard, like other purple photosynthetic bacteria, the photosynthetic units in *R. antarcticus* sp. ANT.BR are highly likely to be in or on certain well-developed cell membrane systems for photosynthesis.

As to why no obvious intracytoplasmic membranes (ICM) were observed in *R. antarcticus*, two possible explanations are proposed here. First, the transmission electron micrograph (TEM) showing the morphology of *R. antarcticus* (Madigan, et al., 2000) has some artifacts that cast doubt on the sample preparation process (Dr. Robert Roberson, Arizona State University, personal communication,). Inappropriate sample fixation and preparation procedure for TEM would lead to membrane disruptions which might explain the absence of obvious ICM in phototrophically grown *R. antarcticus*. Therefore, *R. antarcticus* might have a well-developed regular ICM, but inappropriate sample preparation process might have disrupted it. If that is true, further experiments like a modern form of TEM, called cryo-electron microscopy, might be helpful to reveal the photosynthetic membrane system in *R. antarcticus* because cryo-electron microscopy can make the sample preparation process very fast and at very low temperature to minimize the artifacts as well as membrane disruptions. The second possible scenario is that *R. antarcticus* has a different membrane system to harbor photosynthetic units from the regular ICM. The ICM for most

photosynthetic purple bacteria is arranged in four patterns, the vesicle type (like *Rhodospirillum rubrum*), the tubule type (like *Thiocapsa pfennigii*), the thylakoid-like membrane in regular stacks (like *Rhodospirillum molischianum*), and thylakoid-like membrane, partially stacked and irregular arranged (like *Rhodopseudomonas palustris*) (Drews & Golecki, 1995). But there are still a small group of facultative purple bacteria that contain poorly developed intracytoplasmic membranes, such as *Rhodocyclus* (*Rc.*) *purpureus, Rc. tenuis,* and *Rubrivivax gelatinosus* and others (Drews & Golecki, 1995). For example, freeze-fracture micrographs of *Rc. Tenuis* show that the cell only contains small tubular intrusions of the cytoplasmic membrane to compensate for the lack of intracytoplasmic membrane. Compared with chemotrophic cells, it contains a higher diameter and density of intramembrane particles in the cytoplasmic membrane of phototrophic cells (Drews & Golecki, 1995). *R. antarcticus* might have a similar membrane system for photosynthesis as *Rc. Tenuis*. Freeze-fracture microscopy might be helpful to elucidate this hypothesis.

4

Figure 10 Photosynthetic gene cluster in *R. antarcticus* ANT.BR. Genes are presented as arrows indicating the direction of their transcription. Genes are colored as listed: carotenoid biosynthesis (*crt*) in yellow, chlorophyll biosynthesis (*bch*) in green, reaction-center-related (*puf*) and light-harvesting complexes-related *(puf*, *puh, puc)* in red, regulatory proteins *(aerR, ppsR*) in blue, other functional proteins in purple, and unidentified open reading frames in white.

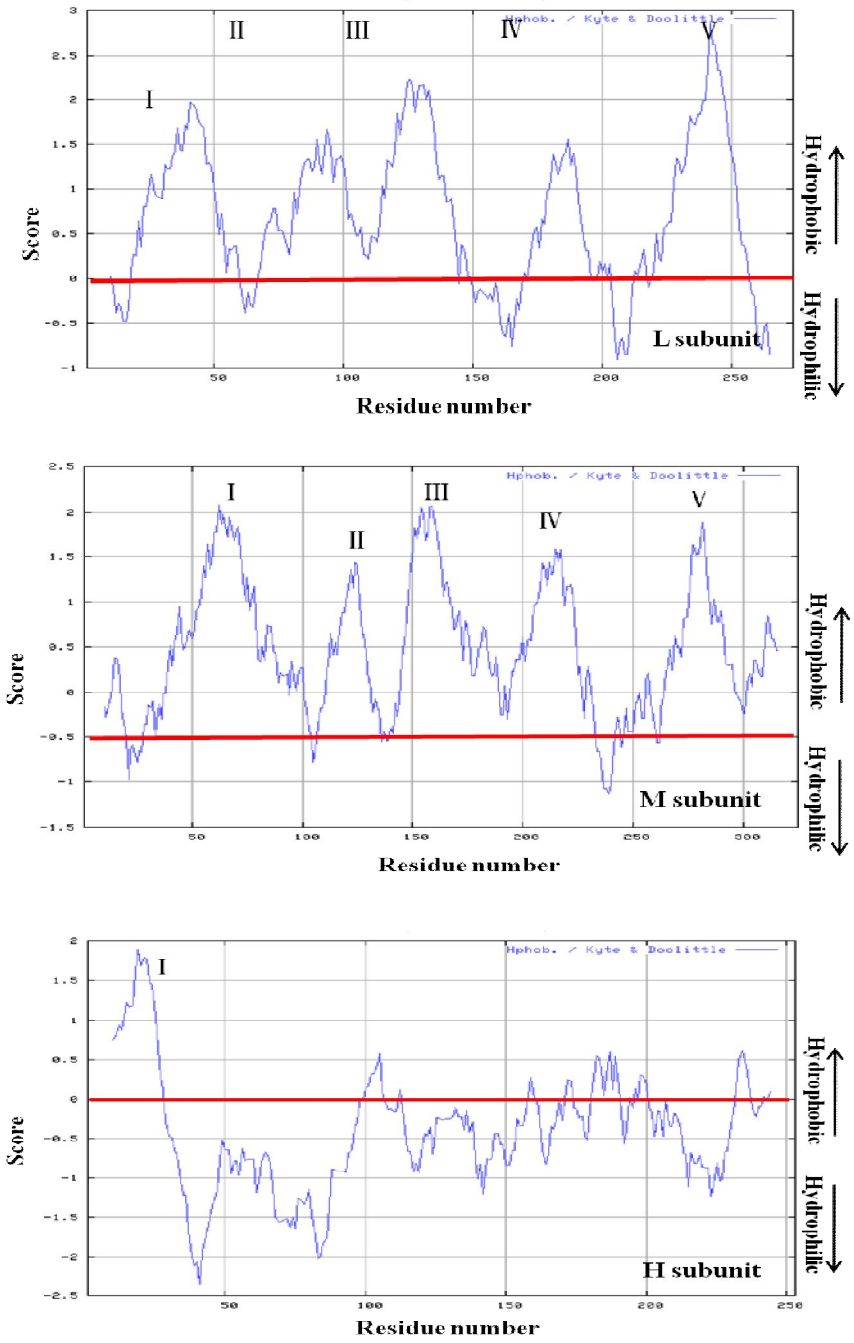


Figure 11 Hydropathy plots of predicted amino acid sequences of L, M, and H subunits, respectively for the identified reaction center in strain ANT.BR.

The absorption spectrum *R. antarcticus* sp. ANT.BR reveals three major peaks the in near-infrared region at 799 nm, 819 nm, and 866 nm, respectively (Madigan, et al., 2000). An hypothesis was proposed that both LH I (B875 type) and LH III (B800/820 type) complexes exist in this strain since the 866nm peak is likely due to absorption by the LH I (B875 type) complex and the 799nm and 819nm peaks are likely due to absorption by the LH II (B800/B820 type) complex. Furthermore, LH II (B800/850 type) complex, although very common in photosynthetic purple bacteria, is predicted to be absent in strain ANT.BR because there was no distinct peak between 850nm and 855nm in the absorption spectrum of intact cells of strain ANT.BR (Madigan, et al., 2000).

However, the genome data described here suggests a different explanation. The genes coding for α subunit (PufA) and β subunit (PufB) of LH I complex in strain ANT.BR have been identified in the annotation results (Figure 10) and when the putative amino acid sequences of the α- and β-subunit of LH I complex are compared with the protein database in the National Center for Biotechnology Information (NCBI), the results support the B875-type LH I complex, which is consistent with part of the previous hypothesis. Surprisingly, the genes coding for the α- and β-subunit of LH III (B800/820) complex are not identified; but instead, the genes encoding the α- and β-subunit of LH II (B800/850) complex, *pucA* and *pucB*, are found according to the annotation result (Figure 10). Also, a phylogenetic tree based on the putative amino acids for the LH complexes in *R.*

antarcticus and other bacteria implies that the LH genes identified in *R. antarcticus* are highly likely to be the genes encoding LH II complex (Figure 12). This finding is entirely in contrast to the previous hypothesis made by Madigan (2000).

