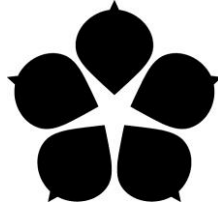


School of Doctoral Studies in Biological Sciences

Faculty of Science, University of South Bohemia



Genome analysis of the haloalkaliphilic bacterium
Rhodobaca barguzinensis

Ph.D. Thesis

Mgr. Karel Kopejtko

Supervisor: doc. Michal Koblížek, Ph.D.

Institute of Microbiology
Czech Academy of Sciences

Scientific advisor: Dr. Jürgen Tomasz

Helmholtz Centre for Infection Research
Braunschweig, Germany

České Budějovice 2019

This thesis should be cited as:

Kopejtka, K., 2019: Genome analysis of the haloalkaliphilic bacterium *Rhodobaca barguzinensis*. Ph.D. Thesis. University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic, 90 p.

Annotation

This PhD thesis presents results of a research focussed on the evolution of phototrophy in the bacterial order *Rhodobacterales* with a special regard to its haloalkaliphilic representatives. The photoheterotrophic bacterium *Rhodobaca barguzinensis* alga05 was used as an organism of choice. Its phylogeny, genome organization, and metabolic potential was characterized. The main result of the thesis is that phototrophy is a genuine trait among the haloalkaliphilic representatives of the *Rhodobacter-Rhodobaca* group inside the *Rhodobacterales* clade.

Declaration [in Czech]

Prohlašuji, že svoji disertační práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury. Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své disertační práce, a to v úpravě vzniklé vypuštěním vyznačených částí archivovaných Přírodovědeckou fakultou elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

České Budějovice, 1.2. 2019

.....

Karel Kopejtka

This thesis originated from a partnership of the Faculty of Science, University of South Bohemia, and the Institute of Microbiology, CAS, Centre Algatech, supporting doctoral studies in the Molecular and Cell Biology and Genetics study programme.

Financial support

This research has been supported by the GAČR projects P501/10/0221 and P501/12/G055, the DAAD project 57155424, the MŠMT projects VaVpI Algatech (ED2.1.00/03.0110) and NPU Algatech Plus (LO1416), and the RFBR grant 16-04-00035. Jürgen Tomasch was supported by the Deutsche Forschungsgemeinschaft (DFG) within the Transregio 51 “Roseobacter“. Dmitry Y. Sorokin was supported by the Gravitation Program (SIAM, grant 24002002) from the Dutch Ministry of Education and Science.

Acknowledgments

Above all I would like to thank to my supervisor Michal Koblížek for giving me the opportunity to be part of his research team and to work on such an interesting topic. I am also very grateful to all my current and former colleagues from the Laboratory of Anoxygenic Phototrophs, as well as to other colleagues and staff from the Algatech Centre, who helped me during my research. My special thanks belongs to my colleague and friend, Jürgen Tomasch, and to my family. I would also wish to thank Jason Dean for language corrections.

List of papers

The thesis is based on the following papers (listed chronologically):

Kopejtka K, Tomasch J, Zeng Y, Tichý M, Sorokin DY, Koblížek M. 2017. Genomic analysis of the evolution of phototrophy among haloalkaliphilic *Rhodobacterales*. *Genome Biol Evol.* 9(7):1950-1962.

KK cultivated the strains, extracted DNA, assembled (with the help of MT) and annotated the genomes, analyzed genomic data, conducted phylogenetic analyses, and wrote part of the paper.

Kopejtka K, Tomasch J, Bunk B, Spröer C, Wagner-Döbler I, Koblížek M. 2018. The complete genome sequence of *Rhodobaca barguzinensis* alga05 (DSM 19920) documents its adaptation for the life in soda lakes. *Extremophiles* 22(6):839-849.

KK cultivated the strain, extracted DNA, conducted phylogenetic and metabolic pathway analyses, and wrote major parts of the paper.

Manuscript:

Kopejtka K, Lin Y, Jakubovičová M, Koblížek M, Tomasch J. Clustered core- and pan-genome content in *Rhodobacteraceae* chromosomes. *Genome Biol Evol.*, under review.

KK gathered genomic data, conducted the orthologous gene cluster analysis and phylogenetic analyses, manually curated the output of Ori-Finder, identified genomic islands using IslandViewer, and wrote major parts of the manuscript.

Michal Koblížek and Jürgen Tomasch, the corresponding authors of listed papers, confirm the contribution of Karel Kopejtka in these papers as described above.

Doc. Mgr. Michal Koblížek, Ph.D.

Dr. Jürgen Tomasch

Contents

List of abbreviations	i
1 Introduction	1
1.1 Bacteria	1
1.1.1 Bacterial genome	2
1.2. Photosynthesis	5
1.2.1 Evolution of photosynthesis	6
1.2.2 Great oxygenation	7
1.3 Aerobic anoxygenic phototrophs	9
1.3.1 Photosynthetic apparatus	10
1.3.2 Photosynthesis gene cluster	11
1.4 Microbial life in soda lakes	12
1.5 Objectives	14
2 Publications.....	15
2.1 Paper 1. The complete genome sequence of <i>Rhodobaca barguzinensis</i> alga05 (DSM 19920) documents its adaptation for life in soda lakes.....	15
2.2 Paper 2. Genomic analysis of the evolution of phototrophy among haloalkaliphilic <i>Rhodobacterales</i>	28
2.3 Paper 3. Clustered core- and pan-genome content on <i>Rhodobacteraceae</i> chromosomes	49
3 Conclusions	78
4 References	79
Curriculum vitae	89

List of abbreviations

16S rRNA – gene(s) coding for a component of the 30S small subunit of prokaryotic ribosomes

AAP – aerobic anoxygenic phototrophic

ATP – adenosine-5'-triphosphate

BChl *a* – bacteriochlorophyll *a*

DNA – deoxyribonucleic acid

HGT – horizontal gene transfer

LH – light-harvesting

ORF – open reading frame

PGC – photosynthesis gene cluster

PNSB – purple non-sulfur bacteria

PS – photosynthetic

RC – reaction center

UV – ultraviolet light

1. Introduction

1.1 Bacteria

Bacteria are prokaryotic microorganisms inhabiting most of the habitats on Earth. They are typically a few micrometres in length, with shape ranging from spheres to rods and spirals. In contrast to eukaryotes, bacterial cells do not possess neither a nucleus nor true organelles. However, some bacteria do contain bacterial microcompartments, which are supposed to act as primitive organelles (Kerfeld *et al.*, 2005).

Based on the source of energy for maintaining of the metabolism, bacteria can use two trophic strategies. Chemotrophic bacteria gain energy by breaking down chemical compounds using oxidation. The second strategy, phototrophy, involves harvesting light energy using photosynthesis. Another criterion reflects the source of carbon used for cellular growth. Bacteria utilizing carbon from organic compounds are called heterotrophs. In contrast, autotrophic bacteria obtain carbon by assimilation of CO₂ (Hellingwerf *et al.*, 1994). Some minor groups of bacteria are able to oxidize methan to gain both energy and carbon (Dalton 2005).

The last common ancestor of Bacteria and Archaea is supposed to be a hyperthermophilic unicellular bacterium that lived 2.5 – 3.2 billion years ago (DiGiulio 2003; Battistuzzi *et al.*, 2004). Phylogenetic studies based on multiple gene phylogenies suggest that bacterial lineage later split from the lineage containing ancestors of both archaeal and eukaryotic species (Brown and Doolittle, 1997; Yan and Wu, 2016; Eme *et al.*, 2017). Bacteria were also crucial for the further evolution of the third domain of life, Eukaryotes. During multiple endosymbiotic events domesticated bacteria gave rise to cellular organelles, such as mitochondria and plastids (Gray 1992; Poole and Penny, 2007; Archibald 2009).

Since most of the bacterial species defy laboratory cultivation, we know of their existence only from molecular (culture-independent) studies (Rappé and Giovannoni, 2003). Based on molecular phylogeny domain Bacteria is divided into major lineages called phyla (or divisions) (Fig. 1). There are currently 34 formally described bacterial phyla (July 2018; Parte 2018), and over 20 candidate phyla established using strains which can not be cultivated (Rappé and Giovannoni, 2003). The total number of bacterial phyla has been estimated to exceed 1,000 (Yarza *et al.*, 2014).



Fig. 1 16S rRNA phylogenetic tree of the Bacteria domain. Phyla containing photosynthetic species are marked by orange wedges. Scale bar represents changes per position. Adapted from Zeng *et al.*, 2014.

1.1.1 Bacterial genome

Genomes contain the complete genetic material of an organism stored in its DNA. Prokaryotic genomes are in general smaller than eukaryotic ones. In contrast to eukaryotes (Tab. 1), bacterial genomes usually exist in the form of single circular chromosomes (Fig. 2). However, there are some bacterial species, which have linear chromosomes (Volff and Altenbuchner, 2000; Chaconas and Chen, 2005) or multiple chromosomes (Suwanto and Kaplan, 1992; Casjens 1998). Additional genomic information can be stored in autonomously replicating extrachromosomal elements, plasmids (Kado 2015). Genes in bacterial genomes are often organised into operons. Genes consist of coding parts (exons) without any non-coding portions of the genes (introns), which are largely present in eukaryotic genes.

Table 1 Comparison of a prokaryotic and an eukaryotic genome.

Feature	<i>E. coli</i> K-12 genome	Human genome
Genome size [bases]	4.6 Mb	3.2 Gb
GC content [%]	50.8	40.9
No. of genes	4,288	20,000
No. of chromosomes	1	46
Chromosome conformation	circular	linear

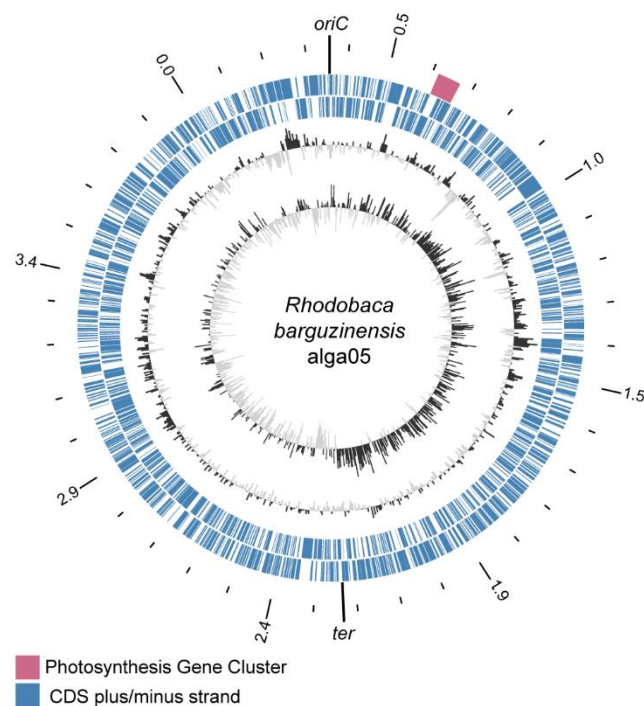


Figure 2 Schematic representation of the circular *Rhodobaca barguzinensis* alga05 chromosome. The outer to inner rings represent: scale of genome size in Mb; features of the chromosome [origin (*oriC*) and terminus (*ter*) of replication]; position of the PGC (in purple); position of ORFs encoded on the plus/minus strand (in blue); GC content (in black/gray); GC skew (in black/gray).

Evolution can be perceived as a process leading to the adaption of a species to its environment. Organisms are under constant selective pressure due to changing life conditions. Evidence for these dynamic changes can be also found on the genomic level. Thanks to the

progress in sequencing techniques, especially with the development of the next-generation sequencing (Ekblom and Wolf, 2014; Land *et al.*, 2015; Vincent *et al.*, 2017), the number of sequenced bacterial genomes steadily grows over time. This allowed, among other things, to study the mechanisms by which their genomes evolve.

Bacterial genomes can evolve through four mechanisms: spontaneous mutations, gene duplications, horizontal gene transfer (HGT) (Treangen and Rocha, 2011; Toussaint and Chandler, 2012) and gene loss (Koonin and Wolf, 2008; Keppen *et al.*, 2013; Kopejtko *et al.*, 2017). Spontaneous mutations are a result of errors occurring during the DNA replication or of an exposure to mutagens. They can lead to either a negative change (protein becomes non-functional), positive change (organism increases its fitness, *e.g.* gains positive traits) or neutral change (no effect on the fitness of the organism) in the nucleotide sequence of the DNA. Gene duplications provide new genetic material for consequent evolutionary processes, such as mutations (Zhang 2003). During the HGT, bacterial cells exchange genetic material using extrachromosomal elements, such as plasmids (Kado 2015) and gene transfer agents (Fig. 3; Lang *et al.*, 2017). Due to the selection pressure combined with significant evolutionary benefits these processes can be followed by a rapid spread and fixation of beneficial traits inside a bacterial population (Wilmes *et al.*, 2008). In balance with the gene gain (genome expansion) is the optimization of genome size using gene loss called genome streamlining (Koonin and Wolf, 2008). Natural selection puts a constant pressure on genome reduction to save energy, time, and nutrients necessary for the genome replication (Rosengarten *et al.*, 2000). This effect is especially pronounced in case of parasitic and endosymbiotic species (Boscaro *et al.*, 2013).

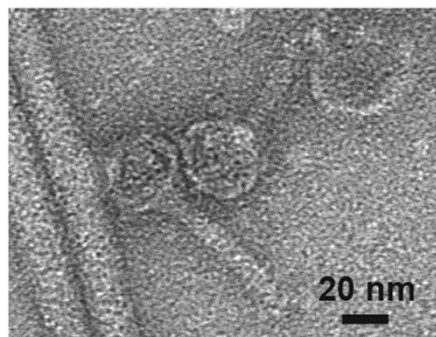


Figure 3 Electron micrograph showing two gene transfer agent particles. Reprinted with permission from Tomasch *et al.*, 2018.

1.2 Photosynthesis

Photosynthesis is an essential biological process in which energy derived from light is transformed into chemical energy (Fig. 4). This energy is consequently utilized by autotrophic organisms for assimilation of CO₂ into biomass. With the exceptions of deep-sea hydrothermal vents and subsurface ecosystems, photosynthetic organisms facilitate the primary production in every habitat on Earth.

Photosynthesis has been found in seven bacterial phyla, *i.e.* Cyanobacteria, Proteobacteria, Chlorobi, Chloroflexi (Pierson and Castenholz, 1974), Firmicutes (Gest and Favinger, 1983), Acidobacteria (Bryant *et al.*, 2007), and a recently identified phylum Gemmatimonadetes (Zeng *et al.*, 2014). These seven phyla comprise only a tiny fraction of the total number of phyla within the Bacteria domain (Yarza *et al.*, 2008; Söhngen *et al.*, 2013). Since photosynthesis is assumed to be a fundamental and therefore an ancient biological process, some scientists hypothesize there still remain some bacterial phototrophic clades undiscovered (Zeng *et al.*, 2014).

Photosynthetic prokaryotes can be classified into two groups based on the type of photosynthesis they conduct. Oxygenic phototrophs perform photosynthesis during which water is used as the electron donor and oxygen is released as a side product (Fig. 4). To date, the only known prokaryotes capable of oxygenic photosynthesis are cyanobacteria. They belong to the major primary producers on Earth (Waterbury *et al.*, 1979; Chisholm *et al.*, 1988) and many of their representatives significantly contribute to the global nitrogen cycle (Zehr *et al.*, 2001; Karl *et al.*, 2002). Oxygenic photosynthesis is performed also by eukaryotes containing chloroplasts, algae and higher plants. Oxygen released by oxygenic phototrophs enabled the development of advanced life forms by providing a ubiquitous terminal oxidant for respiration (Xiong and Bauer, 2002) as well as formation of the ozone layer protecting Earth against the harmful UV radiation (Blankenship 1992). The second group, anoxygenic phototrophs (Fig. 5), represent some of the oldest bacterial species on Earth. They conduct so called anoxygenic photosynthesis, which refers to the fact it is not accompanied with oxygen evolution and it uses electron donors other than water (Bryant and Frigaard, 2006).

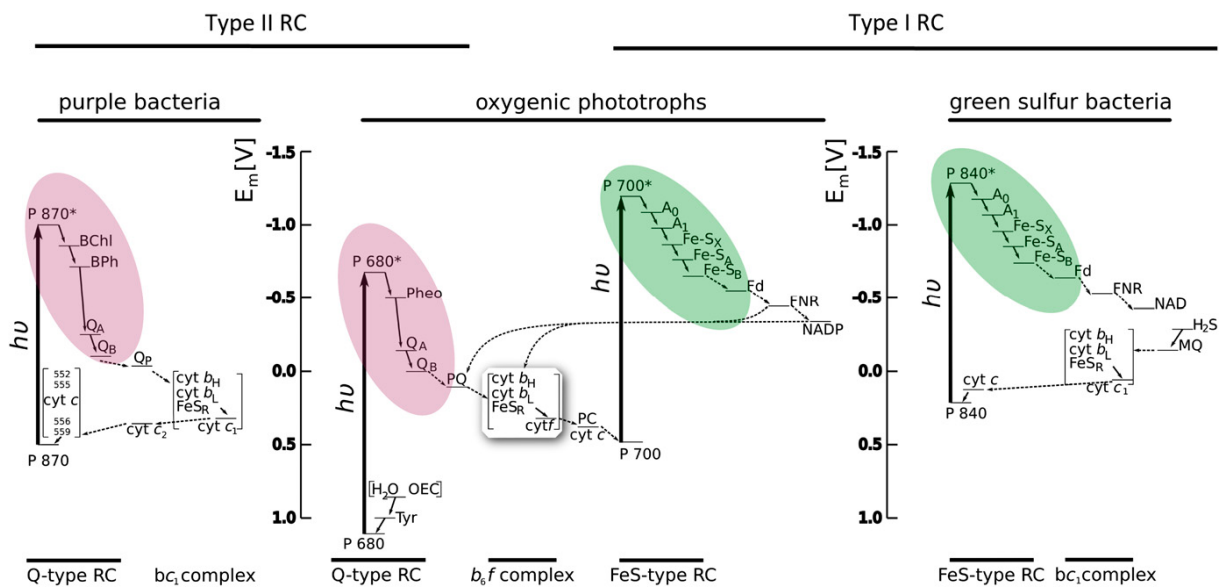


Figure 4 Diagram of the Z-scheme for different groups of photosynthetic organisms. Pink and green colours highlight the electron acceptor portion of the reaction center. Pink, reaction center type I; green, reaction center type II. Adapted from Blankenship 2010.



Figure 5 Anoxygenic phototrophic bacteria. From left to right: *Sphingomonas* sp. AAP5, *Erythrobacter* sp. NAP1, *Roseococcus* sp., *Roseobacter denitrificans*. Figure courtesy of Jason Dean.

1.2.1 Evolution of photosynthesis

The first chemical reactions catalyzed by light energy were abiotic. Abiotic photoreactions produced the first organic molecules on Earth from CO₂ and H₂O and utilized energy of solar UV radiation. These molecules were necessary for the formation of the first cells (Zhang *et al.*, 2007; Mulkidjanian and Galperin, 2009; Mulkidjanian *et al.*, 2012). Under solar radiation

that contained an essential UV component and in the presence of high levels of CO₂ in the primordial atmosphere (Sleep 2010), sulfides of Zn and Mn reduce carbon dioxide to diverse organic molecules (Zhang *et al.*, 2007; Guzman and Martin, 2010). The porous sediments containing ZnS and MnS may served as photosynthesizing habitats of the first phototrophic cells (Mulkidjanian *et al.*, 2012). These ancestral microorganisms may have utilized solar light to assimilate chemical compounds left over from the prebiotic phase of chemical evolution (Zhang *et al.*, 2007).