Hence, based on the absorption spectra (Madigan 2000), the current annotation results and considerable research done in other purple bacteria that also have an absorption peak around 820 nm under certain circumstances (Angerhofer, Cogdell, & Hipkins, 1986; R.J. Cogdell, Durant, Valentine, Lindsay, & Schmidt, 1983; Deinum et al., 1991; A. T. Gardiner, Cogdell, & Takaichi, 1993; A.T. Gardiner, MacKenzie, Barrett, Kaiser, & Cogdell, 1996; Halloren et al., 1995; Mascle-Allemand, Duquesne, Lebrun, Scheuring, & Sturgis, 2010; McLuskey, Prince, Cogdell, & Isaacs, 2001; Tadros & Waterkamp, 1989), a hypothesis is proposed here that the structural genes encoding LH I (B875), LH II (B800/850) and LH III (B800/820) complexes are all present in *R. antarcticus* genome. Detailed supporting evidence is shown below.

Although most purple photosynthetic bacteria contain the LH I and LH II (B800/850) complexes (R. J. Cogdell, et al., 1999), for some species of purple photosynthetic bacteria, the presence of a third type of light-harvesting apparatus, the LH III (B800/820) complex (which is a spectroscopic variant of LH II complex), was detected when the bacteria are under low light intensity and/or low temperature (Angerhofer, et al., 1986; R.J. Cogdell, et al., 1983; A. T. Gardiner, et

al., 1993; Halloren, et al., 1995). Sometimes, the LH III complex would co-exist with the LH II complex in the cell; but sometimes, LH III complex would completely replace the LH II complex as the temperature and/or light intensity decreases. For instance, when grown under high light intensity (160 μmol $s^{-1}m^{-2}$ and above) at 30ºC (which is the normal growth temperature), *Rhodopseudomonas* (*Rps.*) *acidophila* strain 7050 contained only LH I and LH II apparatuses (detected by absorption spectrum). But as the light intensity decreases, a switchover from B800/850 to B800/820 was observed in the absorption spectra (R.J. Cogdell, et al., 1983; A. T. Gardiner, et al., 1993). In other words, the near-infrared absorption peak originally at 850nm gradually moved to 820nm and the ratio of LH II/ LH III moved strongly in favor of LH III as the light intensity was reduced (A. T. Gardiner, et al., 1993). Strain *Rps. acidophila* 7750 had a similar gradual switchover as strain 7050 when the light intensity was lowered (A. T. Gardiner, et al., 1993). Besides, when growing at low light intensity (like 50 μmol $s^{-1}m^{-2}$), as the growth temperature decreased, the switch over from B800/850 to B800/820 was observed in strain 7750 too, and the switchover was complete when the temperature was below 24ºC (A. T. Gardiner, et al., 1993). Thus, the growing conditions (like light intensity and growth temperature) have a remarkable impact on the composition of light-harvesting apparatuses in certain bacteria. Recently, a model has been proposed to explain

how this gradual switchover might occur during those conditions (Mascle-Allemand, et al., 2010).

As to why B800/820 complexes tend to take the place of B800/850 complexes under low light intensity and/or low temperature, Deinum (1991) provided one possible answer. In *Rps. acidophila*, the B800/820 complex, instead of the B800/850 complex, acts as an energy transfer barrier because once the LH I complex is excited, excitation could be transferred back to B800/850, but not to B800/820. The energy transfer from LH III to LH I complex is very fast and efficient. Collectively, the presence of B800/820 complexes enables the absorbed light-energy to be more effectively restricted to the photosynthetic reaction centers so the light-harvesting process is more efficient under those "low" or "stressed" conditions (Deinum, et al., 1991).

And in fact, the structural genes for both LH II and LH III complexes were found in the strain 7050 and strain 7750 (A. T. Gardiner, et al., 1993). So their different phenotypes under different growing conditions were most likely due to regulatory processes (A. T. Gardiner, et al., 1993). For other purple bacteria that have similar growth condition-dependent variability in light-harvesting apparatuses, multiple *puc* clusters (encoding the alpha- and beta-subunits for LH complexes) have been found to be involved in the LH apparatus's assembly as the growth condition changes (A.T. Gardiner, et al., 1996; Mascle-Allemand, et al., 2010; Tadros & Waterkamp, 1989).

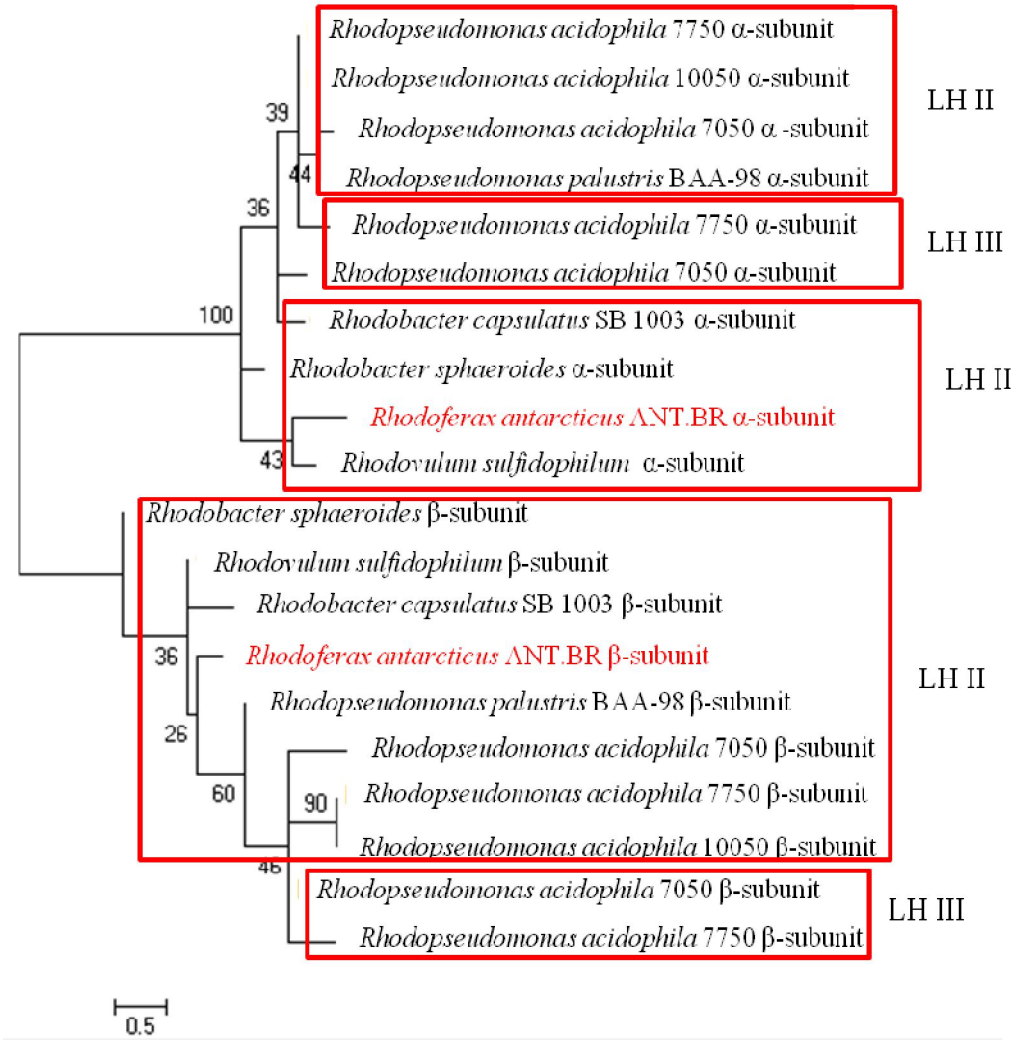


Figure 12 Phylogenetic analysis of α- and β-subunit for LH II and LH III complexes in *R. antarcticus* ANT.BR and other bacteria. Maximum likelihood unrooted tree is based on a multiple protein alignment of α- and β-subunit for LH II and LH III complexes. The sequences found in strain ANT.BR are in red (MEGA is used to build the tree).

According to the details above, a hypothesis is proposed here that the structural genes for LH I (B875), LH II (B800/850) and LH III (B800/820) apparatuses are all present in *R. antarcticus* genome, but because of its special growing conditions (very low light intensity: 1000 lx ~ 19 μmol $s^{-1}m^{-2}$; very low growth temperature: 5ºC) (Madigan, et al., 2000), expression of the genes coding for LH II complex may be inhibited or at least remarkably reduced and expression of the genes coding for LH III complex may be turned on leading to the replacement of LH II with LH III complex, which could make the light-harvesting process more efficient. This hypothesis could well explain the presence of the structural genes for LH II complex in the current genome data and the absorption peaks at 799 nm and 819 nm (probably due to the presence of LH III (B800/820) complex) and the peak at 866 nm (probably due to the presence of LH I (B875) complex). As for not detecting the structural genes for LH III complex in the *R. antarcticus* genome sequence, a possible reason is that they are simply difficult to detect by sequence comparison alone. LH complex proteins are short (40-50 amino acids) and there are only very limited sequences available in databases for comparison. A careful examination of the genome sequence using more sensitive analysis tools might be necessary to detect these genes. And there are two gaps left in the current genome and it is possible that the genes for LH III complexes exist in those two gap regions.

Therefore, further experiments are necessary to elucidate what types of LH complexes exist in the genome of *R. antarcticus*. For example, try to obtain the sequences of the two gap regions in the current genome and then carry out the genomic analysis based on the complete genome data. Also, the effect of light intensity or growth temperature on light-harvesting apparatus in strain ANT.BR would help us to see whether the growth conditions would affect the photosynthetic units in *R. antarcticus*.

Multiple Rubisco Forms

Strong hybridization between a *Rhodobacter sphaeroides* Rubisco large subunit (*cbbL*) probe and strain ANT.BR DNA was observed indicating the presence of Calvin cycle enzymes in strain ANT.BR (Madigan, et al., 2000). And indeed, current genome sequence data of strain ANT.BR reveals two *cbb* clusters (Figure 13), cbb_I and cbb_{II}, encoding enzymatic components involved in the Calvin-Benson-Bassham (CBB) reductive pentose phosphate cycle where CO_2 is combined with ribulose-1,5-bisphosphate (RuBP) to form two molecules of 3-phophoglcerate and further facilitate the conversion of CO_2 to reduced organic carbon (Tabita, 1999). The two *cbb* gene clusters are located far away from each other in the current draft genome with more than 1,300 kb separating them.

The first cluster, cbb_I, as a putative *cbb* operon, contains *cbbR* encoding the Rubisco operon positive transcriptional regulator CbbR, which is actually part of

a wider class of transcriptional regulators, the LysR Transcriptional Regulators, or LTTRs (Schell, 1993). The orientation of *cbbR* gene is usually adjacent to, and divergently transcribed from, the structural genes of the *cbb* operon (Tabita, 1999), which agrees well with the organization of cbb_I operon in ANT.BR (Figure 13). Immediately upstream of *cbbR* are *cbbL* and *cbbS*, coding for the large and small subunits of Form I RuBP carboxylase/oxygenase (Rubisco), respectively. The *cbbO* and *cbbQ* genes, both of which encode Rubisco activation proteins, are immediately downstream of the *cbbLS* genes. Overall, this type of gene structure arrangement of cbb_I operon (Figure 13) indicates a well-recognized subtype of Form I Rubisco, Form IAq (Figure 13), which typically contains L and S gene pairs (*cbbLS*) associated with *cbbQ*/*O* gene pair, with *cbb* metabolic genes located elsewhere in genome (Badger & Bek, 2008). The phylogenetic analysis based on the Rubisco large subunit sequences further confirms the presence of Form IAq Rubisco in strain ANT.BR (Figure 14).

The other cluster cbb_{II}, which contains *cbbM* (a.k.a. *cbbL2*) encoding Form II Rubisco is unusually organized (Figure 13). Normally, Form II Rubiscos are found in two distinct gene organizations (Badger & Bek, 2008). One (*cbbM*) is in association with the *cbb* metabolic genes that encode enzymes of the CBB cycle, such as *cbbK*, *cbbP*, *cbbT*, and *cbbF*, etc. This type of Form II Rubisco is found to be in facultative autotrophs (or mixotrophs) that can readily switch between CBB cycle CO_2 fixation and external organic substrates as carbon source (Badger &

Bek, 2008). It is speculated that it is convenient and efficient for this carbon source switch process if *cbbM* and other *cbb* metabolic genes are located closely together. In the other arrangement (*cbbM-cbbQ-cbbO*), *cbbM* is associated with the downstream presence of the *cbbQ/O* gene pair and *cbb* metabolic genes are located elsewhere in the genome. This type of Form II Rubisco is found in obligate autotrophs that must continually express the CBB cycle and therefore close regulation with Rubisco is not necessary (Badger & Bek, 2008).

Although from the standpoint of gene arrangements, Form II Rubisco in strain ANT.BR seems to be the combination of the two gene organizations mentioned above, with the *cbbQ/O* gene pair located downstream of *cbbM* and *cbb* metabolic genes located upstream (not elsewhere in the genome) (Figure 13). With respect to the transcription synchronization aspect, Form II Rubisco here is, in essence, the *cbbM-cbb* (metabolic) form, which is common in facultative autotrophs. For strain ANT.BR, which can be grown photoautotrophically, photoheterotrophically, and chemoorganotrophically, this gene organization of Form II Rubisco is operationally convenient for properly adjusting the carbon source under different environments.

Furthermore, the gene *cbbR*, which encodes a Rubisco positive transcription regulator, is unusually situated within the cbb_{II} cluster. It is located downstream of *cbbM* and upstream of the *cbb* metabolic genes, and is situated in the same transcriptional orientation as *cbbM*. Although rare, a similar

organization of this Form II Rubisco was observed in another purple non-sulfur bacterium, *Rhodospirillum rubrum* (Falcone & Tabita, 1993; Nargang, McIntosh, & Somerville, 1984). Fragments of different lengths upstream of the *cbbM* gene were deleted to observe *cbbM* expression in *R. rubrum* and the results indicated additional sequences (the coding region of adjacent gene *cbbE*) upstream of *cbbR* are necessary for CbbR to initiate transcription (Falcone & Tabita, 1993). However, how *cbbR* controls transcription and the location of the interaction site in the *cbb* cluster is still unknown. Lastly, there are two other unidentified open reading frames (ORFs) and one gene (*cobB*) encoding cobyrinic acid a,c-diamide synthase, which catalyzes the conversion of cobyrinic acid to cobyrinic acid a,c-diamide, between *cbbR* and *cbbM* s (Figure 13) and it is not clear why *cobB* gene exists in the *cbb* gene cluster.

Different Rubisco forms have been reported to exhibit different kinetic properties. It is speculated that each Rubisco form is adapted to a specific environmental CO_2/O_2 concentration ratio, and that the presence of multiple Rubisco forms in a single organism can be advantageous (Badger & Bek, 2008). This is best exemplified by examining two key kinetic parameters, S_{rel} representing the CO_2/O_2 substrate specificity and k_{cat}. According to the results from previous experiments in other bacteria, Form IAq Rubisco has a significantly higher S_{rel} value (30-53) than Form II Rubisco (9-15), but has a lower reaction rate ($k_{cat} \sim 3.7\ s^{-1}$) than Form II Rubisco ($k_{cat} \sim 5.7\ s^{-1}$), indicating

the adaptation to medium-to-low-CO_2 environments and medium-to-high-CO_2 environments, respectively (Badger & Bek, 2008; Tabita, 1999). In fact, Form II Rubisco generally shows weaker discrimination of CO_2 from O_2 than any subtype of Form I Rubisco (Tabita, Hanson, Satagopan, Witte, & Kreel, 2008). It should also be noted that the gene *cbbZ* encoding phosphoglycolate phosphatase (PGP) (Schaferjohann, Yoo, Kusian, & Bowien, 1993) is located in cbb_{II} cluster to be transcribed together with other *cbb* metabolic genes. *cbbZ* is important, especially in the environment of relatively low CO_2 and high O_2 concentrations. Under that circumstance, Rubisco could catalyze the oxygenation of RuBP with O_2, instead of the carboxylation with CO_2, to produce 2-phosphoglycolate, which is a potential inhibitor of triosephosphate isomerase (TPI), an enzyme in the Calvin cycle (Wolfenden, 1970). Due to the presence of PGP, 2-P-glycolate is dephosphorylated to glycolate, which can be used in the Calvin cycle via the D-glycerate pathway (Figure S2) (Kusian & Bowien, 1997).