It is assumed that the first photosynthetic organisms emerged on Earth more than 3.5 billion years ago (Awramik 1992). Biomarker evidence and microfossil finds suggest that cyanobacteria existed 2.5–2.6 billion years ago (Hedges *et al.*, 2001; Xiong and Bauer, 2002). The first photosynthetic organisms originally evolved in a reducing environment, they did not split water and evolve oxygen (Blankenship 2010). Also phylogenetic analyses of bacterial photosynthetic genes suggest that anoxygenic photosynthesis preceded to oxygenic photosynthesis (Xiong *et al.*, 2000; Olson and Blankenship, 2004). Several studies based on the whole-genome comparisons propose that many of the genes coding for core components of photosynthesis were horizontally transferred among various groups of photosynthetic prokaryotes during the evolution of bacterial phototrophy (Raymond *et al.*, 2002; Zeng *et al.*, 2014, Brinkmann *et al.*, 2018).

Many studies based on comparative genomics of chloroplasts and cyanobacteria have shown that photosynthetic eukaryotes acquired photosynthetic traits through endosymbiosis with cyanobacteria (Morden *et al.*, 1992; Gray 1993). Recent evidence places the origin of eukaryotic photosynthetic microorganisms (and chloroplasts) between 1 and 2 billion years ago (McFadden and van Dooren, 2004; Sánchez-Baracaldo *et al.*, 2017). The evolution of photosynthesis from cyanobacteria to chloroplasts caused many structural changes, but the basic principle of oxygenic photosynthesis remained essentially unchanged (Olson and Blankenship, 2004).

1.2.2 Great oxygenation

During the Archaean period the Earth's atmosphere was mostly anoxic. At the beginning of the Proterozoic era, cyanobacteria started to evolve oxygen (Bekker *et al.*, 2004; Holland 2006). Their world-wide spreading approximately 2.3 billion years ago vastly influenced the geochemistry of Earth and evolution of other organisms (Flannery and Walter, 2012). Through

most of the Proterozoic era, the concentration of oxygen in the Earth's atmosphere stayed relatively constant.

The second rise of oxygen level took place during the Neoproterozoic era (Fig. 6). It seems that during this time Earth went through several significant changes (Shields-Zhou and Och, 2011). The supercontinent Rodinia started to fragment. The increased organic carbon burial during the Neoproterozoic era likely facilitated the rise of oxygen concentration in the Earth's atmosphere. Several large glaciations probably occurred because of the concomitant decline in CO₂ levels (Hoffman *et al.*, 1998).

Further, the rising oxygen level led to the massive and irreversible oxidation of the ocean by the end of the Neoproterozoic. Such dramatic environmental changes put a strong selection pressure on the genuine anaerobic anoxygenic phototrophs. When the competition for environments suitable for anaerobic anoxygenic process became too hard, the facultatively anaerobic PNSB were constrained to develop. It is assumed that the first step was the emergence of heterotrophic metabolic pathways. These metabolic processes allowed them to be independent of phototrophy and consequently occupy new environments. In these habitats, the new developing species disposed of their energetically demanding photosynthetic apparatus, which resulted in the responsible genes being lost. *Proteobacteria* are assumed to have differed from a phototrophic ancestor, according to the diffuse distribution of phototrophy throughout the *Proteobacteria* clade (Woese 1987; Swingley *et al.*, 2008). This hypothesis was proposed by Woese (1987) based on 16S rRNA sequence data. Phylogenetic analysis indicates that the loss of photosynthesis genes has taken place independently in several lineages (Keppen *et al.*, 2013; Koblížek *et al.*, 2013; Brinkmann *et al.*, 2018). The newly emerged non-phototrophs then may have become photoheterotrophs by receiving photosynthetic genes through HGT (Jiang *et al.*, 2009; Yurkov and Csotonyi 2009; Brinkmann *et al.*, 2018). This hypothesis was supported by the fact that a HGT of the entire photosynthesis gene cluster (PGC) between distant bacterial phyla has already been documented (Zeng *et al.*, 2014).

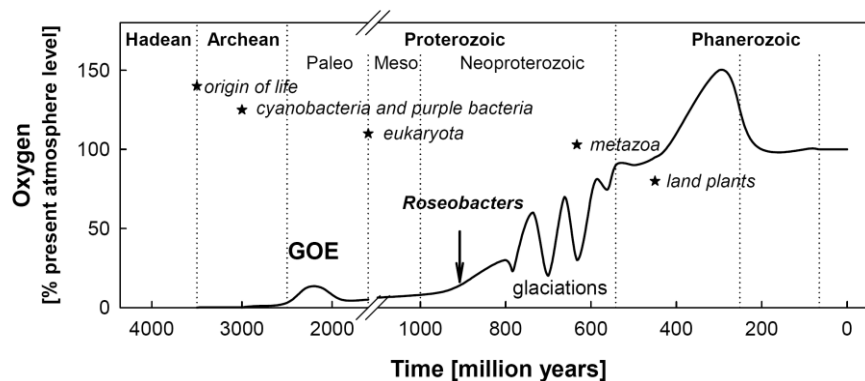


Figure 6 Scheme showing evolution of life on Earth in context to major geological events. Stars represent individual evolutionary events. The bold vertical arrow marks origin of the Roseobacter group. The bold curve indicates the estimated oxygen concentration in atmosphere. GOE, Great Oxidation Event. Reprinted with permission from Koblížek *et al.*, 2013.

1.3 Aerobic anoxygenic phototrophs

Aerobic anoxygenic phototrophic (AAP) bacteria are a functional group of phototrophic organisms which carry out anoxygenic photosynthesis under aerobic conditions (Yurkov and Beatty, 1998). The first AAP bacteria *Erythrobacter longus* and *Roseobacter litoralis* were isolated from the Bay of Tokyo in the 1970s (Shiba *et al.*, 1979). As their name implies, AAP bacteria do not evolve oxygen, because they use electron donors other than water. They are photoheterotrophs; *i.e.* organisms which need a supply of organic substrates for growth, but they also derive a significant amount of their energy requirements from light (Hauruseu and Koblížek, 2012). Light energy is transformed into an electrochemical gradient that can be used for the production of ATP (Yurkov and Beatty, 1998). This ability to utilize light energy likely provides an ecological advantage over their heterotrophic bacterial relatives (Beatty 2002).

The AAP bacteria belong to a phylogenetically diverse genera within *Proteobacteria*. They are closely related to anoxygenic phototrophic purple non-sulfur bacteria as well as other chemotrophic species (Yurkov 2006). AAP bacteria can be considered as a transition between photosynthetic and chemotrophic bacteria, because their ability to utilize light energy can cover only a fraction of the total cellular energy requirements (Hauruseu and Koblížek, 2012). The majority of described AAP species belongs to *Alphaproteobacteria*. There are several AAP species in *Betaproteobacteria* (Yurkov and Csotonyi, 2003; Kasalický *et al.*, 2018), and also several *gammaproteobacterial* representatives (Fuchs *et al.*, 2007).

With an increasing interest in AAP bacteria, it has been found that this group of microorganisms is widely distributed across various environments (Saitoh *et al.*, 1998; Koblížek *et al.*, 2003; Mašín *et al.*, 2008; Yurkov and Csotonyi, 2009) including extreme habitats (Wakao *et al.*, 1996; Suzuki *et al.*, 1999; Yurkov *et al.*, 1999; Rainey *et al.*, 2003). AAP bacteria seem to significantly contribute to the bacterial production in the oligotrophic oceans (Koblížek *et al.*, 2007), but their ecological role remains unclear.

1.3.1 Photosynthetic apparatus

The photosynthetic apparatus of AAP bacteria is supposed to be analogous to that of purple non-sulfur bacteria. It consists of a type II (quinone-type) RC with three structural subunits (L, M and H), four BChls, two bacteriopheophytins, two ubiquinones, a non-heme high-spin Fe^{2+} , and a carotenoid. The photosynthetic RC catalyzes light-induced electron processes leading to a stable charge separation. In summary, it converts solar energy into the biochemical energy (Lancaster and Michel, 1996).

The main light-absorbing pigments in AAP bacteria are BChl *a* and carotenoids. The pigments are bound to special membrane proteins called LH complexes. The LH complexes absorb light, and the energy goes through the pigments of the antenna system to the RC (Cogdell *et al.*, 1996).

All known AAP bacteria synthesize BChl *a* (containing either Mg or Zn) esterified with phytol, but its cellular levels are approximately ten times lower than in purple non-sulfur bacteria (Yurkov and Csotonyi, 2003; Biebl *et al.*, 2005). AAP bacteria produce various species specific carotenoids determining the colour of the microorganism (Yurkov and Beatty, 1998). Moreover many AAP species contain carotenoids that do not serve as a light-harvesting pigment for photosynthesis and its essential function is to protect the photosynthetic apparatus against damaging photochemical reactions (Yurkov *et al.*, 1992; Koblížek *et al.*, 2003). When the BChl is illuminated, triplet-excited BChl molecules are formed; they can react with molecular O_2 to produce singlet oxygen molecule. This singlet form of oxygen is a powerful oxidising agent and rapidly kills bacterial cells. Some kinds of carotenoid pigments are able to protect the photosynthetic apparatus by directly quenching both triplet-excited BChl and singlet molecules of oxygen (Fraser *et al.*, 2001).

BChl synthesis in AAP bacteria is mainly regulated by light (Yurkov and Csotonyi, 2003) and by the presence of O_2 (Suyama *et al.*, 2002). Light represses the BChl synthesis (Yurkov and van Gemerden, 1993). This light repression is more pronounced than in

anaerobic phototrophs (Yurkov and Csotonyi, 2003). On the other hand, the presence of O₂ is required for the BChl synthesis (Suyama *et al.*, 2002). Some AAP species also regulate their pigment expression depending on nutrients, temperature, pH and salinity (Alarico *et al.*, 2002; Koblížek *et al.*, 2003; Rathgeber *et al.*, 2004; Macián *et al.*, 2005; Biebl *et al.*, 2006). Many studies have shown, that AAP bacteria incline to upregulate their pigment synthesis under suboptimal conditions (Alarico *et al.*, 2002; Koblížek *et al.*, 2003; Rathgeber *et al.*, 2004).

1.3.2 Photosynthesis gene cluster

The anoxygenic photosynthetic apparatus consists of a number of genes organised in a so called photosynthesis gene cluster (PGC). This 40-50 kb gene conglomeration has a mosaic structure and consists of five main sets of genes: *bch* and *crt* genes encoding enzymes of the BChl *a* and carotenoids biosynthetic pathways, *puf* operons encoding proteins forming the RC, *puh* operons encoding an RC protein and also responsible for assembly of the RC, and various regulatory genes (Fig. 7). The PGCs of various phototrophic *Proteobacteria* have two conserved subclusters: *crt–bchCXYZ–puf* (ca. 10 kb) and *bchFNBHLM–lhaA–puh* (ca. 12–15 kb) (Waidner and Kirchman, 2005; Liotenberg *et al.*, 2008). PGC contains a core set of 27 photosynthesis genes. Sixteen of them (*bchBCDFGHILMNOPXYZ* and *acsF*) are involved in the BChl *a* biosynthesis. With an exception of 8-vinyl reductase, these genes cover the complete biosynthetic pathway from protoporphyrin XI to BChl *a*. There are 3-8 genes responsible for carotenoid synthesis. The rest of the genes encodes the proteins *pufABLM*, *puhA*, and assembly factors *puhBCE* and *lhaA* of the bacterial photosynthetic units (Zheng *et al.*, 2011).

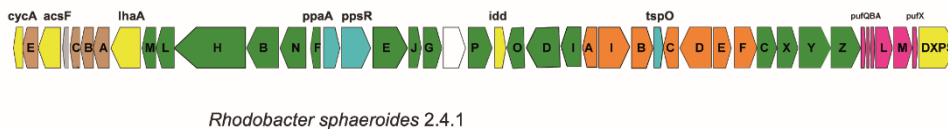


Figure 7 Photosynthesis gene cluster of the purple non-sulfur bacterium *Rhodobacter sphaeroides*. Green, *bch* genes; orange, *crt* genes; pink, *puf* genes; brown, *puh* genes; blue, regulatory proteins; yellow, other photosynthetic genes; grey, hypothetical protein; white, uncertain gene.

1.4 Microbial life in soda lakes

Soda lakes (Fig. 8) are natural water reservoirs characterized by high pH (mostly between 9 – 12) and high concentrations of carbonates and other dissolved salts. They can be found in arid and semiarid areas with high evaporation which leads to high concentration of salts. The majority of life in soda lakes is formed by haloalkaliphilic prokaryotes, which adapted their physiology to these harsh conditions (Sorokin *et al.*, 2014). These organisms often show low genetic similarity to freshwater microbial species (Surakasi *et al.*, 2010; Xiong *et al.*, 2012).

On the other hand, due to the high load of nutrients, soda lakes are often very productive environments. The application of culture-independent techniques documented that the species richness (number of species present) of microorganisms in some soda lakes is often higher when compared with freshwater lakes (Wang *et al.*, 2011). The biomass production is tightly connected with the flow of energy and elements between abiotic and biotic parts of the soda lake system. The carbon cycle is initiated by photosynthetic organisms, which utilize light energy to assimilate inorganic carbon and produce organic molecules available to other organisms. The main photosynthetic organisms in soda lakes are Cyanobacteria, which are responsible for most of the primary production. In addition, there exist also a number of organisms conducting anoxygenic photosynthesis: purple sulfur bacteria, purple non-sulfur bacteria and AAP bacteria. The produced organic carbon is utilized by organotrophic species belonging to the major phyla *Proteobacteria*, *Bacteroidetes*, *Firmicutes*, *Planctomycetes*, *Actinobacteria* etc. (Grant 2006; Lanzén *et al.*, 2013). An important component of soda lake carbon cycle is methane. Methane is produced by methanogenic bacteria which use one-carbon compounds (C1) to gain energy. They are frequently found in soda lake sediments include archaeal genera *Methanocaldococcus*, *Methanohalobium*, *Methanosarcina* and *Methanohalobium* (Grant 2006; Antony *et al.*, 2012). Produced methane can be consequently utilized by methane-oxidizing bacteria such as *Methylobacter* and *Methylomicrobium* (Grant 2006).

The productivity of soda lake crucially depends on the supply of bioavailable nitrogen, which is frequently the limiting nutrient (Carini and Joye, 2008). It seems that some nitrogen cycle pathways (nitrogen fixation, nitrification) known from freshwater and marine environments are not common in hypersaline environments (Sorokin 1998; Oren 1999). Indeed, cyanobacterial species prevalently occurring in soda lakes such as *Arthrospira* are probably not able to fix nitrogen (Grant 2006). One of the nitrogen-containing waste products of biomass decomposition is ammonia. On the other hand, part of the ammonium is utilized

by ammonia-oxidizing bacteria (such as *Thaumarchaea*) which eventually leads to partial loss of bioavailable nitrogen from the system in the form of NO_2 gas (Stevens *et al.*, 1998).

Another important elemental cycle in soda lakes is the cycle of sulfur. Sulfur exists in the environment in the -II to +VI oxidation state. The cycle of sulfur mostly reflects the balance of action of sulfur-reducing bacteria such as *Desulfonatronovibrio* and *Desulfonatronum* and the activity of sulfur-oxidizing species are the genera *Thioalkalivibrio*, *Thiorhodospira*, *Thioalkalimicrobium* and *Natronohydrogenobacter* (Grant 2006). The sulfite or thiosulfate is also required as electron donors for anoxygenic photosynthetic bacteria (Brune 1989; Brune 1995). The high load of organic matter from dead biomass in the bottom part of the lakes frequently leads to anoxia and high concentration of sulfide (Grant 2006).



Figure 8 Saline soda lake in Barguzin Valley, Buryat Republic, Russia. Figure courtesy of Dr. E.N. Boldareva-Nuyanzhina.

1.5 Objectives

The general aim of my PhD thesis was to investigate the evolution of phototrophy in the bacterial order *Rhodobacterales* with a special regard to its haloalkaliphilic representatives from the *Rhodobacter-Rhodobaca* subgroup. For this purpose we used the strain *Rhodobaca barguzinensis* alga05 as an organism of choice.

The initial questions of my work were:

- What is the phylogenetic relationship between *Rca. barguzinensis* and other groups of anoxygenic phototrophic bacteria?
- How has *Rca. barguzinensis* adapted to its environment?
- Is there any specific (marker) gene in genomes of facultatively alkaliphilic *Rhodobacterales*, which is crucial for their adaptation to high pH?
- Is phototrophy an ancestral trait in the studied subgroup of haloalkaliphilic *Rhodobacterales*?

2. Publications

2.1 Paper 1.

The complete genome sequence of *Rhodobaca barguzinensis* alga05 (DSM 19920) documents its adaptation for life in soda lakes.

Soda lakes, characterized by a high pH and high concentrations of carbonates and other dissolved salts, represent a very challenging and often also rapidly changing environment for life. Microorganisms living in these habitats have had to evolve in a way enabling them to dwell under such a tough conditions (Sorokin *et al.*, 2014). Photosynthetic prokaryotes in soda lakes involve also aerobic anoxygenic phototrophs (AAP). These photoheterotrophic organisms transform solar radiation into metabolic energy, but are not able to assimilate inorganic carbon, hence they do not belong among primary producers in soda lakes.

The first article is focussed on the haloalkaliphilic photoheterotrophic bacterium *Rhodobaca (Rca.) barguzinensis* strain alga05. Previous studies suggested that *Rca. barguzinensis* is phylogenetically placed between purple non-sulfur bacteria and AAP species (Keppen *et al.*, 2013). However, this organism is not capable of autotrophic carbon fixation (Boldareva *et al.*, 2008). In the first publication we used single-molecule real time sequencing (SMRT) technology in order to close the *Rca. barguzinensis* genome. The complete genome was analyzed to identify the potential for specific physiological and metabolic adaptations of this organism to the challenging conditions of soda lakes. Another aim of this study was to identify genes responsible for the adaptation to high pH. To fulfill this goal we compared gene inventory in the alga05 genome with genomes of 17 reference strains of *Rhodobacterales*.

Analysis of the genome suggests a large flexibility of the alga05 metabolism. In conclusion, the genetic inventory of *Rca. barguzinensis* corresponds to the adaptations necessary for life in rapidly changing conditions in soda lakes. Further, results of our study shows that the *mrpB* gene coding for the B subunit of the MRP Na⁺/H⁺ antiporter is probably the only gene necessary for the adaptation of haloalkaliphilic *Rhodobacterales* to high pH in soda lakes.



The complete genome sequence of *Rhodobaca barguzinensis* alga05 (DSM 19920) documents its adaptation for life in soda lakes

Karel Kopejtko^{1,2} · Jürgen Tomasch³ · Boyke Bunk⁴ · Cathrin Spröer⁴ · Irene Wagner-Döbler⁵ · Michal Koblížek^{1,2}

Received: 25 April 2018 / Accepted: 9 July 2018 / Published online: 18 July 2018
© Springer Japan KK, part of Springer Nature 2018

Abstract

Soda lakes, with their high salinity and high pH, pose a very challenging environment for life. Microorganisms living in these harsh conditions have had to adapt their physiology and gene inventory. Therefore, we analyzed the complete genome of the haloalkaliphilic photoheterotrophic bacterium *Rhodobaca barguzinensis* strain alga05. It consists of a 3,899,419 bp circular chromosome with 3624 predicted coding sequences. In contrast to most of *Rhodobacterales*, this strain lacks any extrachromosomal elements. To identify the genes responsible for adaptation to high pH, we compared the gene inventory in the alga05 genome with genomes of 17 reference strains belonging to order *Rhodobacterales*. We found that all haloalkaliphilic strains contain the *mrpB* gene coding for the B subunit of the MRP Na⁺/H⁺ antiporter, while this gene is absent in all non-alkaliphilic strains, which indicates its importance for adaptation to high pH. Further analysis showed that alga05 requires organic carbon sources for growth, but it also contains genes encoding the ethylmalonyl-CoA pathway for CO₂ fixation. Remarkable is the genetic potential to utilize organophosphorus compounds as a source of phosphorus. In summary, its genetic inventory indicates a large flexibility of the alga05 metabolism, which is advantageous in rapidly changing environmental conditions in soda lakes.