Although it is still not very clear how multiple forms of Rubisco work in a single bacterium, it is speculated that the meaning of the presence of multiple forms in bacteria, like *R. antarcticus*, is to give bacteria more metabolic flexibility. This agrees with the fact cells of *R. antarcticus* are highly motile and thus may inhabit a wide range of environments.

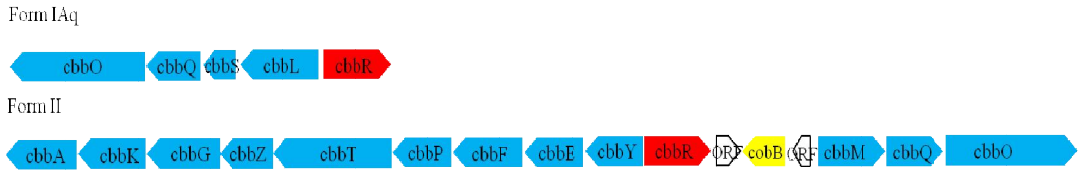


Figure 13 Genomic organizations of two forms of Rubisco in *Rhodoferax antarcticus* ANT.BR. The structural gene arrangements for Form IAq and Form II of Rubisco found in *Rhodoferax antarcticus* ANT.BR are shown. Blue indicates those genes encoding large and small subunits of Rubisco and other CBB cycle enzymes. Red indicates the *cbb* operon transcription regulator. Yellow indicates the gene encoding cobyrinic acid a,c-diamide synthase which is not involved in CBB cycle. White indicates unidentified open reading frames. Those *cbb* metabolic genes are *cbbA*, encoding fructose-1,6-bisphosphate aldolase, *cbbK*, encoding 3-phosphoglycerate kinase, *cbbG*, encoding glyceraldehydes-3-phosphate dehydrogenase, *cbbZ*, encoding 2-phosphoglycolate phosphatase, *cbbT*, encoding transketolase, *cbbP*, encoding phosphoribulokinase, *cbbF*, encoding fructose-1,6/sedoheptulose-1,7-bisphosphatase, *cbbE*, encoding ribulose-5-phosphate-3-epimerase, and *cbbY*, encoding an unknown function.

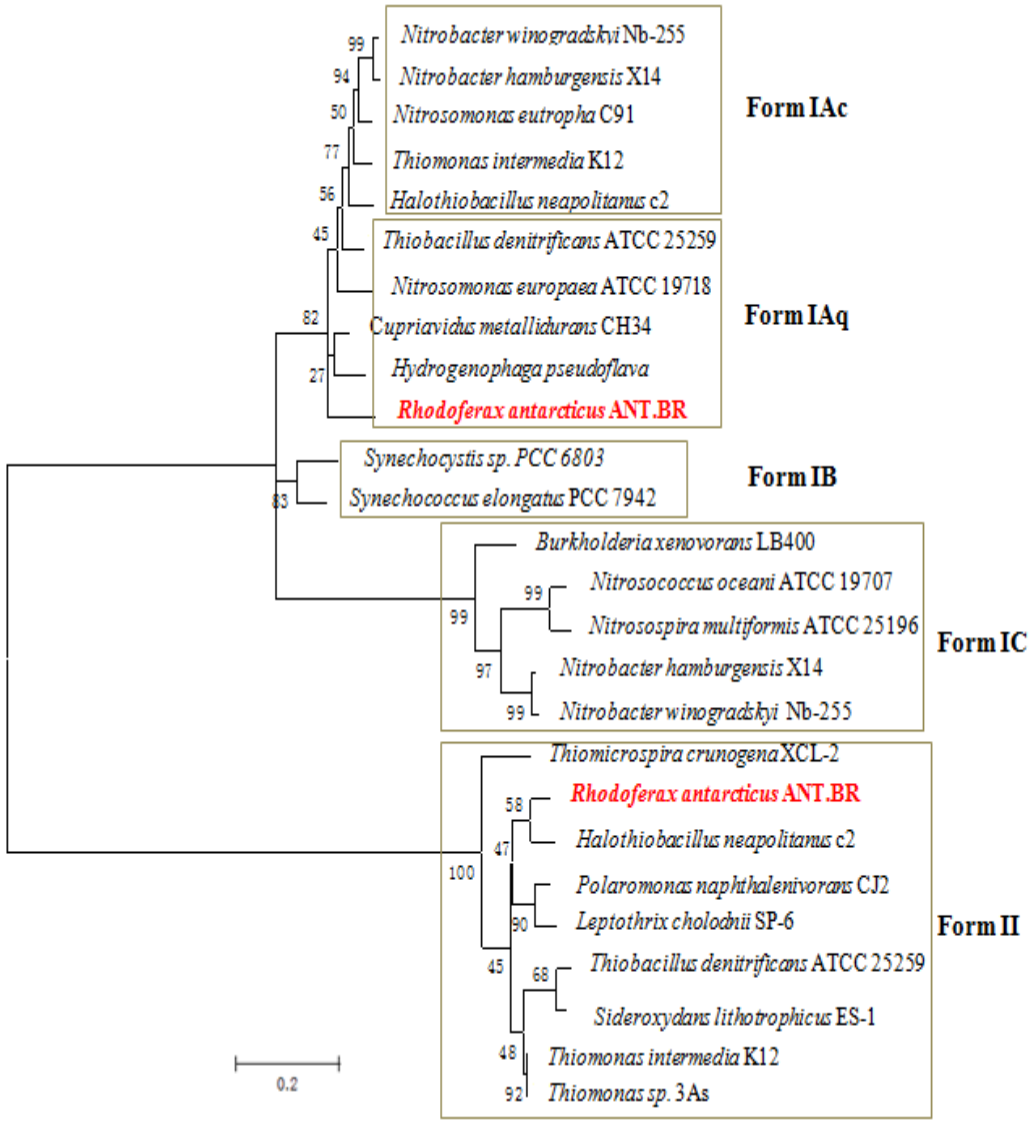


Figure 14 Phylogenetic analysis of two forms of Rubisco in *Rhodoferax antarcticus* ANT.BR. Maximum likelihood unrooted tree of Rubisco Form IAq and Form II based on a multiple protein alignment of the large subunit of Rubiscos. The two forms of Rubisco found in strain ANT.BR are in red (MEGA is used to build the tree).

Flagella, Motility and Chemotaxis

Rhodoferax antarcticus sp. ANT.BR cells were reported to be highly motile based on microscopic observations, and one or more polar flagella are present in this organism (Madigan, et al., 2000). No lateral flagella have been reported so far. And indeed, the annotation result of the genome sequence reveals a large number of flagellar, motility, and chemotaxis genes (71 genes, Table S1) distributed among three separate flagellar gene regions (RG1, RG2, RG3) along the chromosome (Figure 15 (A)).

Most of the structural flagellar or related genes (*flg, flh, fli* genes) are present as a single copy with only one exception, *fliC*, which is duplicated in RG2 (Figure 15(B)). Besides the genes necessary for flagellation, two genes (*motA, motB*) for motility and 14 *che* genes (*cheY*, *cheA*, *cheW*, *cheR*, *cheD*, *cheB* and their duplicates) for chemotaxis have also been identified in the current genome (Figure 15(B)), which completes the regular three gene components (flagellation-motility-chemotaxis) in flagellar gene clusters. There are 12 genes (grey boxes in Figure 15(B)) encoding regulatory and other proteins in the three flagellar gene clusters. One gene (Grey 1) encodes a transcriptional regulator of the ROK family. One gene (Grey 2) encodes a transcriptional regulator of the XRE family containing a DNA-binding helix-turn-helix domain (HTH). This type of XRE family transcriptional regulator along with some other regulators can bind directly to the promoter region of the flagellar master regulator *flhDC* to repress the

flagellar gene expression during fimbrial expression (Pearson & Mobley, 2008) and isa repressor of bacterial motility during fimbrial expression in *Proteus mirabilis* (Li, Rasko, Lockatell, Johnson, & Mobley, 2001; Pearson & Mobley, 2008). Three other genes (Grey3, Grey6, Grey9) encode LuxR family transcriptional regulators containing a CheY-like receiver domain and an HTH DNA-binding domain. Two genes (Grey4, Grey5) encode response regulators with histidine kinase domains. One gene (Grey7) encodes Diguanylate cyclase/phosphodiesterase with PAS/PAC and EAL sensor. One gene (Grey8) encodes signal transduction histidine kinase. Two genes (Grey10, Grey 12) encode methyl-accepting chemotaxis proteins (MCPs) with TarH domains. One gene (Grey11) encodes the protein with an anti-sigma-factor antagonist (STAS) domain.

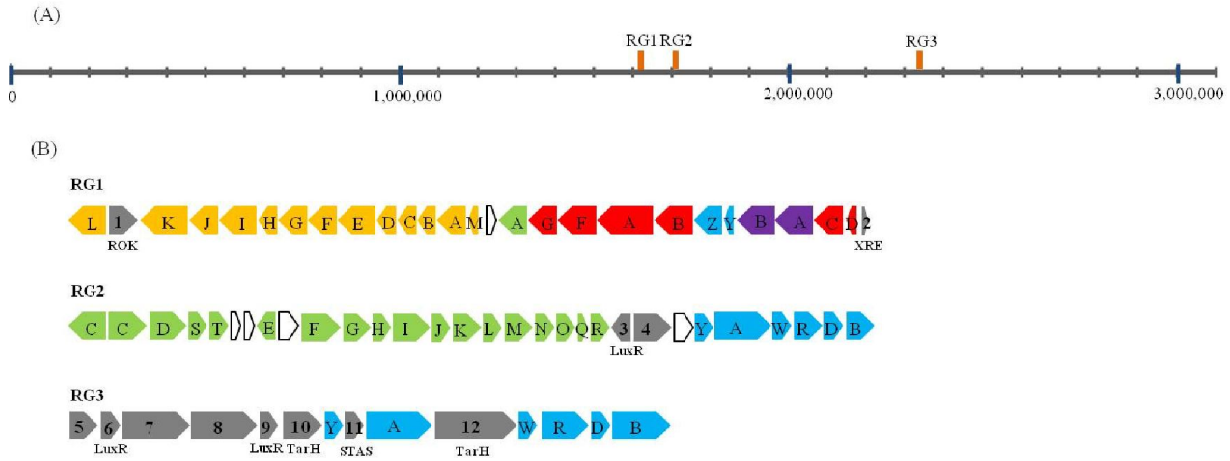


Figure 15 Flagellar gene cluster arrangement. (A) A linear representation of the *Rhodoferax antarcticus* ANT.BR genome with the relative position of the three flagellar gene regions; (B) genes in individual clusters are presented as arrows indicating the direction of transcription and colored as follows: *flg* genes, yellow; *flh* genes, red; *fli* genes, green; *che* genes, blue; *mot* genes, purple; regulatory and other functional proteins, grey; and unidentified open reading frames, white. Under some regulatory proteins are its predicted functional domains.

Type IV Pili

Several clusters of genes encoding the pilin and the assembly for Type IV pili are present in the current genome sequence. These include *pil*A, encoding the major Type IV pilin subunits; *pilE*, *pil*V, *pil*W, and *pil*X, encoding minor pilin-like proteins; *pil*C, encoding a tip adhesin (Rudel, Scheuerpflug, & Meyer, 1995), but also found in the cell membrane (Mattick, 2002); *pil*D, encoding a bifunctional enzyme that can proteolytically remove the leader sequence of prepilin and also methylate the newly created N-terminal amino acid of the pilin; *pil*B, encoding a motor protein that is required for pilus extension; and finally, *pil*T, encoding another motor protein that has an NTP-binding cassette for pilus retraction (Herdendorf, McCaslin, & Forest, 2002; Wu & Kaiser, 1995). Not every Type IV pilus system has the retraction protein PilT, but its presence can enable bacteria to move on solid surface in a process called twitching motility. Additionally, there is a putative operon (*pil*MNOPQ) in ANT.BR that was reported to be required for pili assembly, phage sensitivity and twitching motility in other species (Martin, Watson, McCaul, & Mattick, 1995). The gene *pil*Q encodes an outer-membrane secretin, which is a member of the large secretin family facilitating the passage of filamentous phage particles, DNA, folded proteins, and other macromolecules (Dubnau, 1999; Genin & Boucher, 1994; Linderoth, Simon, & Russel, 1997). Lastly, there is a gene predicted to encode the Type IV pilus assembly protein tip-associated adhesin (PilY1). Figure 16

illustrates a model to show how the products of those pilus-related genes are involved in pilus assembly and retraction.

So far, there are no reports about Type IV pili in strain ANT.BR and its function is not clear. It is likely involved in a flagella-independent form of movement known as twitching motility due to the presence of the motor protein PilT. It might also be involved in some other pili functions like adhesion, biofilm formation, phage transduction (since phage-related genes have been found in strain ANT.BR), and DNA uptake during transformation (the latter two are unique features of Type IV pili compared to other type of pili) (Proft & Baker, 2009). This Type IV pilus-dependent motility may also be used to perform phototaxis (Bhaya, Bianco, Bryant, & Grossman, 2000; Bhaya, Takahashi, & Grossman, 2001).

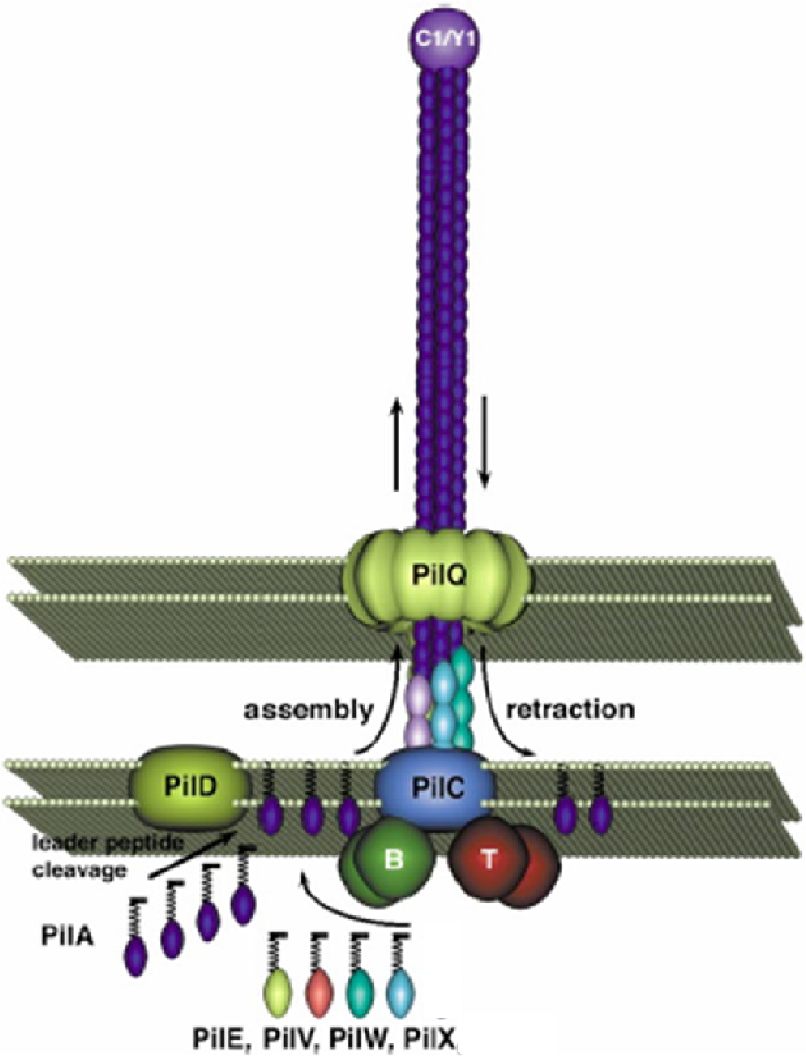


Figure 16 A model for type IV pilus assembly and retraction (Mattick, 2002)