Keywords Element cycles · Genome annotation · Haloalkaliphiles · *Rhodobacterales* · Soda lake

Communicated by A. Oren.

Data deposition: The genome sequence was deposited at NCBI GenBank under the accession number: CP024899.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00792-018-1041-8>) contains supplementary material, which is available to authorized users.

✉ Michal Koblížek
koblizek@alga.cz

¹ Laboratory of Anoxygenic Phototrophs, Center Algatech, Institute of Microbiology CAS, Třeboň, Czech Republic

² Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

³ Department of Molecular Bacteriology, Helmholtz Centre for Infection Research, Braunschweig, Germany

⁴ Department of Microbial Ecology and Diversity Research, Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

⁵ Institute of Microbiology, Braunschweig University of Technology, Braunschweig, Germany

Introduction

Soda lakes are natural water reservoirs characterized by a high pH (mostly between 9 and 11) and high concentrations of carbonates and other dissolved salts. They can be found in arid and semiarid areas with high evaporation, which leads to the high concentrations of salts. The high salinity and high pH represent very challenging conditions for life. The majority of organisms found in soda lakes are haloalkaliphilic prokaryotes, which have adapted their physiology to these harsh conditions (Sorokin et al. 2014). These organisms usually show very little genomic similarity to microbial freshwater species (Surakasi et al. 2010; Xiong et al. 2012).

Due to the high nutrient load, soda lakes are often very productive. The application of culture-independent techniques documented that the species richness (number of species present) of microorganisms in some soda lakes is often higher than in freshwater lakes (Wang et al. 2011). The biomass production is tightly connected with the flow of energy and elements between abiotic and biotic parts of the soda lake system. The carbon cycle is initiated by photosynthetic organisms, which utilize light energy to assimilate inorganic

carbon and produce organic molecules which in turn become available to other organisms. Aerobic anoxygenic phototrophs (AAP) are a separate group of phototrophic organisms, which capture light for their metabolism, but are unable to fix carbon. The photoheterotrophic bacterium *Rhodobaca* (*Rca.*) was isolated from a small soda lake in the Barguzin valley in south-eastern Siberia (Boldareva et al. 2008). Its detailed characterization documented that this organism contains a bacterial photosynthetic apparatus, which is assembled under both aerobic and anaerobic conditions. However, this organism is not capable of carbon fixation (Boldareva et al. 2008). We have recently reported a draft sequence of its genome (Kopejtko et al. 2017). Our analysis suggested that *Rca. barguzinensis* is phylogenetically placed between purple non-sulfur bacteria and AAP species.

In the presented paper, we used single-molecule real-time sequencing (SMRT) technology in order to close the *Rca. barguzinensis* genome. The complete genome was analyzed to identify the potential for specific physiological and metabolic adaptations of this organism to the challenging conditions of soda lakes.

Materials and methods

Bacterial strains and cultivation conditions

Rca. barguzinensis strain alga05 (DSM 19920) was grown under aerobic conditions in the presence of light at 28 °C in 250 mL Erlenmeyer flasks with 100 mL of medium under shaking (150 rpm). The media used were as described in Sorokin et al. (2000) and Boldareva et al. (2008). The purity of the culture was checked by repeated streaking of the individual colonies formed on 2% agar media.

Genome sequencing

DNA extraction

45 mL of dense culture was harvested by centrifugation at 13,000×g for 5 min. The genomic DNA was extracted and purified using a commercial kit TIANamp Genomic DNA Kit (TIANGEN Biotech (Beijing) Co., Ltd., China). DNA quantity and quality were determined using a NanoDrop ND-1000. The identity of the cultures was verified by amplification and sequencing of their 16S rRNA gene.

Illumina library preparation and sequencing

The Illumina whole-genome shotgun sequencing was performed using the Illumina HiSeq 2000 platform at Macrogen (Seoul, South Korea). Procedures for DNA shearing, library preparation and quality control, sample loading, and

sequencer operation were performed according to the Macrogen's standard protocols (<http://dna.macrogen.com>).

PacBio library preparation and sequencing

SMRTbell™ template library was prepared according to the instructions from Pacific Biosciences, Menlo Park, CA, USA, following the Procedure & Checklist—Greater Than 10 kb Template Preparation. Briefly, for preparation of 15 kb libraries, 8 µg of genomic DNA was sheared using g-tubes™ from Covaris, Woburn, MA, USA according to the manufacturer's instructions. DNA was end repaired and ligated overnight to hairpin adapters applying components from the DNA/Polymerase Binding Kit P6 from Pacific Biosciences, Menlo Park, CA, USA. Reactions were carried out according to the instructions of the manufacturer. BluePippin™ Size-Selection to greater than 4 kb was performed according to the manufacturer's instructions (Sage Science, Beverly, MA, USA). Conditions for annealing sequencing primers and binding of polymerase to purified SMRTbell™ template were assessed with the Calculator in RS Remote, Pacific Biosciences, Menlo Park, CA, USA. SMRT sequencing was carried out on a PacBio RSII (Pacific Biosciences, Menlo Park, CA, USA) taking one 240-min movie for one SMRT cell using the P6 Chemistry. Sequencing resulted in 52,405 post-filtered reads with a mean read length of 12,764 bp.

Genome assembly, error correction

SMRT Cell data were assembled using the “RS_HGAP_Assembly.3” protocol included in SMRT Portal version 2.3.0 using default parameters. The assembly revealed one circular chromosome. The chromosome was circularized and, particularly, artificial redundancies at the ends of the contigs were removed and adjusted to *dnaA* as the first gene. Error correction was performed by a mapping of 7 million paired-end Illumina reads of 2×100 bp onto finished genomes using BWA (Li and Durbin 2009) with subsequent variant and consensus calling using VarScan (Koboldt et al. 2012). A consensus concordance of QV60 could be confirmed for the genome.

Annotation

Finally, an annotation was carried out using NCBI Prokaryotic Genome Annotation Pipeline (PGAP, released 2013, http://www.ncbi.nlm.nih.gov/genome/annotation_prok/). The genome sequence was deposited in the NCBI GenBank under Accession Number CP024899. Methylation analysis was performed using the “RS_Modification_and_Motif_Analysis.1” protocol included in SMRT Portal version 2.3.0.

For metabolic pathway predictions, we combined the automatic annotations provided by the Kyoto Encyclopedia

of Genes and Genomes (Kanehisa et al. 2016) with manual searches using BLAST (Altschul et al. 1990).

Genomic analyses

16S rRNA gene sequences obtained from NCBI GenBank™ were aligned using ClustalX version 2.1. Ambiguously aligned regions and gaps were manually excluded from further analyses. The 16S rRNA tree was constructed by PhyML/MEGA 6.0 software using maximum-likelihood (ML) algorithm with HKY85 nucleotide substitution model and bootstrap 1000×. *Rhodospirillum rubrum* was used as an outgroup.

For the MrpA tree, amino acid sequences were retrieved from GenBank™ and aligned using ClustalX version 2.1. The sites containing gaps and ambiguously aligned regions were stripped from the final alignment manually. The phylogenetic tree was inferred by MEGA 6.0 software using the maximum-likelihood (ML) algorithm with the LG model and bootstrap 1000×. *Agrobacterium tumefaciens* was used as an outgroup.

Comparative genomic analysis

Protein-coding sequences (CDS) of *Rca. barguzinensis* were blasted against CDS from 17 sequenced *Rhodobacterales* strains (Fig. 1; NCBI Genome IDs: 49837, 50480, 57336, 49809, 1658, 49411, 39779, 16178, 41971, 39778, 1096, 509, 40652, 49663, 49965, 44411, 32275) and vice versa. Bidirectional best BLAST hits were obtained using a custom python script downloaded from <http://scriptomika.wordpress.com/2014/01/28/extract-best-reciprocal-blast-match-es/>. Bidirectional best BLAST hits were only considered in further analysis if *e* values were below 10^{-10} .

Origin and terminus of replication were predicted using the web-based tool Ori-finder (Gao and Zhang 2008).

Results and discussion

Genome characteristics

The genome of *Rca. barguzinensis* alga05 consists of one circular chromosome (3,899,419 bp) (Fig. 2). The chromosome contains three copies of rRNA operons (5S, 16S, 23S), 47 tRNA genes, and 3624 protein-coding sequences (CDS). Overall, GC content is 59% (Table 1). In contrast to many other members of *Rhodobacterales*, it lacks any extrachromosomal elements such as plasmids or chromids. The majority of the protein-encoding genes (75.5%) were assigned to one of the known COG categories (Supp. Table S1).

pH homeostasis

To analyze how *Rca. barguzinensis* adapted to high pH conditions, we performed a comparative genomic analysis of 17 representative *Rhodobacterales* species (Fig. 1). The analyzed set of genomes was divided into two groups: alkaliphilic species (with pH optimum > 8.5) and non-alkaliphilic ones. Both groups were queried for trait-specific genes. The analysis (for details see “Materials and methods”) revealed *mrpB* (BG454_08265) to be present in all the genomes of alkaliphilic strains and absent in the others.

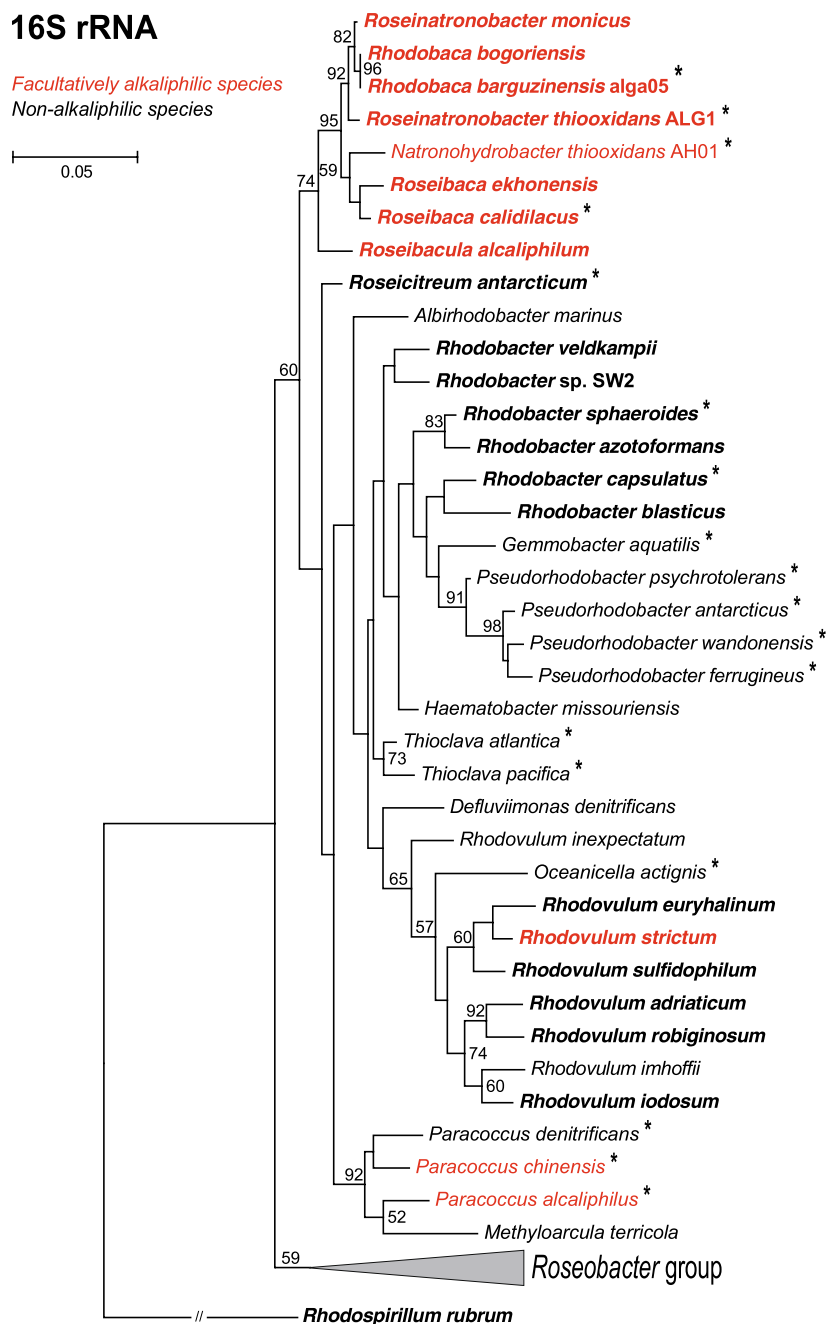
The gene is a part of the *mrp* operon encoding the monovalent cation/proton antiporter system. It has been shown before that *mrp* may contain either seven (*mrpA-G*) or six (*mrpA, mrpC-G*) genes (Swartz et al. 2005), but all the seven genes are required to form a heterooligomeric complex and to facilitate alkaliphily (Kajiyama et al. 2007; Morino et al. 2010). We identified the complete *mrp* operon containing all seven genes in the alga05 genome (Supp. Table S2). The other genes of the operon such as *mrpA* (BG454_08270) or *mrpF* (G454_08245) were present in all 18 analyzed genomes. In the phylogenetic tree founded on the alignment of amino acid sequences of the MrpA protein, almost all (facultatively) alkaliphilic strains (with exception of *P. chinensis*) clearly cluster into Group 1. Group 2 encompasses non-alkaliphilic strains missing the *mrpB* gene in the *mrp* operon (Fig. 3). All this documents that the acquisition of the *mrpB* gene facilitated adaptation of *Rca. barguzinensis* to alkaliphilic conditions.

Salinity

Soda lakes are characterized not only by high pH but also by high salinity. Moreover, in shallow soda lakes mineral content largely fluctuates depending on the balance between freshwater input from rain or melting snow, and evaporation. Therefore, haloalkaliphilic bacteria had to develop several strategies to cope with a high degree of osmotic stress caused by high concentrations of salts (Oren 2011; Sorokin et al. 2014). Their membranes are typically stable and poorly permeable to protons and ions (van de Vossenberg et al. 1999). To deal with osmotic stress and to regulate intracellular pH, bacteria use several groups of transmembrane antiporter proteins. Unfortunately, to date, there is only one study focussed on the characterization of an antiporter from a haloalkaliphilic bacterium isolated from a soda lake (Liu et al. 2005).

Based on the BLAST search using already characterized antiporter proteins, we identified four putative homologs of sodium/proton antiporters in the genome of alga05 (BG454_02710, BG454_05195, BG454_05975, BG454_08255). Based on the published literature (Mesbah et al. 2009; Fujisawa et al. 2005; Yang et al. 2006), we

Fig. 1 16S rRNA phylogenetic tree showing the distribution of facultatively alkaliphilic representatives (highlighted in red) within *Rhodobacterales*. Strains used for the comparative genomic analysis are marked with an asterisk. Phototrophic strains are in bold. The phylogenetic tree was calculated using maximum-likelihood algorithm with HKY85 nucleotide substitution model and bootstrap $\times 1000$. *Rhodospirillum rubrum* was used as an outgroup. Scale bar represents changes per position. Bootstrap values $> 50\%$ are shown



suppose that these proteins, together with the Mrp-proteins (see previous paragraph), are involved in cellular osmotic regulation.

Another strategy to deal with osmotic stress is to synthesize osmoprotectants. Organic compounds such as glycine betaine, glutamine, proline, ectoine, and hydroxyectoine have been found to act as osmoprotectants (Hoffmann et al.

2012; Sorokin et al. 2013). We identified genes coding for an ATP-binding cassette (ABC) transporter for both proline and glycine betaine (BG454_07245, BG454_07250, BG454_07255), as well as for key enzymes of proline (BG454_08430), glycine betaine (BG454_08855) and ectoine (BG454_10030, BG454_10035, BG454_10040) biosynthetic pathways.

Fig. 2 The *Rhodobaca barguzinensis* alga05 genome. Circular representation of the *Rca. barguzinensis* alga05 3.90 Mb chromosome. The outer to inner rings represent: scale of genome size in Mb; features of the chromosome [origin (*oriC*) and terminus (*ter*) of replication]; position of the PGC (in purple); position of ORFs encoded on the plus/minus strand (in blue); GC content (in black/gray); GC skew (in black/gray)

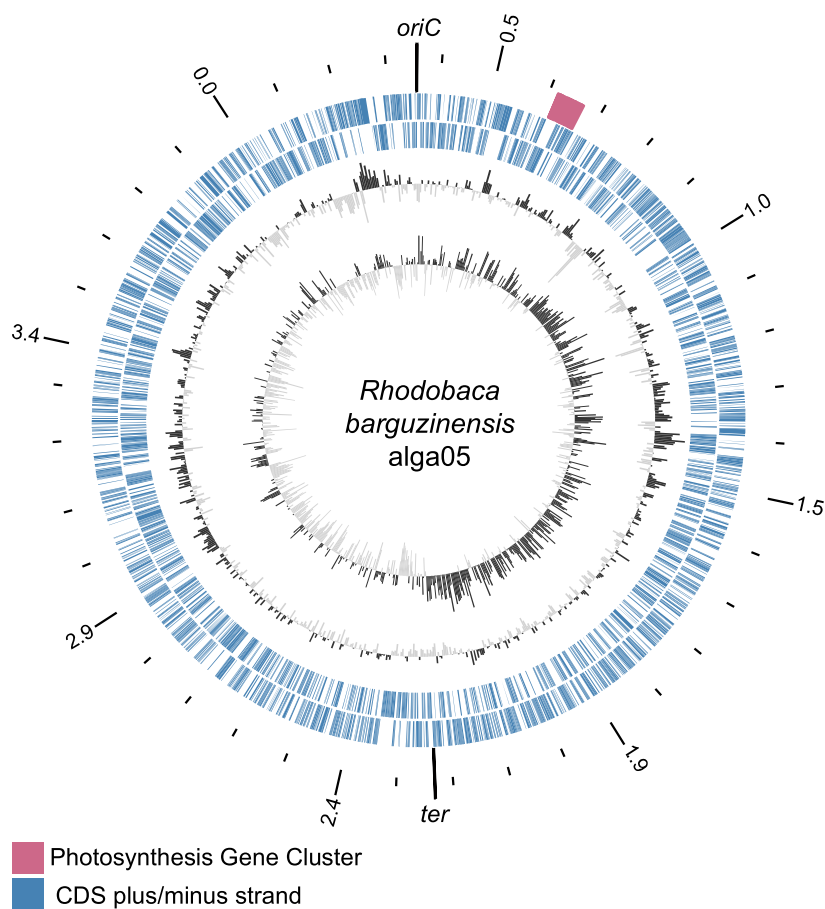


Table 1 General characteristics of *Rca. barguzinensis* genome after NCBI PGAP

Attribute	Value
Total bases	3,899,419
No. of contigs	1
GC content (%)	59
No. of RNAs	59
ncRNAs	3
tRNAs	47
Chromosomes	1
Plasmids	None
Genes (total)	3734
CDS	3624 (97%)
Hypo. proteins	700 (18%)
Pseudogenes	51
PGC length (kbp)	45.4

Energy and carbon metabolism

Alga05's genome contains a standard set of genes associated with the tricarboxylic acid (TCA) cycle and genes

coding for five main protein complexes involved in oxidative phosphorylation. Genes coding for all enzymes involved in the gluconeogenesis/glycolysis (Embden–Meyerhof–Parnas) pathway are present (Table S2). Furthermore, we identified genes coding the ATP-binding cassette (ABC) transporters for amino acids as well as for mono- and oligosaccharides (Fig. 4, Supp. Table S2). Hence, these organic compounds can be taken up into the cell and used as substrates and carbon sources in alga05's energetic metabolism. As documented by Boldareva and colleagues, this metabolism utilizes mainly aerobic respiration and anoxygenic photosynthesis for energy generation (Boldareva et al. 2008).

Five out of the 17 *Rhodobacterales* species, which we used for comparative genomic analysis, possess a complete photosynthetic apparatus. Scattered occurrences of phototrophy among the representatives of the order *Rhodobacterales* (Fig. 1) are one of the characteristic features of this clade. We focussed on the evolutionary aspects of phototrophy among haloalkaliphilic species in our previous paper (Kopejtká et al. 2017).

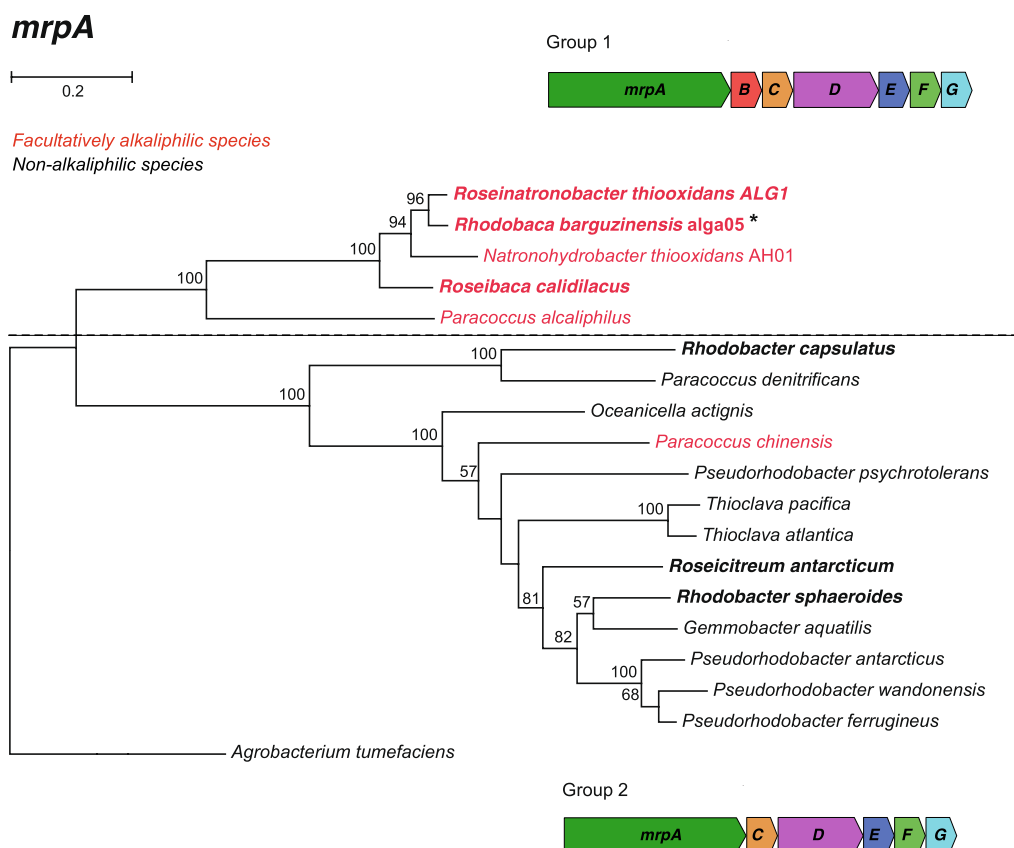


Fig. 3 Phylogenetic tree based on the alignment of amino acid sequences of the MrpA protein (720 common amino acid positions) from representative *Rhodobacteriales* species. The studied strain is marked with an asterisk. Phototrophic strains are in bold. Facultatively alkaliphilic strains are highlighted in red. The phylogenetic tree was calculated using the maximum-likelihood (ML) algorithm with

the LG model and bootstrap $\times 1000$. *Agrobacterium tumefaciens* was used as an outgroup. Scale bar represents changes per position. Bootstrap values $> 50\%$ are shown. The horizontal dashed line divides the strains into two groups (groups 1 and 2) based on different *mrp* operations

Rca. barguzinensis lacks the RuBisCO genes, which is consistent with the absence of carbon fixation in this organism (Boldareva et al. 2008). Recently, it has been shown that the marine AAP bacterium *Dinoroseobacter shibae* employs the ethylmalonyl-CoA (EMC) pathway facilitating CO_2 incorporation. The transition from heterotrophy to photoheterotrophy was accompanied by increased incorporation of fixed carbon into intermediates of the EMC and finally into amino acids (Bill et al. 2017). We identified the presence of all necessary genes involved in the EMC pathway in the genome of alga05 (Fig. 4, Supp. Table S2) as well as in all 17 strains we used for comparative genome analysis (Fig. 1); thus, this pathway seems to be common in *Rhodobacteriales*. To elucidate whether this organism is able to fix carbon to a significant extent and which pathway(s) can potentially be employed for this purpose, additional experiments must be performed.

The carbon cycle in soda lakes also involves the oxidation of methane. In methanotrophic bacteria, this process is initiated by the enzyme methanemmonooxygenase (MMO) (Sazinsky and Lippard 2015). We did not identify any genes coding for MMOs in alga05's genome. None of the 17 *Rhodobacteriales* reference strains belongs among the methanotrophs.

Nitrogen metabolism

The productivity of soda lakes crucially depends on the supply of bioavailable nitrogen, which is frequently the limiting nutrient (Carini and Joye 2008). It seems that some nitrogen cycle pathways (nitrogen fixation, nitrification) known from freshwater and marine environments are not common in hypersaline environments (Sorokin 1998; Oren 1999; Namsaraev et al. 2018). Physiological experiments done by Boldareva and colleagues revealed that under

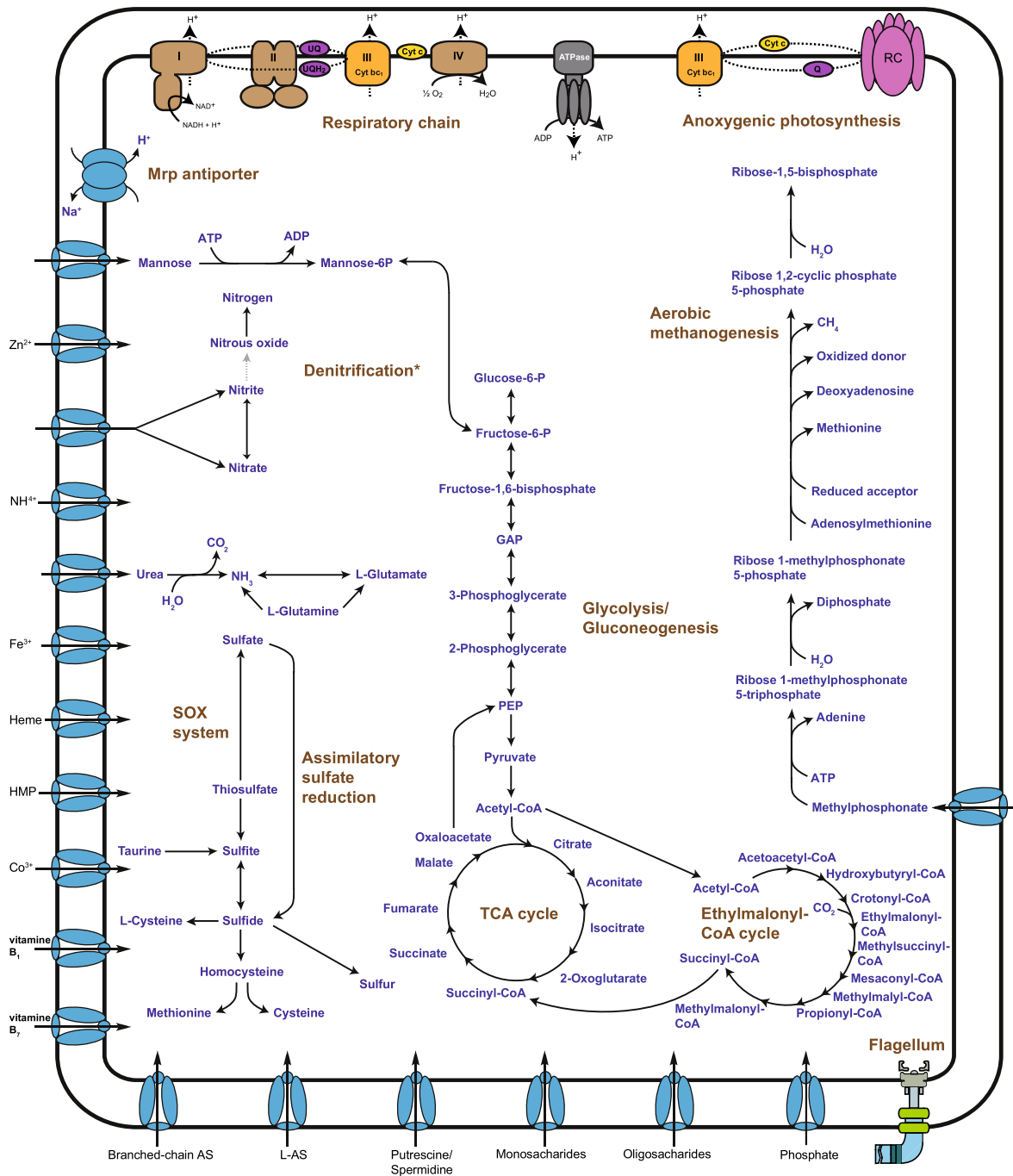


Fig. 4 Reconstruction of metabolic pathways deduced from the genome of *Rca. barguzinensis*. GAP glyceraldehyde-6-phosphate, HMP hydroxymethyl pyrimidine, L-AS L-amino acids, PEP phosphoe-

nolpyruvate. Asterisk indicates denitrification pathway is incomplete; the missing part is highlighted in gray dashed arrow

photoheterotrophic conditions, *Rca. barguzinensis* employs urea, ammonium chloride, glutamate and serine as nitrogen sources (Boldareva et al. 2008).

We did not identify any *nif* genes coding for nitrogenases in alga05's genome, which is consistent with its inability to fix N₂. Also, our comparative analysis documented that none of all 5 other facultatively alkaliphilic strains contain nitrogenase, and among the 12 non-alkaliphiles only two representatives (freshwater photosynthetic bacteria *Rhodobacter sphaeroides* and *Rhodobacter capsulatus*) are nitrogen-fixing bacteria (Table 2).

The reduction of nitrate is possible through two anaerobic respiration pathways resulting in different products. Whereas denitrification yields only the unreactive N₂, dissimilatory nitrate reduction (known as nitrate respiration) produces two bioavailable nitrogen compounds—ammonium and nitrous oxide (N₂O). In the alga05 strain, both nitrate-reducing pathways are incomplete. Indeed, there is a *narGHI* operon present encoding for nitrate reductase enabling the reduction of nitrate to nitrite, but other enzymes reducing further nitrate to nitrous oxide or ammonia are missing (Supp. Table S2). This is consistent with the data of Boldareva et al. (2008), which demonstrated that alga05 is capable of reduction of nitrate to nitrite under anaerobic conditions when illuminated and under aerobic conditions when grown in the dark (Boldareva et al. 2008). Surprisingly, the gene *nosZ*

(BG454_18025), that encodes the terminal enzyme in denitrification for N₂O reductase, is present in alga05 genome (Supp. Table S2). When compared with the five facultatively alkaliphilic reference strains, each anaerobic respiration pathway was present in two representatives. In the group of non-alkaliphilic strains, the distribution of positive genomes for denitrification and dissimilatory nitrate reduction was 5 and 3 out of 12, respectively (Table 2).

Nitrification is an aerobic process involving biotransformation of ammonia to nitrate to gain metabolic energy. Although we identified a gene homolog of the *amoA* gene (BG454_04260) in the genome of alga05 (Supp. Table S2), the genes coding for other subunits of AMO are missing. Also, genes coding for other enzymes oxidizing hydroxylamine further to nitrite and consequently to nitrate are not present in the genome. Only two out of the 17 *Rhodobacterales* species which we used for comparative genomic analysis contain the complete nitrification pathway for oxidation of ammonia up to nitrate. From the group of five facultatively alkaliphilic representatives, only one (*Paracoccus alcaliphilus*) belongs among the nitrifying bacteria. The absence of nitrification among alkaliphilic strains is not surprising as ammonium is frequently a limiting nutrient in soda lakes. Moreover, alga05 can generate energy through photosynthesis, which is much more effective than nitrification.

Table 2 Presence of selected metabolic pathways in the studied organism and other representative facultatively alkaliphilic and non-alkaliphilic *Rhodobacterales* species

	alga05	Facultatively alkaliphilic <i>Rhodobacterales</i> representatives (n = 5)	Non-alkaliphilic <i>Rhodobacterales</i> representatives (n = 12)
Energy metabolism			
EMC pathway (<i>ccrA</i>)	+	5	12
Anoxygenic photosynthesis (<i>pufM</i>)	+	2	3
Nitrogen metabolism			
Nitrogen fixation (<i>nifH</i>)	–	0	2
Dissimilatory nitrate reduction	–	2	5
Denitrification	–	2	3
Nitrification	–	1	1
Sulfur metabolism			
Sulfur oxidation (<i>sox</i>)	+	3	8
Sulfide oxidation (<i>sqr</i>)	+	3	9
DMSP lyase (<i>ddl</i>)	+	2	3
Phosphorus metabolism			
C–P lyase (<i>phnJ</i>)	+	2	8
Iron metabolism			
Hemin uptake	+	2	5
Motility			
Flagellar assembly	+	2	4

Numbers indicate how many representatives contain selected enzyme/pathway
+, enzyme/metabolic pathway present; –, enzyme/metabolic pathway absent

Sulfur metabolism

One of the main processes in the microbial sulfur cycle is the oxidation of reduced inorganic sulfur compounds to yield energy for growth. When using thiosulfate as a substrate, the sulfur oxidizing (Sox) multienzyme complex is employed (Friedrich 1997; Kelly et al. 1997). The Sox complex is encoded by 15 genes, 7 of which (*soxA*, *B*, *C*, *D*, *X*, *Y*, and *Z*) are crucial for the ability to catalyze the oxidation of thiosulfate to sulfate without forming free intermediates (Friedrich et al. 2005; Meyer et al. 2007).

Our search confirmed that there is a complete *sox* gene cluster in the *Rca. barguzinensis* genome (Supp. Table S2); thus, thiosulfate could be potentially used for energy generation in this organism. When comparing the occurrence of the *sox* genes among facultatively alkaliphilic versus non-alkaliphilic reference strains, there were three positive out of five strains and eight positive out of 12 strains, respectively (Table 2).

Interestingly, although the genome of *alga05* contains all the *sox* genes necessary for oxidation of this compound, the actual utilization of thiosulfate has yet to be reported. In the marine AAP gammaproteobacterium *Congregibacter litoralis*, it was suggested that the Sox system was adjusted during its evolution to the oxidation of submillimolar concentrations of the thiosulfate and is non-functional at higher concentrations of this compound (Spring 2014). We speculate that in the *alga05* strain the Sox complex might have been optimized in a similar way.

In the purple non-sulfur bacterium *Rhodobacter capsulatus*, sulfide is oxidized into elemental sulfur using the enzyme sulfide:quinone oxidoreductase (SQR) (Schütz et al. 1997). We discovered a candidate gene (BG454_10075) coding for SQR in the genome of *alga05* (Supp. Table S2) as well as among most of the studied *Rhodobacterales*. When compared, it was found in 3 out of 5 haloalkaliphilic strains and 9 out of 12 non-alkaliphilic strains (Table 2).

Dimethylsulfoniopropionate (DMSP) is an important osmoprotectant for organisms in soda lakes (Antony et al. 2010; Oremland 2013). Regarding DMSP catabolism, *alga05* possesses a homolog of the gene *dddL* (BG454_06175) coding for the enzyme DMSP lyase (Curson et al. 2008), which cleaves DMSP into the gas dimethylsulfide (DMS) and the acrylate (Supp. Table S2). The occurrence of this enzyme among the facultatively alkaliphilic *Rhodobacterales* was 2 positive out of 5 strains and in the group of non-alkaliphilic representatives there were 3 positive out of 12 strains (Table 2). The activity of the DMSP lyase has not been experimentally tested in this organism (Boldareva et al. 2008).

Phosphorus metabolism

Phosphorus (P) is an indispensable element playing an important role in cell metabolism. As the availability of phosphorus may be a limiting growth factor, the capacity to use non-conventional phosphorus sources may represent an important competitive advantage. Phosphonates are a group of organophosphorus compounds containing a single carbon-to-phosphorus (C–P) bond. The purple non-sulfur bacterium *Rba. capsulatus* is able to degrade phosphonates using the C–P lyase (Schowanek and Verstraete 1990). This enzyme catalyzes the breakdown of organophosphorus compounds into inorganic phosphate and the relevant hydrocarbon compound (Kononova and Nesmeyanova 2002).

We identified all the genes necessary for the transport of methylphosphonate (MPn) into the cytoplasm and its cleavage to methane and phosphate in the genome of *alga05* (Supp. Table S2). We also identified the gene coding for the enzyme C–P lyase (BG454_02325) in genomes of 10 out of 17 other *Rhodobacterales* strains, which we used for comparative genomic analysis. The occurrence of this enzyme among the facultatively alkaliphilic *Rhodobacterales* reference strains was 2 out of 5 strains and in the group of non-alkaliphilic representatives it was 7 out of 12 strains (Table 2, Supp. Table S4).

We hypothesize that MPn might serve as an (accessory) P source for *Rca. barguzinensis*, since it has all the enzymes necessary for degrading it to phosphate and methane. This selective advantage can be even more pronounced in the case of other non-haloalkaliphilic freshwater *Rhodobacterales* species, since in a freshwater non-hypersaline environment P is a limiting nutrient (Elser et al. 2007). Soda lakes are anomalous in this aspect; they have high concentrations of dissolved inorganic P. It has already been shown that some freshwater lakes contain up to 10% of dissolved organic P (DOP) in the form of phosphonates (Cheesman et al. 2012; Coupe et al. 2012), thus phosphonates as substrates are potentially available in these ecosystems. Further, it was shown that the phosphonate degradation pathway is common in freshwater lakes (Yao et al. 2016). Hence, after the depletion of dissolved phosphates, P-starved bacteria may acquire P from phosphonates.

Iron metabolism

Iron serves as a biocatalyst for many enzymes and regulatory proteins. *Alga05* genome analysis led us to the identification of two operons coding for proteins involved in two different iron uptake mechanisms (Supp. Table S2). *Alga05* can acquire iron ions from the environment in ferric form (Fe^{3+}) or bound in porphyrins. The first mechanism encompasses transport of Fe^{3+} ions chelated to siderophores into the cytoplasm. In the second iron-uptake pathway, a heme-binding

protein (hemin) is transported into the cell. Whereas the first mechanism utilizes free iron ions present in the environment, the hemin-uptake system enables the transfer of Fe^{3+} ions in the form of heme released from lysed cells.

We found homologs of genes coding for the siderophore-mediated iron uptake in all reference genomes. When compared the occurrence of the hemin-uptake system among facultatively alkaliphilic versus non-alkaliphilic reference strains, there were 2 positive out of 5 strains and 5 positive out of 12 strains, respectively (Table 2).

Vitamins and cofactors and their synthesis

We found complete biosynthetic pathways for thiamine (vit. B1), riboflavin (vit. B2), niacin (vit. B3), pyridoxin (vit. B6), cobalamin (vit. B12), and retinol (vit. A1). No known biosynthetic pathway for biotin (vit. B7) was found. Moreover, an ABC transporter for this crucial vitamin was detected (Supp. Table S2) indicating that *alga05* requires an external supply of biotin. We identified complete pathways for the cofactors folate, pyridoxal phosphate, heme, lipoic acid, molybdenum cofactor, coenzyme A and pantothenate. Transporters for thiamine and heme were also present (Supp. Table S2).