Chapter 4

SUMMARY

Rhodoferax antarcticus sp. ANT.BR, isolated from a microbial mat in Antarctica, was chosen for complete genome sequencing and analysis in my study. It is the first photosynthetic β-proteobacterium that is under complete genome sequencing and it is also the first described psychrophilic anoxygenic phototrophic bacterium. Current annotation results reveal a chromosome of about 3.8Mb and a plasmid of about 200kb. Further genomic analyses indicate the presence of the genes for Molybdenum nitrogenase as well as two forms of Rubiscos, Form IAq and Form II, both of which have been proven to exhibit different kinetic properties in other bacteria. Two Rubisco forms may imply metabolic flexibility that *R. antarcticus* might exhibit to adapt to diverse environments. Previous experimental results suggest the presence of structural genes for three types of LH apparatuses in *R. antarcticus*, LH I (B875), LH II (B800/850), and LH III (B800/820), although only LH I and LH II have been found in the genomic sequence. The LH III complex in *R. antarcticus* might be a growth-condition dependent light-harvesting apparatus as other research results suggest. Lastly, the genes encoding polar flagella and type IV pili have also been identified, both of which give *R. antarcticus* high flexibility in motility.

EFERENCES

Angerhofer, A., Cogdell, R. J., & Hipkins, M. F. (1986). A spectral characterisation of the light-harvesting pigment-protein complexes from *Rhodopseudomonas acidophila*. *Biochimica et Biophysica Acta (BBA)-Bioenergetics, 848*(3), 333-341.

Badger, M. R., & Bek, E. J. (2008). Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO_2 acquisition by the CBB cycle. *Journal of Experimental Botany, 59*(7), 1525-1541. doi: 10.1093/jxb/erm297

Bhaya, D., Bianco, N. R., Bryant, D., & Grossman, A. (2000). Type IV pilus biogenesis and motility in the cyanobacterium *Synechocystis* sp PCC6803. *Molecular Microbiology, 37*(4), 941-951.

Bhaya, D., Takahashi, A., & Grossman, A. R. (2001). Light regulation of type IV pilus-dependent motility by chemosensor-like elements in *Synechocystis* PCC6803. *Proceedings of the National Academy of Sciences of the United States of America, 98*(13), 7540-7545.

Blankenship, R. E. (2001). Molecular evidence for the evolution of photosynthesis. *Trends in Plant Science, 6*(1), 4-6.

Brenchley, J. E. (1996). Psychrophilic microorganisms and their cold-active enzymes. *Journal of Industrial Microbiology & Biotechnology, 17*(5-6), 432-437.

Brinkmeyer, R., Glockner, F. O., Helmke, E., & Amann, R. (2004). Predominance of beta-proteobacteria in summer melt pools on Arctic pack ice. *Limnology and Oceanography, 49*(4), 1013-1021.

Cogdell, R. J., Durant, I., Valentine, J., Lindsay, J. G., & Schmidt, K. (1983). The isolation and partial characterisation of the light-harvesting pigment-protein complement of *Rhodopseudomonas acidophila*. *Biochimica et Biophysica Acta (BBA)-Bioenergetics, 722*(3), 427-435.

Cogdell, R. J., Isaacs, N. W., Howard, T. D., McLuskey, K., Fraser, N. J., & Prince, S. M. (1999). How photosynthetic bacteria harvest solar energy. *Journal of Bacteriology, 181*(13), 3869-3879.

Deinum, G., Otte, S. C. M., Gardiner, A. T., Aartsma, T. J., Cogdell, R. J., & Amesz, J. (1991). Antenna organization of *Rhodopseudomonas acidophila*: a study of the excitation migration. *Biochimica Et Biophysica Acta, 1060*(1), 125-131.

Deisenhofer, J., Epp, O., Miki, K., Huber, R., & Michel, H. (1985). Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3Å resolution. *Nature, 318*, 618-624.

Drews, G., & Golecki, J. (1995). Structure, molecular organization, and biosynthesis of membranes of purple bacteria. In Blankenship RE, Madigan MT, and Bauer CE (Eds.) *Anoxygenic photosynthetic bacteria (pp.231-257)*. Dordrecht, The Netherlands: Kluwer Academic Publishers

Dubnau, D. (1999). DNA uptake in bacteria. *Annual Review of Microbiology, 53*, 217-244.

Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research, 8*(3), 186-194.

Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research, 8*(3), 175-185.

Falcone, D. L., & Tabita, F. R. (1993). Complementation analysis and regulation of CO_2 fixation gene expression in a ribulose 1,5-bisphosphate carboxylase-oxygenase deletion strain of *Rhodospirillum rubrum*. *Journal of Bacteriology, 175*(16), 5066-5077.

Feher, G., Allen, J. P., Okamura, M. Y., & Rees, D. C. (1989). Structure and function of bacterial photosynthetic reaction centers. *Nature, 339*(6220), 111-116.

Garcia, D., Richaud, P., Breton, J., & Vermeglio, A. (1994). Structure and function of the tetraheme cytochrome associated to the reaction centers of *Roseobacter denitrificans*. *Biochimie, 76*(7), 666-673.

Gardiner, A. T., Cogdell, R. J., & Takaichi, S. (1993). The effect of growth conditions on the light-harvesting apparatus in *Rhodopseudomonas acidophila*. *Photosynthesis Research, 38*(2), 159-167.

Gardiner, A. T., MacKenzie, R. C., Barrett, S. J., Kaiser, K., & Cogdell, R. J. (1996). The purple photosynthetic bacterium *Rhodopseudomonas acidophila* contains multiple *puc* peripheral antenna complex (LH2) genes: Cloning and initial characterisation of four β / α pairs. *Photosynthesis Research, 49*(3), 223-235.

Genin, S., & Boucher, C. A. (1994). A superfamily of proteins involved in different secretion pathways in gram-negative bacteria: Molecular structure and specificity of the N-terminal domain. *Molecular & General Genetics, 243*(1), 112-118.

Gounot, A. M. (1991). Bacterial life at low temperature: physiological aspects and biotechnological implications. *Journal of Applied Bacteriology, 71*(5), 386-397.

Halloren, E., McDermott, G., Lindsay, J. G., Miller, C., Freer, A. A., Isaacs, N. W., & Cogdell, R. J. (1995). Studies on the light-harvesting complexes from the thermotolerant purple bacterium *Rhodopseudomonas cryptolactis*. *Photosynthesis Research, 44*(1), 149-155.

Herdendorf, T. J., McCaslin, D. R., & Forest, K. T. (2002). *Aquifex aeolicus* PilT, homologue of a surface motility protein, is a thermostable oligomeric NTPase. *Journal of Bacteriology, 184*(23), 6465-6471. doi: 10.1128/jb.184.23.6465-6471.2002

Hiraishi, A., Hoshino, Y., & Satoh, T. (1991). *Rhodoferax fermentans* gen. nov., and sp. nov., a phototrophic purple nonsulfur bacterium previously referred to as the "Rhodocyclus gelatinous-like" group. *Archives of microbiology, 155*(4), 330-336.

Imhoff, J. F. (1992). The family Ectothiorhodospiraceae. In T. H. Balows A, Dworkin M, Harder W, Schleifer K-H (Ed.), *The Prokaryotes* (2nd ed., pp. 3222 – 3229). Berlin Heidelberg New York: Springer.

Irgens, R. L., Gosink, J. J., & Staley, J. T. (1996). *Polaromonas vacuolata* gen. nov., and sp. nov., a psychrophilic marine gas vacuolate bacterium from Antarctica. *International Journal of Systematic Bacteriology, 46*(3), 822-826.