Motility

For some bacteria, motility is an important physiological process (Moens and Vanderleyden 1996). This can be facilitated by the bacterial flagellum, which is a complex rotating structure allowing the cell to swim in a liquid or semisolid environment (Berg 2003). The flagellum is composed of a long helical filament which is attached to a cellular motor embedded in the cytoplasmic membrane through a flexible linker known as the hook.

We found homologs of genes coding for all three main parts of the bacterial flagellum in the genome of *Rca. barguzinensis* (Supp. Table S2). Also, previous morphological observation done by Boldareva et al. (2008) confirmed the motility of individual cells (Boldareva et al. 2008).

Further, we identified genes coding for flagellar proteins in genomes of 6 out of 17 other *Rhodobacterales* strains.

Conclusions

Analysis of the complete genome of *Rca. barguzinensis* revealed only the *mrpB* gene coding for the B subunit of the MRP antiporter as potentially genetic determinant for haloalkaliphilic *Rhodobacterales*. In general, the metabolic strategy indicates that it gains most of its basic nutrients (C, N, P) from organic matter. Noteworthy is the potential to utilize phosphonate as a P-source. Regarding energy sources,

alga05 is very versatile. It is able to live in both aerobic and anaerobic conditions, employing several strategies— aerobic respiration, photophosphorylation and sulfur oxidation under aerobic conditions and denitrification and photosynthesis under anoxic (or semiaerobic) conditions. Overall, the *Rca. barguzinensis* genome reflects the adaptation of the organism to the very dynamic conditions in soda lakes, characterized by high productivity, steep gradients, and rapid changes of oxygen concentration and nutrient availability.

Acknowledgements The authors are beholden to Dr. Katya Boldareva-Nuyanzina for providing her strain *Rca. barguzinensis* strain *alga05*. We also thank Jason Dean B.Sc. for the language revision. We thank Simone Severitt and Nicole Heyer for excellent technical assistance. This research has been supported by the GAČR project P501/12/G055, the DAAD project 57155424, and the MŠMT project Algatech Plus (LO1416). J.T. was supported by the Deutsche Forschungsgemeinschaft (DFG) within the Transregio 51 “Roseobacter”.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Antony CP et al (2010) Active methylotrophs in the sediments of Lonar Lake, a saline and alkaline ecosystem formed by meteor impact. *ISME J* 4:1470–1480
- Berg HC (2003) The rotary motor of bacterial flagella. *Annu Rev Biochem* 72:19–54
- Bill N et al (2017) Fixation of CO_2 using the ethylmalonyl-CoA pathway in the photoheterotrophic marine bacterium *Dinoroseobacter shibae*. *Environ Microbiol* 19:2645–2660
- Boldareva EN et al (2008) *Rhodobaca barguzinensis* sp. nov., a new alkaliphilic purple nonsulfur bacterium isolated from a soda lake of the Barguzin Valley (Buryat Republic, Eastern Siberia). *Microbiology* 77:206–218
- Carini SA, Joye SB (2008) Nitrification in Mono Lake, California: activity and community composition during contrasting hydrological regimes. *Limnol Oceanogr* 53:2546–2557
- Cheesman AW, Turner BL, Ramesh Reddy K (2012) Soil phosphorus forms along a strong nutrient gradient in a tropical ombrotrophic wetland. *Soil Sci Soc Am J* 76:1496–1506
- Coupe RH, Kalkhoff SJ, Capel PD, Gregoire C (2012) Fate and transport of glyphosate and aminomethylphosphonic acid in surface waters of agricultural basins. *Pest Manag Sci* 68:16–30
- Curson ARJ, Rogers R, Todd JD, Brearley CA, Johnston AWB (2008) Molecular genetic analysis of a dimethylsulfoniopropionate lyase that liberates the climate-changing gas dimethylsulfide in several marine α -proteobacteria and *Rhodobacter sphaeroides*. *Environ Microbiol* 10:757–767
- Elser JJ et al (2007) Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecol Lett* 10:1135–1142
- Friedrich CG (1997) Physiology and genetics of sulfur-oxidizing bacteria. *Adv Microb Physiol* 39:235–289
- Friedrich CG, Bardischewsky F, Rother D, Quentmeier A, Fischer J (2005) Prokaryotic sulfur oxidation. *Curr Opin Microbiol* 8:253–259
- Fujisawa M, Kusumoto A, Wada Y, Tsuchiya T, Ito M (2005) NhaK, a novel monovalent cation/H⁺ antiporter of *Bacillus subtilis*. *Arch Microbiol* 183:411–420

- Gao F, Zhang CT (2008) Ori-finder: a web-based system for finding oriC s in unannotated bacterial genomes. *BMC Bioinform* 9:79
- Hoffmann T et al (2012) Synthesis, release, and recapture of compatible solute proline by osmotically stressed *Bacillus subtilis* cells. *Appl Environ Microbiol* 78:5753–5762
- Kajiyama Y, Otagiri M, Sekiguchi J, Kosono S, Kudo T (2007) Complex formation by the *mrpABCDEF* gene products, which constitute a principal Na⁺/H⁺ antiporter in *Bacillus subtilis*. *J Bacteriol* 189:7511–7514
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462
- Kelly DP, Shergill JK, Lu WP, Wood AP (1997) Oxidative metabolism of inorganic sulfur compounds by bacteria. *Antonie Van Leeuwenhoek* 1:95–107
- Koboldt DC et al (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22:568–576
- Kononova SV, Nesmeyanova MA (2002) Phosphonates and their degradation by microorganisms. *Biochemistry (Moscow)* 67:184–195
- Kopejka K et al (2017) Genomic analysis of the evolution of phototrophy among haloalkaliphilic *Rhodobacterales*. *Genome Biol Evol* 9:1950–1962
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760
- Liu J et al (2005) The activity profile of the NhaD-type Na⁺ (Li⁺)/H⁺ antiporter from the soda lake haloalkaliphile *Alkalimonas amylolytica* is adaptive for the extreme environment. *J Bacteriol* 187:7589–7595
- Mesbah NM, Cook GM, Wiegel J (2009) The halophilic alkalithermophile *Natranaerobius thermophilus* adapts to multiple environmental extremes using a large repertoire of Na⁺ (K⁺)/H⁺ antiporters. *Mol Microbiol* 74:270–281
- Meyer B, Imhoff JF, Kuever J (2007) Molecular analysis of the distribution and phylogeny of the soxB gene among sulfur-oxidizing bacteria—evolution of the Sox sulfur oxidation enzyme system. *Environ Microbiol* 9:2957–2977
- Moens S, Vanderleyden J (1996) Functions of bacterial flagella. *Crit Rev Microbiol* 22:67–100
- Morino M et al (2010) Single site mutations in the hetero-oligomeric Mrp antiporter from alkaliphilic *Bacillus pseudofirmus* OF4 that affect Na⁺/H⁺ antiporter activity, sodium exclusion, individual Mrp protein levels, or Mrp complex formation. *J Biol Chem* 285:30942–30950
- Namsaraev Z et al (2018) Effect of salinity on diazotrophic activity and microbial composition of phototrophic communities from Bitter-1 soda lake (Kulunda Steppe, Russia). *Extremophiles* 22:651–663
- Oremland RS (2013) A random biogeochemical walk into three soda lakes of the western USA: with an introduction to a few of their microbial denizens. *Polyextremophiles. Cellular Origin. Life in Extreme Habitats and Astrobiology*. Springer, Dordrecht, Netherlands, pp 179–199
- Oren A (1999) Bioenergetic aspects of halophilism. *Microbiol Mol Biol Rev* 63:334–348
- Oren A (2011) Thermodynamic limits to microbial life at high salt concentrations. *Environ Microbiol* 13:1908–1923
- Sazinsky MH, Lippard SJ (2015) Methane monooxygenase: functionalizing methane at iron and copper. *Sustaining life on planet earth: Metalloenzymes mastering dioxygen and other chewy gases. Metal Ions in Life Sciences*. Springer, Cham, pp 205–256
- Schowaneck D, Verstraete W (1990) Phosphonate utilization by bacterial cultures and enrichments from environmental samples. *Appl Environ Microbiol* 56:895–903
- Schütz M, Shahak Y, Padan E, Hauska G (1997) Sulfide-quinone reductase from *Rhodobacter capsulatus* purification, cloning, and expression. *J Biol Chem* 272:9890–9894
- Sorokin DY (1998) Occurrence of nitrification in extremely alkaline natural habitats. *Microbiology* 67:404–408
- Sorokin DY et al (2013) Halophilic and haloalkaliphilic sulfur-oxidizing bacteria. *The prokaryotes*. Springer, Berlin, Heidelberg, pp 529–554
- Sorokin DY, Turova TP, Kuznetsov BB, Briantseva IA, Gorlenko VM (2000) *Roseinatronobacter thiooxidans* gen. nov., sp. nov., a new alkaliphilic aerobic bacteriochlorophyll- α -containing bacteria from a soda lake. *Mikrobiologiya* 69:89–97
- Sorokin DY et al (2014) Microbial diversity and biogeochemical cycling in soda lakes. *Extremophiles* 18:791–809
- Spring S (2014) Function and evolution of the sox multienzyme complex in the marine gammaproteobacterium *Congregibacter litoralis*. *ISRN Microbiol* 2014:597418. <https://doi.org/10.1155/2014/597418>
- Surakasi VP, Antony CP, Sharma S, Patole MS, Shouche YS (2010) Temporal bacterial diversity and detection of putative methanotrophs in surface mats of Lonar crater lake. *J Basic Microbiol* 50:465–474
- Swartz TH, Ikewada S, Ishikawa O, Ito M, Krulwich TA (2005) The Mrp system: a giant among monovalent cation/proton antiporters? *Extremophiles* 9:345–354
- van de Vossenberg JL, Driessen AJ, Grant D, Konings WN (1999) Lipid membranes from halophilic and alkali-halophilic Archaea have a low H⁺ and Na⁺ permeability at high salt concentration. *Extremophiles* 3:253–257
- Wang J et al (2011) Do patterns of bacterial diversity along salinity gradients differ from those observed for macroorganisms? *PLoS One* 6:e27597
- Xiong J et al (2012) Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Environ Microbiol* 14:2457–2466
- Yang LF et al (2006) A Na⁺/H⁺ antiporter gene of the moderately halophilic bacterium *Halobacillus dabanensis* D-8T: cloning and molecular characterization. *FEMS Microbiol Lett* 255:89–95
- Yao M, Henny C, Maresca JA (2016) Freshwater bacteria release methane as a by-product of phosphorus acquisition. *Appl Environ Microbiol* 82:6994–7003

Supplementary information

The online version of this article (<https://doi.org/10.1007/s00792-018-1041-8>) contains supplementary material, which is available to authorized users.

2.2 Paper 2.

Genomic analysis of the evolution of phototrophy among haloalkaliphilic *Rhodobacterales*.

Evolution of a prokaryotic genome is driven by three fundamental processes: random mutations, gain of genes *via* HGT, and gene loss. Consequent natural selection decides if the newly coded trait persists in the lineage. In bacteria, for functional photosynthetic apparatus, usually a range of specialized genes is needed. Moreover, these genes are often organized in operons and this clustering in operons is usually crucial for their optimal expression. Hence, for evolution of phototrophy inside a certain bacterial clade, only the latter two listed processes are substantial.

The order *Rhodobacterales* (Alphaproteobacteria) contains both photoautotrophic and photoheterotrophic species, these phototrophic species are phylogenetically mixed with chemotrophs (Simon *et al.*, 2017). Representatives of this clade perform essential metabolic processes such as anaerobic fermentation, aerobic respiration, autotrophic carbon fixation, nitrogen fixation, sulfur oxidation, or hydrogen production (Garrity *et al.*, 2005; Androga *et al.*, 2012). In context of such a variability, the question how their trophic strategies evolved occurs.

In the second publication, we studied the evolution of phototrophy among the haloalkaliphilic members of the *Rhodobacter-Rhodobaca* (*RR*) group inside the *Rhodobacterales* clade. We tested two possible evolutionary scenarios (1) the “regressive evolution” scenario, where the ancestors of the haloalkaliphilic *RR*-group were photoautotrophic organisms, which later lost part (or all) of their photosynthesis genes, or (2) the ancestors of the *RR*-group were heterotrophs, which adopted their photosynthesis genes via HGT of photosynthesis genes. To test the two proposed evolutionary scenarios, we analysed the presence of photosynthesis genes, their organization and phylogeny in three closely related representatives with different phototrophic capacities and oxygen preferences. Our analyses show a general trend in the reduction of photosynthetic apparatus (loss of photosynthesis genes), hence a shift towards (photo)heterotrophic lifestyle. Interestingly, a singlet oxygen defense mechanism, which is common in phototrophic species, was found in the heterotrophic representative. In conclusion, these findings indicate that the ancestors of the haloalkaliphilic members of the *RR*-group were phototrophic species.

Genomic Analysis of the Evolution of Phototrophy among Haloalkaliphilic *Rhodobacterales*

Karel Kopejtko^{1,2}, Jürgen Tomasch³, Yonghui Zeng⁴, Martin Tichý¹, Dimitry Y. Sorokin^{5,6}, and Michal Koblížek^{1,2,*}

¹Laboratory of Anoxygenic Phototrophs, Institute of Microbiology, CAS, Center Algatech, Treboň, Czech Republic

²Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

³Research Group Microbial Communication, Helmholtz Centre for Infection Research, Braunschweig, Germany

⁴Aarhus Institute of Advanced Studies, Aarhus, Denmark

⁵Winogradsky Institute of Microbiology, Research Centre of Biotechnology, Russian Academy of Sciences, Moscow, Russia

⁶Department of Biotechnology, Delft University of Technology, The Netherlands

*Corresponding author: E-mail: koblizek@alga.cz.

Accepted: July 26, 2017

Data deposition: This project has been deposited at NCBI GenBank under the accession numbers: GCA_001870665.1, GCA_001870675.1, GCA_001884735.1.

Abstract

A characteristic feature of the order *Rhodobacterales* is the presence of a large number of photoautotrophic and photoheterotrophic species containing bacteriochlorophyll. Interestingly, these phototrophic species are phylogenetically mixed with chemotrophs. To better understand the origin of such variability, we sequenced the genomes of three closely related haloalkaliphilic species, differing in their phototrophic capacity and oxygen preference: the photoheterotrophic and facultatively anaerobic bacterium *Rhodobaca barguzinensis*, aerobic photoheterotroph *Roseinatronobacter thiooxidans*, and aerobic heterotrophic bacterium *Natronohydrobacter thiooxidans*. These three haloalkaliphilic species are phylogenetically related and share many common characteristics with the *Rhodobacter* species, forming together the *Rhodobacter-Rhodobaca* (RR) group. A comparative genomic analysis showed close homology of photosynthetic proteins and similarity in photosynthesis gene organization among the investigated phototrophic RR species. On the other hand, *Rhodobaca barguzinensis* and *Roseinatronobacter thiooxidans* lack an inorganic carbon fixation pathway and outer light-harvesting genes. This documents the reduction of their photosynthetic machinery towards a mostly photoheterotrophic lifestyle. Moreover, both phototrophic species contain 5-aminolevulinic synthase (encoded by the *hemA* gene) incorporated into their photosynthesis gene clusters, which seems to be a common feature of all aerobic anoxygenic phototrophic *Alphaproteobacteria*. Interestingly, the *chrR-rpoE* (sigma24) operon, which is part of singlet oxygen defense in phototrophic species, was found in the heterotrophic strain *Natronohydrobacter thiooxidans*. This suggests that this organism evolved from a photoheterotrophic ancestor through the loss of its photosynthesis genes. The overall evolution of phototrophy among the haloalkaliphilic members of the RR group is discussed.

Key words: Anoxygenic photosynthesis, bacteriochlorophyll, horizontal gene transfer, photosynthesis gene cluster, *Rhodobacteraceae*.

Introduction

The order *Rhodobacterales* (*Alphaproteobacteria*) encompasses a highly diverse ensemble of species (Simon et al. 2017). They largely differ in their phenotype, metabolic traits, and ecological niches they inhabit. Its members conduct many fundamental metabolic processes such as aerobic respiration,

anaerobic fermentation, sulfur oxidation, autotrophic carbon fixation, nitrogen fixation, or hydrogen production in various combinations (Garrity et al. 2005; Androga et al. 2012). One of the most interesting features of *Rhodobacterales* is the presence of many species that perform anoxygenic photosynthesis. Members of the *Rhodobacter* (*Rba.*) and *Rhodovulum* (*Rdv.*)

© The Author 2017. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

genera represent classical examples of purple nonsulfur bacteria—organisms capable of photoautotrophic growth under anaerobic conditions utilizing the Calvin cycle, or heterotrophic growth in the presence of oxygen (Androga et al. 2012). On the other hand, there exist many marine photoheterotrophic species belonging to the so called *Roseobacter* group, which grow and conduct photosynthesis in the presence of oxygen (Wagner-Döbler and Biebl 2006; Moran et al. 2007; Brinkhoff et al. 2008), do not contain RuBisCO and use light only as an additional energy source (Koblížek et al. 2013).

In the last 20 years, seven haloalkaliphilic photoheterotrophic species were isolated from several saline and soda lakes in Africa, America, Asia, and Antarctica. Remarkably, these organisms are phylogenetically closer to freshwater *Rhodobacter* species than to marine photoheterotrophs. *Rhodobaca (Rca.) bogoriensis* was isolated from soda lakes in the East African Rift Valley (Milford et al. 2000), and is capable of both aerobic and anaerobic growth on organic substrates, however it does not contain Calvin cycle enzymes and thus cannot grow photoautotrophically. A closely related organism with the same physiology, *Rca. barguzinensis*, was later isolated from a small soda lake in the Barguzin valley in south-eastern Siberia (Boldareva et al. 2008). In contrast, the closely related aerobic anoxygenic phototrophic (AAP) bacterium *Roseinatronobacter (Rna.) thiooxidans*, isolated from a soda lake in the Kunkurskaya steppe (Chita region, south-eastern Siberia) grows solely under aerobic conditions (Sorokin et al. 2000b; Stadnichuk et al. 2009). Later, a very similar species, *Rna. monicus* was isolated from the hypersaline Soda Mono lake in California, USA (Boldareva et al. 2006). Closely related to the above mentioned species are two Antarctic AAP isolates *Roseibaca ekhonensis* and *Roseicitreum antarcticum*, which were collected from Lake Ekho (Labrenz et al. 2009), and sandy intertidal sediments (Yu et al. 2011), respectively. The last described AAP bacterium in this group is *Roseibacula alcaliphilum*, which was isolated from Lake Doroninskoe, East Transbaikal region, Russia (Boldareva and Gorlenko 2014). This photoheterotrophic group is complemented by some non-phototrophic species such as the haloalkaliphilic heterotrophic bacterium *Natronohydrobacter (Nhb.) thiooxidans* isolated from a soda lake in Kenya.

Based on the close phylogenetic proximity of photoautotrophic, photoheterotrophic, and lithoautotrophic species, Keppen et al. (2013) suggested that the photoheterotrophic species such as *Rhodobaca* and *Roseinatronobacter* represent the intermediate species on the regressive evolutionary pathway leading from photoautotrophic purple nonsulfur bacteria to heterotrophic species. A similar scenario has also been proposed for photoheterotrophic and heterotrophic species in the marine *Roseobacter* group (Koblížek et al. 2013). However, the existence of photoheterotrophic species can be explained using an alternative scenario where the originally heterotrophic species acquired their photosynthesis genes *via* horizontal gene transfer (HGT). A typical feature of purple

phototrophic bacteria is that most of their genes involved in the light-phase of photosynthesis are organized in photosynthesis gene clusters (PGCs) (Zsebo and Hearst 1984; Zheng et al. 2011). It has been speculated that clustering of photosynthesis genes in PGCs can facilitate HGT. In line with this hypothesis is the report that the whole PGC was likely transferred between *Proteobacteria* and *Gemmatimonadetes* (Zeng et al. 2014). *Rhodobacterales* species often carry various extra chromosomal elements (Petersen et al. 2013), transposable elements (Vollmers et al. 2013), or gene transfer agents (Luo and Moran 2014), which can facilitate gene transfer. Moreover, PGC-containing plasmids have been identified in some *Roseobacter*-group representatives (Petersen et al. 2013). HGT has also been proposed for *puf* genes encoding the protein subunits of the bacterial reaction centers within the *Roseobacter* group (Koblížek et al. 2015).

To elucidate the most likely evolutionary pathway, we decided to sequence genomes of three closely related (on average 98.7% 16S rRNA pairwise similarity) haloalkaliphilic organisms differing in their metabolic capacities and oxygen preference. Photoheterotrophic *Rca. barguzinensis* strain alga05 exhibits both aerobic and anaerobic growth. It lacks RuBisCO and nitrogenase, but it has the capacity to utilize thiosulfate and sulfide as electron donors (Boldareva et al. 2008). The second organism, *Rna. thiooxidans* strain ALG1, is a photoheterotrophic bacterium combining aerobic respiration and photophosphorylation, but it can also oxidize thiosulfate and sulfide during aerobic lithoheterotrophic growth (Sorokin et al. 2000b). The last organism, *Nhb. thiooxidans* strain AH01, is a strictly aerobic, nonphototrophic bacterium. It is capable of growing lithoheterotrophically with acetate and thiosulfate, sulfide, polysulfide, and elemental sulfur which are oxidized to sulfate, and in addition it can oxidize H₂ (Sorokin et al. 2000a). We mostly focused on the presence of photosynthesis genes, their organization and phylogeny in order to test the two proposed evolutionary scenarios.

Materials and Methods

Bacterial Strains and Cultivation Conditions

Rna. thiooxidans strain ALG1, *Rca. barguzinensis* strain alga05 and *Nhb. thiooxidans* strain AH01 were obtained from the UNIQEM Culture Collection of the Winogradsky Institute of Microbiology, Russian Academy of Sciences. The phototrophic strains were grown under aerobic conditions in the presence of light at 28 °C in 250 ml Erlenmeyer flasks with 100 ml of medium under shaking (150 rpm). The media used was as described in Sorokin et al. (2000b) and Boldareva et al. (2008). The chemotrophic strain AH01 was grown in a 5 l gas bottle with 1 l of acetate-containing medium and with a 20% H₂/80% air gas phase (Sorokin et al. 2000a). The purity of the cultures was controlled by repeated streaking of the individual colonies formed on the medium solidified with 2% (wt/vol) agar and verified by microscopy.

Genome Sequencing and Data Analyses

About 45 ml of dense culture was harvested by centrifugation at $13,000 \times g$ for 5 min. The genomic DNA was extracted and purified using a commercial kit (Tiangen Biotech, Beijing, China). DNA quantity and quality were determined using a NanoDrop ND-1000. The identity of the cultures was verified by amplification and sequencing their 16S rRNA genes. The whole-genome shotgun sequencing was performed using the Illumina HiSeq 2000 platform and a Roche 454 GS FLX+ system at Macrogen (Seoul, South Korea). Procedures for DNA shearing, library preparation and quality control, sample loading, and sequencer operation were performed according to Macrogen's standard protocols (<https://dna.macrogen.com>).

Raw reads produced from both types of whole-genome shotgun sequencing were trimmed and de novo assembled into contigs using CLC Genomics Workbench 9.0 software (QIAGEN Aarhus, Denmark). To decrease the number of final contigs, they were further processed using the "Genome Finishing Module" of the CLC software.

Annotation of contigs was performed with the RAST server (Aziz et al. 2008). The obtained genomic data was compared with other *Rhodobacterales* species [*Rba. sphaeroides* 2.4.1, *Rba. capsulatus* SB 1003, *Rdv. sulfidophilum* DSM 2351, *Paracoccus (Pco.) denitrificans* PD 1222] obtained from the NCBI GenBank.

To detect clustered regularly interspaced short palindromic repeats (CRISPR) in studied genomes, we used the web-based tool CRISPRFinder (Grissa et al. 2007).

Orthologous gene cluster analysis was performed using the OrthoVenn web server (Wang et al. 2015). Pairwise sequence similarities between all input protein sequences were calculated with an *e*-value cut-off of $1e^{-5}$. An inflation value ($-$) of 1.5 was used to define orthologous cluster structure.

Phylogenetic Analyses

16S rRNA gene sequences obtained from GenBank™ were aligned using ClustalX version 2.1. Ambiguously aligned regions and gaps were excluded from further analyses using Gblocks (Talavera and Castresana 2007). Phylogenetic trees based on 16S rRNA gene sequences were constructed using PhyML/MEGA 6.0 software using a maximum likelihood (ML) algorithm with a HKY85 nucleotide substitution model and a bootstrap of $1,000 \times$. For the BchIDHLNB concatenated tree, amino acid sequences were retrieved from GenBank™ and aligned using ClustalX version 2.1. The sites containing gaps and ambiguously aligned regions were stripped from the final alignment using Gblocks. Amino acid sequence alignments of each peptide were concatenated with Geneious version 8.1.2 (Biomatters Ltd.). The 5-aminolevulinic synthase (ALAS) phylogenetic tree involves amino acid sequences derived by BLASTing from ALAS sequences of *Dinoroseobacter (Drb.) shibae* (Dshi_3546, Dshi_1182, and Dshi_3190).

Phylogenetic trees were inferred by MEGA 6.0 software using ML algorithm with LG model. 16S rRNA sequence pairwise similarities were calculated using GGDC tool under the recommended settings (Meier-Kolthoff et al. 2013).

Sequence Accession Numbers

The *Rca. barguzinensis* strain alga05, *Rna. thiooxidans* strain ALG1 and *Nhb. thiooxidans* strain AH01 genome sequences have been deposited at NCBI GenBank under the accession numbers: GCA_001870665.1, GCA_001870675.1, and GCA_001884735.1.

Results

The sequencing and assembly of the genomes of the three studied organisms resulted in draft assemblies consisting of 5–52 contigs (table 1). The smallest genome (~ 3.5 Mb) was found in *Rna. thiooxidans* ALG1, while *Rca. barguzinensis* alga05 and *Nhb. thiooxidans* AH01 had somewhat larger genomes (~ 3.9 and 4.1 Mb). The three strains share 2,240 orthologs which account for $\sim 80\%$ of their genes (supplementary fig. 1, Supplementary Material online).

The 16S rRNA sequences of the three sequenced strains were compared with other members of order *Rhodobacterales*. The obtained phylogenetic tree documents that the sequenced haloalkaliphilic strains are related not only to *Roseibaca*, *Roseicitreum*, and *Roseibacula* species but more distantly also to *Rhodobacter* species *Rba. sphaeroides*, *Rba. capsulatus*, *Rba. sp. SW2*, and *Rba. veldkampii* (fig. 1). We denote this group *Rhodobacter-Rhodobaca* group, or, in shorthand *RR* group. Freshwater photoautotrophic bacterium *Rdv. sulfidophilum*, and *Paracoccus* species, were closely related to the *RR*-group. The marine species belonging to the so-called *Roseobacter* group were phylogenetically more distant (fig. 1).

As expected, a complete PGC was found in alga05 and ALG1 genomes. Its organization was identical in both strains, and it was very similar to the organization of the PGC in *Rhodobacter* species (fig. 2). The only difference was the presence of the *hemA* gene in the haloalkaliphilic strains and the transposition of *tspO-crtBIA-bchIDO* genes. The PGC in *Drb. shibae* and *Rvu. sulfidophilum* has a clearly different organization. Moreover, in *Rvu. sulfidophilum*, it is split in two parts placed ~ 300 kb apart. A typical feature for all phototrophic *RR*-species is the divergent orientation ($\leftarrow \rightarrow$) of superoperons *bchFNBHLM-lhaA-puh* and *crt-bchCXYZ-puf*, while in other *Rhodobacterales* groups they have convergent or colinear orientation (Zheng et al. 2011).

The major difference between the investigated species and *Rhodobacter* species was the complete absence of genes for carbon and nitrogen fixation. We detected a RuBisCO-like protein in the genome of the AH01 strain, but due to low homology (36% amino-acid identity) and absence of the

Table 1

General Features of the Studied Genomes as Compared with the Published Genomes of the Closest Relatives

	<i>Rca. barguzinensis</i> alga05	<i>Rna. thiooxidans</i> ALG1	<i>Nhb. thiooxidans</i> AH01	<i>Rba. sphaeroides</i> 2.4.1	<i>Rdv. sulfidophilum</i> DSM 2351	<i>Rba. capsulatus</i> SB 1003	<i>Pco. denitrificans</i> PD 1222
Total bases	3,893,644	3,504,907	4,075,255	4,628,173	4,732,772	3,871,920	5,236,194
No. of contigs	5	52	39	Closed	Closed	Closed	Closed
GC content [%]	59	60	63	69	67	67	67
No. of RNAs	53	42	43	60	59	66	61
chromosomes	<i>n.a.</i>	<i>n.a.</i>	<i>n.a.</i>	2	1	1	2
plasmids	<i>n.a.</i>	<i>n.a.</i>	<i>n.a.</i>	5	3	1	1
ORFs	3,805	3,385	3,885	4,347	4,398	3,685	5,121
proteins	2,760 (73%)	2,502 (74%)	2,834 (73%)	3,076 (71%)	2,837 (65%)	3,100 (84%)	4,032 (79%)
hyp.proteins	1,045 (27%)	883 (26%)	1,051 (27%)	1,271 (29%)	1,561 (35%)	610 (16%)	1,089 (21%)
PGC length [kbp]	45.4	45.1	–	44.7	42.1	44.3	–
No. of CRISPR spacers	4	38	29	0	11	59	9

other necessary enzymes it does not seem to be involved in carbon fixation. All three studied strains also lack the *pucBAC* operon encoding the outer light harvesting complexes (table 2), which is consistent with the lack of LH2 bands in their absorption spectra (Sorokin et al. 2000a).

Tetrapyrrole Biosynthesis Genes

Bacteriochlorophylls are the main light-harvesting pigments in anoxygenic phototrophs (Willows and Kriegel 2009). The genes responsible for their biosynthesis have been routinely used as suitable phylogenetic markers for the investigation of the origin of photosynthesis in bacteria (Raymond et al. 2002; Koblížek et al. 2013).

The biosynthesis of all tetrapyrrols, including BChl *a*, starts with the synthesis of 5-aminolevulinic acid (Willows and Kriegel 2009; Dailey et al. 2017). This compound is synthesized in all Proteobacteria via the Shemin pathway by 5-aminolevulinic acid synthase (ALAS). In *Rba. sphaeroides* 2.4.1., there exist two homologous isoenzymes coded by the *hemA* and *hemT* genes (Fanica-Gaignier and Clement-Metral 1973; Bolt et al. 1999). Similarly, there were both ALAS genes in the phototrophic strains alga05 and ALG1 (table 2). Only one gene was found (*hemA* form) in the heterotrophic strain AH01, in photoautotrophic bacteria *Rba. capsulatus*, *Rba. sp. SW2*, as well as in *Rvu. sulfidophilum*. To better understand the origin of the multiple ALAS forms, we constructed a maximum-likelihood tree with all ALAS genes found in various members of *Rhodobacterales*. The constructed tree clearly splits into two major groups (fig. 3). The first group encompasses *hemA*-like genes from both heterotrophic and phototrophic species, including the *hemA* gene from *Rba. sphaeroides*. It also contains *hemA* genes from *Rba. capsulatus*, *Rba. sp. SW2*, and *Rvu. sulfidophilum*. The second group contains *hemT* related genes, mostly from phototrophic *Rhodobacterales*. Both, alga05 and ALG1 contain one ALAS gene from the *hemA*

group, and one from the *hemT* group. Noteworthy, the latter is a part of the PGC (fig. 2), which seems to be a universal characteristic of all AAP *Rhodobacterales*. Interestingly, a homologue of this gene was also found in the PGC of AAP bacterium *Erythrobacter* sp. NAP1 (order *Sphingomanadales*), though, here the gene was likely acquired through the HGT.

The tetrapyrrole pathway then proceeds from 5-aminolevulinic acid to protoporphyrin IX by action of six different enzymes. The last two are oxidases, which exist in oxygen-dependent and oxygen-independent forms. The coproporphyrinogen III oxidase is present in both forms (*hemF*, *hemN*) in all studied strains except *Rba. capsulatus*, which contains only the oxygen-independent form of this enzyme. In the case of protoporphyrinogen IX oxidase, all the investigated strains contained only the oxygen-dependent form encoded by the *hemY* gene (table 2).

Bacteriochlorophyll Biosynthesis

The BChl *a* pathway on its own starts with magnesium chelatase. This enzyme catalyzes the insertion of Mg into protoporphyrin IX. It is composed of three subunits encoded by the *bchl*, *bchD*, and *bchH* genes. Another enzyme of the BChl *a* biosynthesis pathway is light-independent protochlorophyllide reductase (subunits encoded by the *bchl*, *bchN*, and *bchB* genes). This protein is required for the light-independent reduction of protochlorophyllide to chlorophyllide *a* (Willows and Kriegel 2009). To investigate the phylogenetic relationship of the photosynthesis genes between studied and reference strains, we calculated the tree using the concatenated amino acid sequences of these two enzymes. The inferred phylogeny documents a very close relationship of the two photoheterotrophic strains, alga05 and ALG1, and all the *RR*-species (fig. 4).

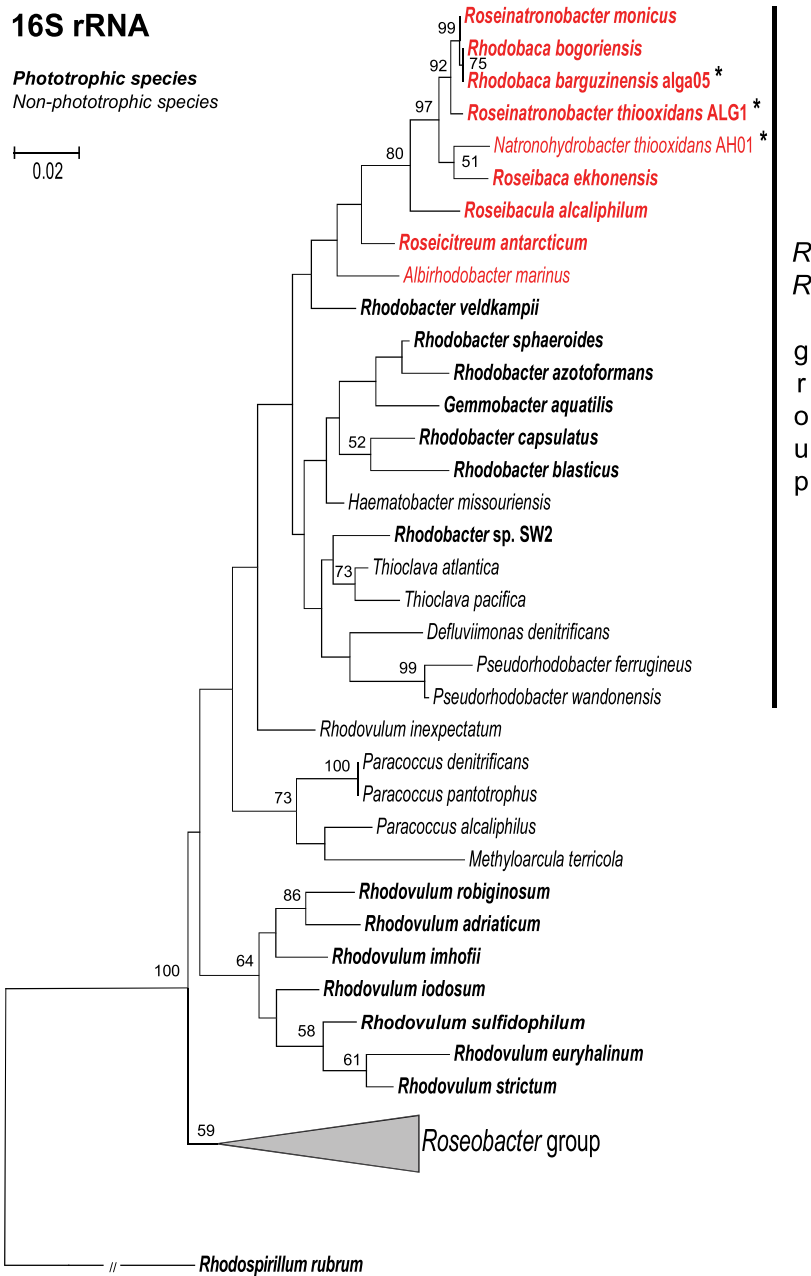


Fig. 1.—16S rRNA phylogenetic tree showing the position of studied strains (marked with an asterisk) within the *Rhodobacterales* clade. Halophilic strains are highlighted in red. The phylogenetic tree was calculated using maximum likelihood algorithm with HKY85 nucleotide substitution model and bootstrap 1,000×. *Rhodospirillum rubrum* was used as an outgroup organism. Scale bar represents changes per position. Bootstrap values >50% are shown.

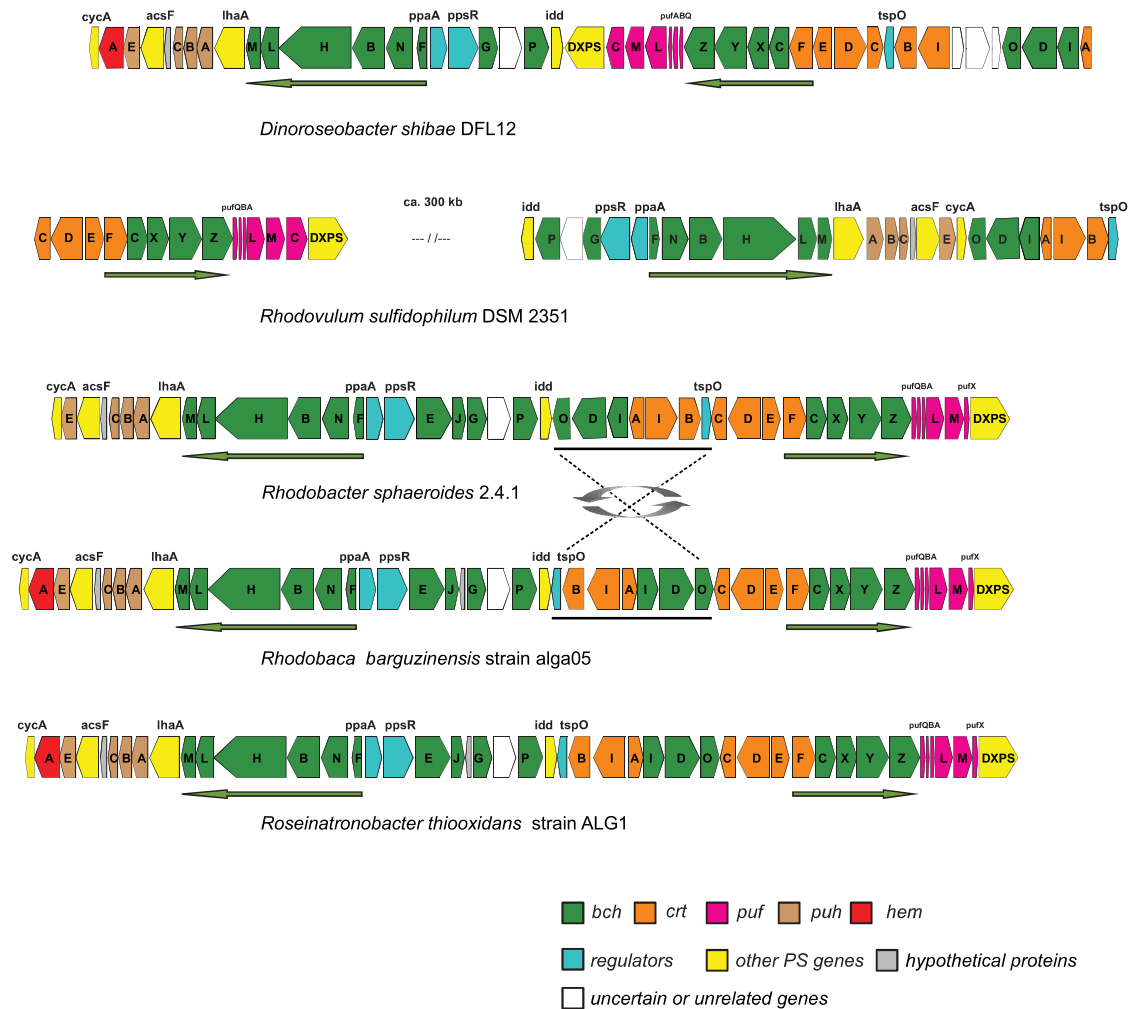


Fig. 2.—Comparison of gene organization in PGCs of the studied and reference genomes. Arrows show directions of superoperons *bchFNBHLM-lhaA-puh* and superoperons *crt-bchCXYZ-puf*. Note: *Rba.* sp. SW2 has the same PGC organization as *Rba. sphaeroides*. PGC in *Rba. capsulatus* differs from *Rba. sphaeroides* only by the absence of the *acsF* gene.