Jung, D. O., Achenbach, L. A., Karr, E. A., Takaichi, S., & Madigan, M. T. (2004). A gas vesiculate planktonic strain of the purple non-sulfur bacterium *Rhodoferax antarcticus* isolated from Lake Fryxell, Dry

Valleys, Antarctica. *Archives of microbiology, 182*(2-3), 236-243. doi: 10.1007/s00203-004-0719-8

Ke, B. (2001). Light-harvesting pigment-protein complexes of photosynthetic bacteria. In Govindjee (Ed.), *Photosynthesis: photobiochemistry and photobiophysics* (Vol. 10, pp. 66-86): Kluwer Academic Publishers.

Kusian, B., & Bowien, B. (1997). Organization and regulation of *cbb* CO_2 assimilation genes in autotrophic bacteria. *Fems Microbiology Reviews, 21*(2), 135-155.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology, 157*(1), 105-132.

Li, X., Rasko, D. A., Lockatell, C. V., Johnson, D. E., & Mobley, H. L. T. (2001). Repression of bacterial motility by a novel fimbrial gene product. *Embo Journal, 20*(17), 4854-4862.

Liebetanz, R., Hornberger, U., & Drews, G. (1991). Organization of the genes coding for the reaction-centre L and M subunits and B870 antenna polypeptides alpha and beta from the aerobic photosynthetic bacterium *Erythrobacter* species OCH114. *Mol Microbiol, 5*(6), 1459-1468.

Linderoth, N. A., Simon, M. N., & Russel, M. (1997). The filamentous phage pIV multimer visualized by scanning transmission electron microscopy. *Science, 278*(5343), 1635-1638.

Madigan, M. T. (2003). Anoxygenic phototrophic bacteria from extreme environments. *Photosynthesis Research, 76*(1-3), 157-171.

Madigan, M. T., Jung, D. O., Woese, C. R., & Achenbach, L. A. (2000). *Rhodoferax antarcticus* sp. nov., a moderately psychrophilic purple nonsulfur bacterium isolated from an Antarctic microbial mat. *Archives of microbiology, 173*(5-6), 449-449.

Madigan, M. T., & Marrs, B. L. (1997). Extremophiles. *Scientific American, 276*(4), 82-87.

Martin, P. R., Watson, A. A., McCaul, T. F., & Mattick, J. S. (1995). Characterization of a five cluster required for the biogenesis of type 4

fimbriae in *Pseudomonas aeruginosa*. *Molecular Microbiology, 16*(3), 497-508.

Mascle-Allemand, C., Duquesne, K., Lebrun, R., Scheuring, S., & Sturgis, J. N. (2010). Antenna mixing in photosynthetic membranes from *Phaeospirillum molischianum*. *Proceedings of the National Academy of Sciences, 107*(12), 5357.

Mattick, J. S. (2002). Type IV pili and twitching motility. *Annual Review of Microbiology, 56*, 289-314.

McLuskey, K., Prince, S. M., Cogdell, R. J., & Isaacs, N. W. (2001). The crystallographic structure of the B800-820 LH3 light-harvesting complex from the purple bacteria *Rhodopseudomonas acidophila* strain 7050. *Biochemistry, 40*(30), 8783-8789. doi: 10.1021/bi010309a

Miller, R. W., & Eady, R. R. (1988). Molybdenum and vanadium nitrogenase of *Azotobacter chroococcum*: Low temperature favours N_2 reduction by vanadium nitrogenase. *Biochemical Journal, 256*(2), 429-432.

Nargang, F., McIntosh, L., & Somerville, C. (1984). Nucleotide sequence of the ribulosebisphosphate carboxylase gene from *Rhodospirillum rubrum*. *Molecular and General Genetics MGG, 193*(2), 220-224.

Okamura, K., Mitsumori, F., Ito, O., Takamiya, K. I., & Nishimura, M. (1986). Photophosphorylation and oxidative phosphorylation in intact cells and chromatophores of an aerobic photosynthetic bacterium, *Erythrobacter* sp. strain OCh114. *Journal of Bacteriology, 168*(3), 1142-1146.

Okamura, K., Takamiya, K., & Nishimura, M. (1985). Photosynthetic electron transfer system is inoperative in anaerobic cells of Erythrobacter species strain OCh 114. *Archives of microbiology, 142*(1), 12-17.

Pearson, M. M., & Mobley, H. L. T. (2008). Repression of motility during fimbrial expression: identification of 14 *mrpJ* gene paralogues in *Proteus mirabilis*. *Molecular Microbiology, 69*(2), 548-558. doi: 10.1111/j.1365-2958.2008.06307.x

Pfennig, N. (1969). *Rhodopseudomonas acidophila*, sp.n., a new species of the budding purple nonsulfur bacteria. *Journal of Bacteriology, 99*(2), 597-602.

Pfennig, N. (1974). *Rhodopseudomonas globiformis*, sp.n., a new species of the Rhodospirillaceae. *Archives of microbiology, 100*(3), 197-206.

Pierson, B., & Castenholz, R. (1995). Taxonomy and physiology of filamentous anoxygenic phototrophs. In Blankenship RE, Madigan MT and Bauer CE (Eds.). *Anoxygenic Photosynthetic Bacteria (pp. 31-47)*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Pierson, B. K., & Castenho.Rw. (1974). A phototrophic gliding filamentous bacterium of hot springs. *Chloroflexus aurantiacus*, gen. and sp. nov. *Archives of microbiology, 100*(1), 5-24.

Pierson, B. K., & Castenholz, R. W. (1974). Studies of pigments and growth in *Chloroflexus aurantiacus*, a phototrophic filamentous bacterium. *Archives of microbiology, 100*(1), 283-305.

Proft, T., & Baker, E. (2009). Pili in Gram-negative and Gram-positive bacteria—structure, assembly and their role in disease. *Cellular and molecular life sciences, 66*(4), 613-635.

Rudel, T., Scheuerpflug, I., & Meyer, T. F. (1995). Neisseria PilC protein identified as type 4 pilus tip-located adhesin. *Nature, 373*(6512), 357-359.

Schaferjohann, J., Yoo, J., Kusian, B., & Bowien, B. (1993). The cbb operons of the facultative chemoautotroph Alcaligenes eutrophus encode phosphoglycolate phosphatase. *Journal of Bacteriology, 175*(22), 7329.

Schell, M. A. (1993). Molecular biology of the LysR family of transcriptional regulators. *Annual Review of Microbiology, 47*, 597-626.

Tabita, F. R. (1999). Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective. *Photosynthesis Research, 60*(1), 1-28.

Tabita, F. R., Hanson, T. E., Satagopan, S., Witte, B. H., & Kreel, N. E. (2008). Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Philosophical Transactions of the Royal Society B-Biological Sciences, 363*(1504), 2629-2640. doi: 10.1098/rstb.2008.0023

Tadros, M., & Waterkamp, K. (1989). Multiple copies of the coding regions for the light-harvesting B800-850 α-and β-polypeptides are present in the *Rhodopseudomonas palustris* genome. *The EMBO Journal, 8*(5), 1303.

Takaichi, S. (1999). Carotenoids and carotenogenesis in anoxygenic photosynthetic bacteria. In C. R. Frank HA, Young A, Britton G (Ed.), *The photochemistry of carotenoids: applications in biology* (pp. 39-69). Dordrecht: Kluwer.

Vincent, W. F., Castenholz, R. W., Downes, M. T., & Howard-Williams, C. (1993). Antarctic cyanobacteria: light, nutrients, and photosynthesis in the microbial mat environment. *Journal of Phycology, 29*(6), 745-755.

Vincent, W. F., Downes, M. T., Castenholz, R. W., & Howard-Williams, C. (1993). Community structure and pigment organisation of cyanobacteria-dominated microbial mats in Antarctica. *European Journal of Phycology, 28*(4), 213-221.

Ward, D., & Castenholz, R. (2000). Cyanobacteria in geothermal habitats. *The Ecology of Cyanobacteria: Their Diversity in Time and Space*, 37-59.

Willey, J. M., Sherwood, L. M., Woolverton, C. J., Prescott, L. M., Harley, J. P., & Klein, D. A. (2008). Procaryotic Cell Structure and Function. *Prescott, Harley, and Klein's microbiology* (pp. 39-78): McGraw-Hill.

Wolfenden, R. V. (1970). Binding of substrate and transition state analogs to triosephosphate isomerase. *Biochemistry, 9*(17), 3404-3407.