Another important step of the BChl a biosynthetic pathway is the cyclization of its fifth ring by Mg-protoporphyrin IX monomethyl ester oxidative cyclase (Willows and Kriegel 2009). This enzyme has two forms: an oxygen-dependent and oxygen-independent form encoded by *acsF* and *bchE* gene, respectively. *Rba. sphaeroides*, *Rba.* sp. SW2, *Rvu. sulfidophilum*, *alga05* and *ALG1* contain both forms of this enzyme, while *Rba. capsulatus* contains only the oxygen independent form. Heterotrophic strains *AH01* and *Pco. denitrificans* obviously do not contain these genes. Interestingly these two forms of the cyclase have a very different phylogeny. The oxygen-dependent form of the cyclase (*acsF* gene product) was used previously as a convenient gene marker for

the detection of AAPs (Boldareva et al. 2013; Zeng et al. 2014). Indeed, all *RR*-species formed a distinct group, clearly separated from the marine species as well as *Rvu. sulfidophilum* (supplementary fig. 2, Supplementary Material online). The situation was very different for the oxygen-independent form (*bchE* gene product), where in *Rba.* spp., *ALG1*, and *alga05*, the *bchE* gene is part of the PGC, while in *Rvu. sulfidophilum* and in many marine AAP species it is outside the PGC.

Reaction Center Proteins

One part of the PGC comprises of *puf* (photosynthesis unit forming), an operon encoding the proteins for the inner

Table 2
The Presence of Genes Related to Photosynthesis in the Studied and Reference Genomes

	DSM 2351	2.4.1	SB 1003	alga05	ALG1	AH01	PD 1222
Carbon fixation							
RubisCO (large subunit)	○	○	○	–	–	–/?	○
<i>puf</i> genes							
<i>pufC</i>	●	–	–	–	–	–	–
<i>pufX</i>	–	●	●	●	●	–	–
<i>pufH</i> genes							
<i>pufABC, pufE</i>	●	●	●	●	●	–	–
<i>puc</i> genes							
<i>pucBAC</i>	○	○	○	–	–	–	–
Other PS genes							
<i>cycA</i>	●	●	○	●	●	○	○
<i>lhaA</i>	●	●	●	●	●	–	–
<i>DXPS</i>	●○	●○	●○	●○	●○	○	○
Porphyrin biosynthesis							
<i>hemA</i>	○	○○	○	●○	●○	○	○
<i>hemF*</i>	○	○	–	○	○	○	○
<i>hemN</i>	○	○	○	○	○	○	○
<i>hemY*</i>	○	○	○	○	○	○	○
<i>hemG</i>	–	–	–	–	–	–	–
<i>acsF*</i>	●	●	–	●	●	–	–
<i>bchE</i>	○	●	●	●	●	–	–
Regulatory proteins							
<i>ppaA/aerR</i>	●	●	●	●	●	–	–
<i>appA</i>	–	○	–	–	–	–	–
<i>ppsR/crtJ</i>	●	●	●	●	●	–	–
<i>hvrB</i>	–	–	●	–	–	○	–
<i>tspO</i>	●	●	●	●	●	–	–
<i>pufQ</i>	●	●	●	●	●	–	–
Transcription factors							
Sigma factors	○	○	○	○	○	○	○
RpoE, RpoH _I , RpoH _{II}							
Anti-sigma factor ChrR	○	○	–	○	○	○	–

NOTE.—The PS genes present inside the PGC are marked “●”; the PS genes located outside the PGC are marked “○”; “○ ○” means two forms of gene are present; “–” means gene is absent in the genome. Genes coding for the oxygen-dependent form of the enzyme are marked with asterisk. *hemF* and *hemN* are aerobic and anaerobic forms of Coproporphyrinogen III oxidase, *hemY* and *hemG* are aerobic and anaerobic forms of Protoporphyrinogen IX oxidase. *acsF* and *bchE* are aerobic and anaerobic forms of Mg-protoporphyrin IX monomethyl ester oxidative cyclase. ‘/?’ = RubisCO-like protein identified in the genome of AH01.

antenna (*pufBA*) and the photosynthesis reaction center subunits L and M (*pufLM*). In all *Rhodobacter* species (*Rba. sphaeroides*, *Rba. capsulatus*, and *Rba. sp. SW2*) these genes are accompanied by a *pufX* gene encoding the light-harvesting associated protein PufX, with the typical *pufQBALMX* operon organization. This gene was also present in the two phototrophic strains alga05 and ALG1. In contrast, *Rdv. sulfidophilum* contains the *pufC* gene encoding the cytochrome *c* subunit of the complex (table 2).

Since the *pufM* gene represents a convenient genetic marker of anoxygenic phototrophs with type 2 reaction centers, it is frequently used to infer phylogeny of new phototrophic organisms. The computed PufM phylogenetic tree of phototrophic *Rhodobacterales* shows a major split into two branches. The first encompasses *Rhodobacter* species, the phototrophic species alga05 and ALG1, and a few members

of the marine *Roseobacter* group. The second group contains *Rvu. sulfidophilum*, *Roseobacter (Rsb.) denitrificans*, *Drb. shibae* and most of the marine AAP species. This phylogeny further confirms the close relationship of the studied haloalkalophiles with *Rhodobacter* species, and their distinction from *Rhodovulum* and *Roseobacter* species (supplementary fig. 3, Supplementary Material online).

Regulatory Proteins

All photosynthetic organisms control expression of their photosynthesis genes in response to changes in the environment. All seven studied strains contain these common regulators: AhcY, FnrL, HvrA, PrrA/RegA, PrrB/RegB, and PrrC/SenC. All studied phototrophic strains accommodate *ppaA/aerR*, *ppsR/crtJ*, *tspO*, and *pufQ* genes inside their PGCs. *Rba. sphaeroides*

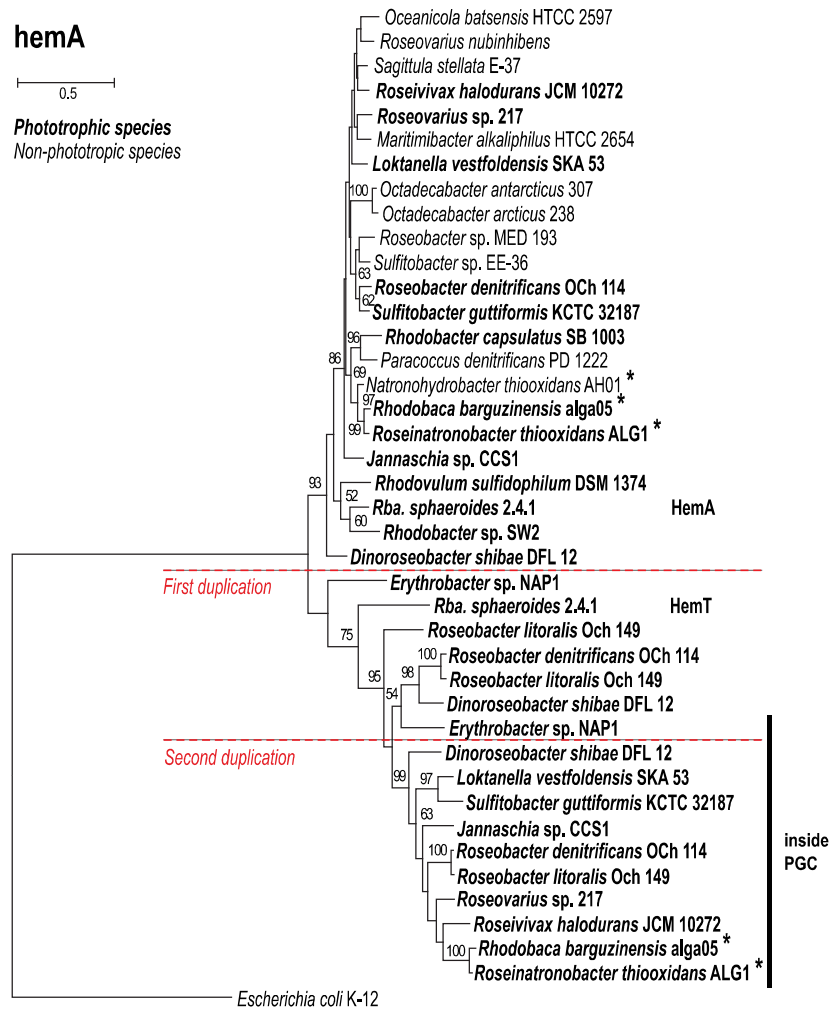


Fig. 3.—Phylogenetic analysis based on alignment of amino acid sequences of ALAS (HemA or HemT). Maximum likelihood (ML) tree with bootstrap 500× was constructed for studied strains and other representative species from the *Rhodobacterales* clade. *Escherichia coli* K-12 was used as an outgroup organisms. Scale bars represent changes per position. Bootstrap values >50% are shown. Horizontal bar marks sequences found inside the PGC. Studied strains are marked with an asterisk. Hypothetical gene duplication events are marked with horizontal dashed lines.

is the only strain with *appA* gene, which is located outside the PGC. Interestingly, *Rba.* sp. SW2 also does not possess this gene in the genome (data not shown). Two investigated strains, that is, phototrophic *Rba. capsulatus* and chemotrophic strain AH01 possess the gene coding for the HvrB protein. The phototrophic strain contains this gene inside its PGC (table 2).

One of the most important physiological processes in phototrophic species is the protection of cells against reactive oxygen species. Singlet oxygen is generated under aerobic conditions by light-excited BChl *a* (Borland et al. 1989).

In *Rba. sphaeroides* three alternative sigma factors are involved in response to the singlet oxygen, the sigma-24 factor RpoE, and the sigma-32 factors RpoH_I and RpoH_{II}. While RpoH is part of a more general oxidative stress response, RpoE and ChrR have been shown to be triggered specifically by the oxygen singlet. In these organisms, RpoE and ChrR are encoded in an operon and together with a second operon transcribed from the same promoter in the opposite direction form a conserved gene cluster. This gene cluster has also been found in all investigated phototrophic species, but it was absent in most of the heterotrophic species. Surprisingly, the

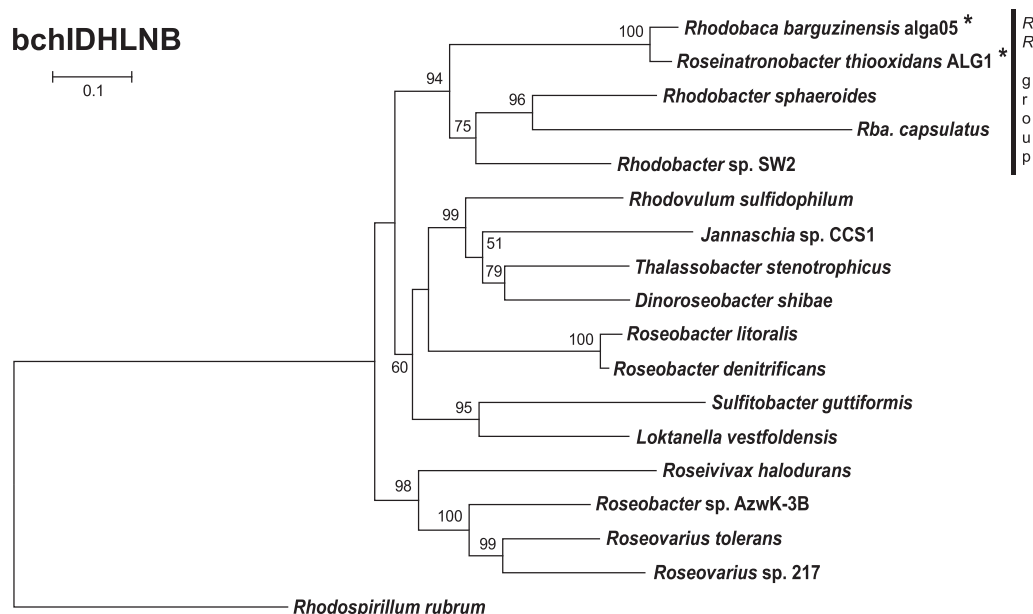


Fig. 4.—Phylogenetic tree based on concatenated alignments of amino acid sequences of magnesium chelatase and light-independent protochlorophyllide reductase (*bchIDHLNB*; 2317 common amino acid positions). Maximum-likelihood (ML) tree with bootstrap 1,000× was constructed for studied strains and other representative species from the *Rhodobacterales* clade. *Rhodospirillum rubrum* was used as an outgroup organism. Scale bars represent changes per position. Bootstrap values >50% are shown. Studied strains are marked with an asterisk.

rpoE-chrR cluster was also present in the studied heterotrophic strain AH01, as well as in several marine heterotrophic bacteria *Maritimibacter alkaliphilus*, *Octadecabacter antarcticus*, *Octadecabacter arcticus*, *Oceanicola batsensis*, and *Sagittula stellata* which should not be prone to singlet oxygen formation (fig. 5).

Discussion

Two evolutionary scenarios explaining the presence of phototrophic and heterotrophic strains among haloalkaliphilic species in the *RR*-group were considered in this study (1) the “regressive evolution” scenario, where the ancestors of the haloalkaliphilic *RR*-group were photoautotrophic organisms, which later lost part (or all) of their photosynthesis gene, or (2) the ancestors of the *RR*-group were heterotrophs, which adopted their photosynthesis genes *via* HGT of photosynthesis genes. Our data indicates that phototrophy is ancestral in the *RR*-group. There are two main lines of evidence. First, all the studied phototrophic organisms belonging to the *RR*-group shared an almost identical organization of their PGCs (fig. 2). A unique feature for all phototrophic *RR*-species is the divergent orientation of superoperons *bchFNBHLM-IhaA-puh* and *crt-bchCXYZ-puf*. This orientation has not been found in any other *Rhodobacterales* species (Zheng et al. 2011), which

represents the main evidence against the hypothesis that the *RR*-species adopted the PGC through HTG from other *Rhodobacterales* genera. Another distinctive feature present in all studied phototrophic members of the *RR*-group is the presence of the *pufX* gene, and the absence of the *pufC* gene (table 2). The gene *pufX* is also present in several marine AAP species, but they have likely received it *via* the HGT as discussed before (Koblížek et al. 2015). Second, the performed phylogenetic analyses on the photosynthesis genes *bchIDHLNB*, *acsF* and *pufM* (fig. 4, supplementary figs. 2 and 3, Supplementary Material online) document that the *RR*-group always cluster together, with an exception of the *pufM* tree where the *RR*-group is mixed with several marine AAP species (supplementary fig. 3, Supplementary Material online).

An important finding is the presence of genes connected with the singlet oxygen defense system in chemotrophic strain AH01. This strain contains the same singlet-oxygen-stress-response mechanism (encoded in *rpoE-chrR* gene operon) as *Rba. sphaeroides* (Anthony et al. 2004), as well as in ALG1 and alga05 strains. When bound to its anti-sigma factor ChrR, RpoE is inactive in the cell. The presence of singlet oxygen leads to a dissociation of the heterodimer, thus the activation of RpoE and subsequently its more than 180 target genes (Anthony et al. 2004). A similar response has been

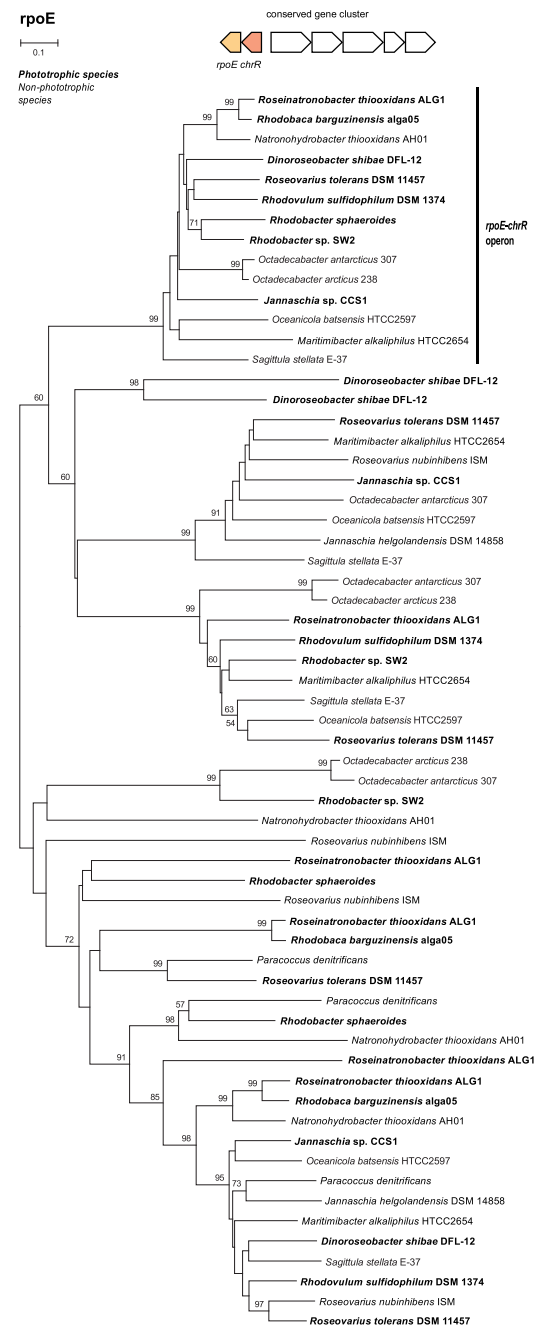


Fig. 5.—Phylogenetic tree of sigma-24 factors in *Rhodobacterales*. Species with a photosynthesis gene cluster are in bold. Sigma-24 factors in an operon with *chrR* form a distinct phylogenetic group, marked by a vertical bar. The tree has been constructed using the Neighbour-joining (NJ) method with pairwise deletion of gaps and 500 bootstraps.

shown for the AAP species *Rsb. littoralis* (Berghoff et al. 2011) and *Drb. shibae* (Tomasch et al. 2011). As the main source of singlet oxygen in bacteria is BChl a (Borland et al. 1989), all the phototrophic *Rhodobacterales* contain the RpoE–ChrR system to protect their cellular machinery against damage. However, chemotrophic organisms (which lack BChl a) should not be prone to the formation of singlet oxygen, and the presence of RpoE–ChrR system is unnecessary. This suggests that the ancestor of AH01 was originally a phototrophic bacterium, which later lost its photosynthetic genes, but yet retained the RpoE–ChrR system.