Wu, S. S., & Kaiser, D. (1995). Genetic and functional evidence that type IV pili are required for social gliding motility in *Myxococcus xanthus*. *Molecular Microbiology, 18*(3), 547-558.

Yang, S., Lin, Z., Cui, X., Lian, J., Zhao, C., & Qu, Y. (2008). Current taxonomy of anoxygenic phototrophic bacteria. *Wei Sheng Wu Xue Bao, 48*(11), 1562-1566.

Yurkov, V., Gad'on, N., & Drews, G. (1993). The major part of polar carotenoids of the aerobic bacteria <i>Roseococcus thiosulfatophilus</i> RB3 and <i>Erythromicrobium ramosum</i> E5 is not bound to the bacteriochlorophyll a-complexes of the photosynthetic apparatus. *Archives of microbiology, 160*(5), 372-376. doi: 10.1007/bf00252223

Yurkov, V., Gad'on, N., Angerhofer, A. and Drews, G. (1994). Light-harvesting complexes of aerobic bacteriochlorophyll-containing bacteria *Roseococcus thiosulfatophilus*, RB3 and *Erythromicrobium ramosum*, E5

and the transfer of excitation energy from carotenoids to bacteriochlorophyll. *Z Naturforsch Teil C, 49*, 579-586.

Yurkov, V., Schoepp, B., & Vermeglio, A. (1995). Electron transfer carriers in obligately aerobic photosynthetic bacteria from genera *Roseococcus* and *Erythromicrobium*. *Photosynthesis: From Light to Biosphere (ed. Matthis, P.), Dordrecht: Kluwer Academic Publishers*, 543-546.

Yurkov, V. V., & Beatty, J. T. (1998). Aerobic anoxygenic phototrophic bacteria. *Microbiology and Molecular Biology Reviews, 62*(3), 695-724.

APPENDIX A

SUPPLEMENTAL FIGURE S1

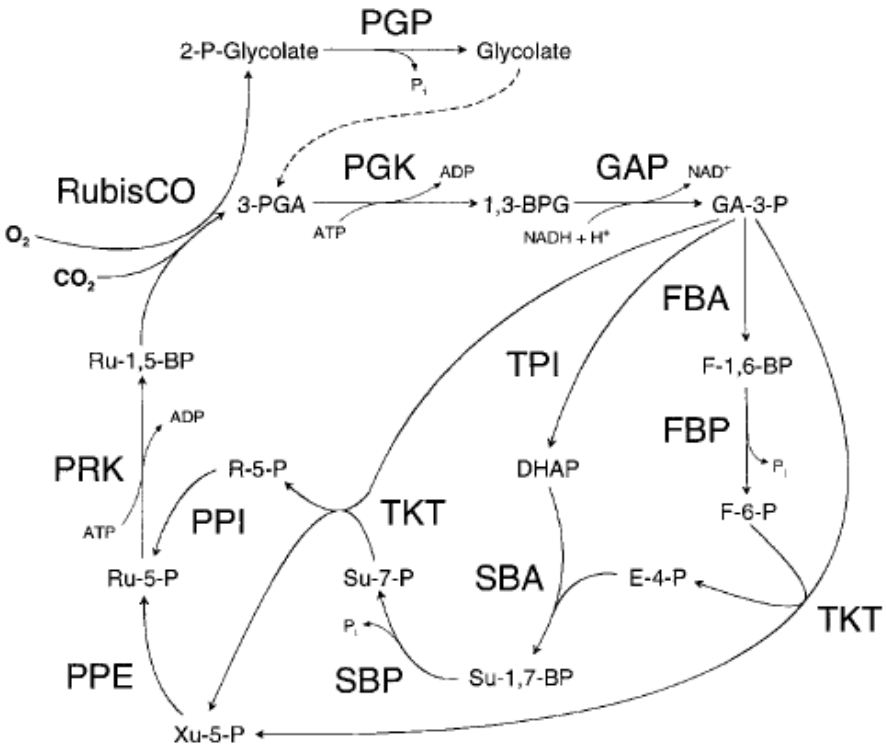


Figure S1 A general scheme for the Calvin cycle. Adapt from (Kusian & Bowien, 1997). Abbreviations used above are: 1,3-BPG, 1,3-bisphosphoglycerate; DHAP, dihydroxyacetone phosphate; E-4-P, erythrose-4-phosphate; F-1,6-BP, fructose-1,6-bisphosphate; F-6-P, fructose-6-phosphate; GA-3-P, glyceraldeyde-3-phosphate; 3-PGA, 3-phosphoglycerate; R-5-P, ribose-5-phosphate; RuBP, ribulose-1,5-bisphosphate; Ru-5-P, ribulose-5-phosphate; Su-1,7-BP, sedoheptulose-1,7-bisphosphate; Su-7-P, sedoheptulose-7-phosphate; Xu-5-P, xylulose-5-phosphate; FBA, fructose-1,6-bisphophate aldolase; FBP, fructose-1,6-bisphosphatase; GAP, glyceraldehyde-3-phosphate dehydrogenase; PGK, 3-phosphoglycerate kinase; PGP, phosphoglycolate phosphatase; PPE, pentose-5-phosphate 3-epimerase; PPI, pentose-5-phosphate isomerase; PRK, phosphoribulokinase; RubisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase; SBA, sedoheptulose-1,7-bisphosphate aldolase; SBP, sedoheptulose-1,7-bisphosphatase; TKT, transketolase; TPI, triosephosphate isomerase.

APPENDIX B

SUPPLEMENTAL TABLE S1

Table S1
Characterization of *Rhodoferax antarcticus* ANT.BR flagellar genes

Gene nomenclature	Copy number	Predicted function
***flg*A**	1	Flagellar basal body P-ring formation protein
***flg*B**	1	Cell-proximal portion of flagellar basal body rod protein
***flg*C**	1	Cell-proximal portion of flagellar basal body rod protein
***flg*D**	1	Flagellar basal body rod modification protein
***flg*E**	1	Flagellar hook protein
***flg*F**	1	Cell-proximal portion of flagellar basal body rod protein
***flg*G**	1	Cell-distal portion of flagellar basal body rod protein
***flg*H**	1	Flagellar L-ring protein
***flg*I**	1	Flagellar P-ring protein
***flg*J**	1	Unknown function
***flg*K**	1	Flagellar hook-associated protein 1
***flg*L**	1	Flagellar hook-associated protein 3
***flg*M**	1	Anti-FliA (anti-σ) factor
***flh*A**	1	Flagellar biosynthesis protein
***flh*B**	1	Flagellar biosynthesis protein
***flh*C**	1	Master regulator of the flagellar regulon
***flh*D**	1	Master regulator of the flagellar regulon
***flh*G**	1	Flagellar biosynthesis protein
***flh*H**	1	Flagellar biosynthesis protein
***fli*A**	1	Transcription initiation (σ) factor
***fli*C**	2	Filament (flagellin protein)
***fli*D**	1	Filament cap (flagellar hook-associated protein 2)
***fli*E**	1	Flagellar hook-basal body complex protein
***fli*F**	1	Flagellar M-ring protein
***fli*G**	1	Flagellar motor switch protein
***fli*H**	1	Flagellar assembly protein
***fli*I**	1	H^+-transporting two-sector ATPase
***fli*J**	1	Flagellar biosynthesis protein
***fli*K**	1	Flagellar hook-length control protein
***fli*L**	1	Flagellar basal body-associated protein
***fli*M**	1	Flagellar motor switch protein
***fli*N**	1	Flagellar motor switch protein
***fli*O**	1	Flagellar assembly protein
***fli*Q**	1	Flagellar biosynthesis protein
***fli*R**	1	Flagellar biosynthesis protein
***fli*S**	1	A chaperon of FliC
***fli*T**	1	A chaperon for flagellum assembly
***mot*A**	1	Motility protein
***mot*B**	1	Motility protein
***che*A**	2	CheY and CheB kinase
***che*B**	2	Demethylation of receptors

***che*D**	2	Methyl-accepting chemotaxis protein
***che*R**	2	Methylation of receptors
***che*W**	2	Positive regulator of CheA activity
***che*Y**	3	Switch regulator, placing it in clockwise state
***che*Z**	1	CheY phosphatase. Antagonist of CheY as switch regulator