From the presented pieces of evidence, we conclude that the ancestors of the haloalkaliphilic members of the *RR*-group were phototrophic organisms. A more difficult question to answer is whether these ancestral species were photoautotrophic and contained RubisCO. There are two possible scenarios (fig. 6). The first [compatible with Keppen et al. (2013)] assumes that the ancestors of the *RR*-group were photoautotrophs similar to most of the modern *Rhodobacter* species. Then, during evolution, some *RR*-lineages lost a part or all of their photosynthesis genes producing photoheterotrophic, lithoheterotrophic, or chemoheterotrophic species. In the second scenario, the ancestors of the *RR*-group were photoheterotrophs. Here, some lineages (e.g., *Rhodobacter* species) gained RubisCO genes and became photoautotrophs, whereas some lineages lost all their photosynthesis genes and became heterotrophs (fig. 6). These two scenarios are difficult to reconcile based on the available data, but based on the performed phylogenetic analyses we prefer the photoautotrophic ancestor scenario. Indeed, the photoautotrophic *Rhodobacter* species (especially *Rba. veldkampii*) are located closer to the root of the *RR*-group in our 16S rRNA tree, whereas the photoheterotrophic species branch off later (fig. 1). The weakness of this argument is the very low statistical support of the 16S phylogenetic tree. However, a similar conclusion can also be made based on the ALAS phylogeny. The ALAS (*hemA/hemT* gene) tree shows a major split, documenting the existence of two different forms of the gene in *Rhodobacterales* (*hemA*-like and *hemT*-like form). These two forms probably originate from a gene duplication event. Since the *hemA* and *hemT* genes of *Rba. sphaeroides* lay close to the split, it seems that the gene duplication occurred in the ancestors of *Rhodobacter*-like species and all the subsequent photoheterotrophic ancestors retained the two forms of the gene, which gradually diverged. Photoheterotrophic species always contain both forms, which indicate that they only evolved after *Rba. sphaeroides*. Later, the AAP species incorporated the *hemT*-like form of ALAS into the PGC. The presence of ALAS in the PGC seems to be a common characteristic of all the AAP species belonging to *Alphaproteobacteria* (Zheng et al. 2011). This arrangement is probably advantageous for convenient regulation of BChl a synthesis in AAP species, which need to tightly control the initial step of the tetrapyrrol pathway with the final part of the

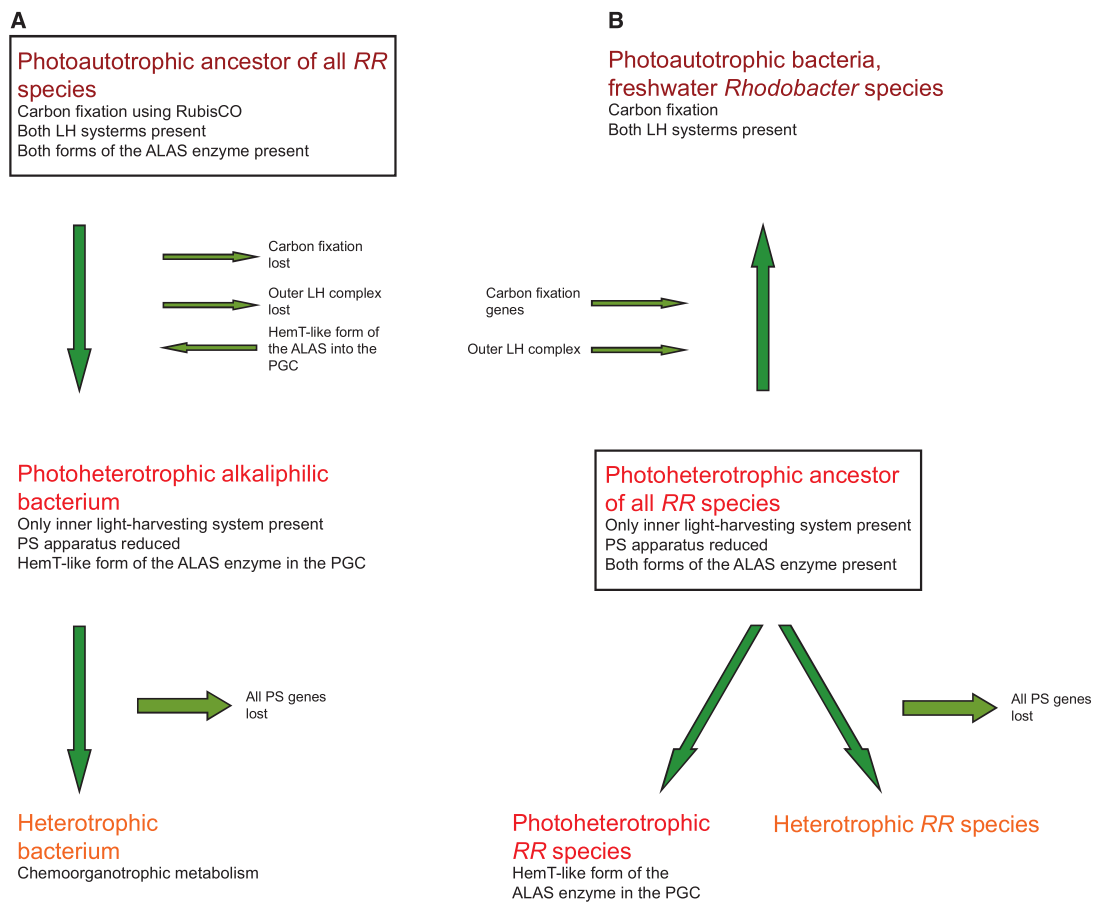


Fig. 6.—Proposed schemes of the evolution of phototrophy among the members of the *RR*-group: (A) Regressive evolution model compatible with Keppen et al. (2013). (B) Mixed model assuming a photoheterotrophic ancestor of all modern *RR*-species—heterotrophs, photoheterotrophs, and photoautotrophs.

BChl *a* pathway. Indeed, this has been documented in *Drb. shibae*, where the inducible ALAS gene located in the PGC (*hemT*-like form) was under strong light regulation together with all *bch* genes (Tomasch et al. 2011). Interestingly, all these ALAS genes present in the PGC cluster together form a distinct subclade of the entire *hemT*-like group (fig. 3). This indicates that the *hemA* split occurred first in photoautotrophic *Rhodobacter*-like species, which probably needed to differentially regulate the tetrapyrrol biosynthesis pathway for heme (*hemA* form) and BChl *a* synthesis (*hemT* form). Later, after the evolution of the AAP species, the *hemT* form of the gene was moved under the common regulation of the PGC cluster. Moreover, in some species (*Drb. shibae* and *Roseobacter* species), the *hemT* form of ALAS underwent a second duplication, indicating the need for even more delicate regulation of the gene.

In conclusion, the presented data indicate that the ancestors of the haloalkaliphilic members of the *RR*-group were phototrophic species. The heterotrophic species have evolved through the regressive loss of their photosynthetic apparatus. In addition, we have demonstrated that the important step enabling the evolution of photoheterotrophic species was the duplication of the *hemA* gene, and its later incorporation into the PGC of the AAP species.

Many interesting questions were left to be answered such as if there is an alternative pathway of descent from photoautotrophic species through the facultatively autotrophic aerobic or facultatively anaerobic lithotrophs, such as *Paracoccus* (loss of the PGC), to aerobic chemolithoheterotrophs (loss of RubisCO) and finally organoheterotrophs? What was the role of the hydrogenase in this alternative scenario? Did *Rba. capsulatus* evolve regressively from a *Rba. sphaeroides*-like

species through the loss of genes (*acsF*, *hemF*, *hemT*, and anti-sigma factor), or does it represent a former (more ancient) photoautotrophic species? We believe, that the constantly expanding genomic data, will allow us to address these questions in the near future.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

The authors are grateful to Dr Katya Boldareva-Nuyanzhina for providing her strains *Rca. barguzinensis* and *Rna. thiooxidans*. We also thank Jason Dean BSc. for the language correction. This research has been supported by the GAČR project P501/12/G055, the DAAD project 57155424, the EC-funded project Algattech Plus (LO1416), and the RFBF grant 16-04-00035. J.T. was supported by the Deutsche Forschungsgemeinschaft (DFG) within the Transregio 51 "Roseobacter". D.Y.S. was supported by the Gravitation Program (SIAM, grant 24002002) from the Dutch Ministry of Education and Science.

Literature Cited

- Androga DD, Özgür E, Eroglu I, Yücel M, Gündüz U. 2012. Photofermentative hydrogen production in outdoor conditions. Croatia: INTECH Open Access Publisher.
- Anthony JR, Newman JD, Donohue TJ. 2004. Interactions between the *Rhodobacter sphaeroides* ECF sigma factor, σE , and its anti-sigma factor, ChrR. *J Mol Biol.* 341(2):345–360.
- Auch AF, von Jan M, Klenk HP, Göker M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci.* 2(1):117–134.
- Aziz RK, et al. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9(1):1.
- Berghoff BA, et al. 2011. Anoxygenic photosynthesis and photooxidative stress: a particular challenge for Roseobacter. *Environ Microbiol* 13(3):775–791.
- Boldareva EN, et al. 2006. The new alkaliphilic bacteriochlorophyll *a*-containing bacterium *Roseinatronobacter monicus* sp. nov. from the hypersaline Soda Mono Lake (California, United States). *Microbiology* 76:82–92.
- Boldareva EN, et al. 2008. *Rhodobaca barguzinensis* sp. nov., a new alkaliphilic purple nonsulfur bacterium isolated from a soda lake of the Barguzin Valley (Buryat Republic, Eastern Siberia). *Microbiology* 77(2):206–218.
- Boldareva EN, Bláhová Z, Sobotka R, Koblížek M. 2013. Distribution and origin of oxygen-dependent and oxygen-independent forms of Mg-protoporphyrin monomethylester cyclase among phototrophic proteobacteria. *Appl Environ Microbiol.* 79(8):2596–2604.
- Boldareva EN, Gorlenko VM. 2014. *Roseibacula alcaliphilum* gen. nov. sp. nov., a new alkaliphilic aerobic anoxygenic phototrophic bacterium from a meromictic soda lake Doroninskoe (East Siberia, Russia). *Microbiology* 83:381–390.
- Bolt EL, et al. 1999. Characterization of the *Rhodobacter sphaeroides* 5-aminolaevulinic acid synthase isoenzymes HemA and HemT, isolated from recombinant *Escherichia coli*. *Eur J Biochem.* 265(1):290–299.
- Borland CF, Cogdell RJ, Land EJ, Truscott TG. 1989. Bacteriochlorophyll *a* triplet state and its interactions with bacterial carotenoids and oxygen. *J Photochem Photobiol B* 3(2):237–245.
- Brinkhoff T, Giebel HA, Simon M. 2008. Diversity, ecology, and genomics of the Roseobacter clade: a short overview. *Arch Microbiol.* 189(6):531–539.
- Brocks JJ, et al. 2005. Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. *Nature* 437(7060) 866–870.
- Canfield DE, et al. 2008. Ferruginous conditions dominated later Neoproterozoic deep-water chemistry. *Science* 321(5891):949–952.
- Dailey HA, et al. 2017. Prokaryotic heme biosynthesis: multiple pathways to a common essential product. *Microbiol Mol Biol Rev.* 81(1): e00048–16.
- Fanica-Gaignier M, Clement-Metral J. 1973. 5-Aminolevulinic acid synthetase of *Rhodospseudomonas sphaeroides* Y. *Eur J Biochem.* 40(1): 13–18.
- Fitch WM. 1970. Distinguishing homologous from analogous proteins. *Syst Zool.* 19:99–113.
- Garrity GM, Bell JA, Lilburn T. 2005. Order III. *Rhodobacterales* ord. nov. In: Vos P, et al., editors. *Bergey's Manual® of Systematic Bacteriology*. New York: Springer US. p. 270–323.
- Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35(Web Server):52–57.
- Keppen OI, Krasil'nikova EN, Lebedeva NV, Ivanovskii RN. 2013. Comparative study of metabolism of the purple photosynthetic bacteria grown in the light and in the dark under anaerobic and aerobic conditions. *Microbiology* 82(5):547–553.
- Koblížek M, Moulisová V, Muroňová M, Oborník M. 2015. Horizontal transfers of two types of puf operons among phototrophic members of the Roseobacter clade. *Folia Microbiol.* 60(1):37–43.
- Koblížek M, Zeng Y, Horák A, Oborník M. 2013. Regressive evolution of photosynthesis in the Roseobacter clade. *Adv Bot Res.* 66:385–405.
- Labrenz M, Lawson PA, Tindall BJ, Hirsch P. 2009. *Roseibaca ekhonensis* gen. nov., sp. nov., an alkalitolerant and aerobic bacteriochlorophyll *a*-producing alphaproteobacterium from hypersaline Ekho Lake. *Int J Syst Evol Microbiol.* 59 (Pt 8):1935–1940.
- Lietenberg S, et al. 2008. Organization and expression of photosynthesis genes and operons in anoxygenic photosynthetic proteobacteria. *Environ Microbiol.* 10(9):2267–2276.
- Luo H, Moran MA. 2014. Evolutionary ecology of the marine Roseobacter clade. *Microbiol Mol Biol Rev.* 78(4):573–587.
- Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP. 2013. When should a DDH experiment be mandatory in microbial taxonomy? *Arch Microbiol.* 195(6):413–418.
- Milford AD, Achenbach LA, Jung DO, Madigan MT. 2000. *Rhodobaca bogoriensis* gen. nov. and sp. nov., an alkaliphilic purple nonsulfur bacterium from African Rift Valley soda lakes. *Arch Microbiol.* 174(1–2):18–27.
- Moran MA, et al. 2007. Ecological genomics of marine Roseobacters. *Appl Environ Microbiol.* 73(14):4559–4569.
- Petersen J, Frank O, Göker M, Pradella S. 2013. Extrachromosomal, extraordinary and essential—the plasmids of the Roseobacter clade. *Appl Microbiol Biotechnol.* 97(7):2805–2815.
- Raymond J, Zhaxybayeva O, Gogarten JP, Gerdes SY, Blankenship RE. 2002. Whole-genome analysis of photosynthetic prokaryotes. *Science* 298(5598):1616–1620.
- Simon M, et al. 2017. Phylogenomics of *Rhodobacteraceae* reveals evolutionary adaptation to marine and non-marine habitats. *ISME J.* 11(6):1483–1499.
- Sorokin DY, Tourova TP, Kuenen JG. 2000a. A new facultatively autotrophic hydrogen- and sulphur-oxidizing bacterium from alkaline environment. *Extremophiles* 4(4):237–245.

- Sorokin DY, Tourova TP, Kuznetsov BB, Bryantseva IA, Gorlenko VM. 2000b. *Roseinatrobacter thiooxidans* gen. nov., sp. nov., a new alkaliphilic aerobic bacteriochlorophyll-a-containing bacteria from a soda lake. *Microbiology* 69(1):89–97.
- Stadnichuk IN, et al. 2009. Photosynthetic activity and components of the electron transport chain in the aerobic bacteriochlorophyll a-containing bacterium *Roseinatrobacter thiooxidans*. *Microbiology* 78(1):7–15.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 56(4):564–577.
- Tomasch J, Gohl R, Bunk B, Diez MS, Wagner-Döbler I. 2011. Transcriptional response of the photoheterotrophic marine bacterium *Dinoroseobacter shibae* to changing light regimes. *ISME J.* 5(12):1957–1968.
- Vollmers J, et al. 2013. Poles apart: Arctic and Antarctic Octadecabacter strains share high genome plasticity and a new type of xanthorhodopsin. *PLoS One* 8(5):e63422.
- Wagner-Döbler I, Biebl H. 2006. Environmental biology of the marine Roseobacter lineage. *Annu Rev Microbiol.* 60:255–280.
- Wang Y, Coleman-Derr D, Chen G, Gu YQ. 2015. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 43(W1):78–84.
- Willows RD, Krieger AM. 2009. Biosynthesis of bacteriochlorophylls in purple bacteria. In: Hunter CN, Daldal F, Thurnauer MC, and Beatty JT, editors. *The purple phototrophic bacteria*. Netherlands: Springer. p. 57–79.
- Yu Y, Yan SL, Li HR, Zhang XH. 2011. *Roseicitreum antarcticum* gen. nov., sp. nov., an aerobic bacteriochlorophyll a-containing alphaproteobacterium isolated from Antarctic sandy intertidal sediment. *Int J Syst Evol Microbiol.* 61(Pt 9):2173–2179.
- Zeng Y, Feng F, Medová H, Dean J, Koblížek M. 2014. Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. *Proc Natl Acad Sci USA.* 111(21):7795–7800.
- Zheng Q, et al. 2011. Diverse arrangement of photosynthetic gene clusters in aerobic anoxygenic phototrophic bacteria. *PLoS One* 6(9):e25050.
- Zsebo KM, Hearst JE. 1984. Genetic-physical mapping of a photosynthetic gene cluster from *R. capsulata*. *Cell* 37(3):937–947.

Associate editor: Eric Baptiste

Suppl. Table 1. Comparison of RAST gene distribution in *Rhodobacterales* genomes.

	DSM 2351	2.4.1	alga05	ALG1	AH01	PD 1222
Cofactors, Vitamins, Pigments	8.32%	8.32%	10.06%	10.23%	8.40%	10.07%
Cell Wall and Capsule	4.23%	3.19%	3.72%	4.20%	2.65%	3.47%
Virulence, Disease and Defense	2.22%	2.31%	2.03%	1.76%	1.91%	2.26%
Potassium metabolism	0.74%	0.65%	0.63%	0.52%	0.46%	0.30%
Photosynthesis	0.46%	0.50%	0.28%	0.20%	0.00%	0.00%
Miscellaneous	1.45%	1.30%	1.75%	1.92%	1.48%	1.98%
Phages, Prophages, Plasmids	1.48%	1.43%	1.30%	0.84%	0.88%	0.92%
Membrane Transport	8.57%	7.96%	7.78%	6.83%	7.55%	7.09%
Iron acquisition and metabolism	1.90%	0.62%	1.02%	0.16%	1.27%	1.79%
RNA Metabolism	4.72%	4.36%	4.59%	5.24%	4.80%	3.92%
Nucleosides and Nucleotides	3.77%	3.90%	3.96%	4.08%	3.63%	3.32%
Protein Metabolism	7.51%	8.29%	7.47%	8.43%	7.02%	6.13%
Cell Division and Cell Cycle	1.06%	0.94%	7.47%	1.08%	0.95%	0.87%
Motility and Chemotaxis	1.94%	3.67%	1.54%	1.84%	2.61%	0.40%
Regulation and Cell signaling	1.69%	1.89%	1.44%	1.24%	1.80%	2.06%
Secondary Metabolism	0.14%	0.13%	0.14%	0.16%	0.14%	0.15%
DNA Metabolism	4.19%	3.35%	2.80%	3.64%	3.14%	2.36%
Fatty Acids, Lipids, Isoprenoids	5.60%	4.45%	5.29%	5.48%	4.55%	4.76%
Nitrogen Metabolism	1.34%	1.07%	1.96%	1.40%	1.41%	2.36%
Dormancy and Sporulation	0.00%	0.07%	0.04%	0.04%	0.07%	0.02%
Respiration	3.84%	6.57%	5.26%	5.12%	4.73%	4.17%
Stress Response	4.27%	4.13%	4.35%	4.48%	5.01%	4.54%
Metabolism of Aromatic Compounds	0.63%	0.68%	1.93%	2.00%	1.52%	2.46%
Amino Acids and Derivates	14.80%	13.85%	14.48%	14.67%	13.90%	16.10%
Sulfur Metabolism	0.81%	0.72%	1.19%	1.28%	1.38%	1.86%
Phosphorus Metabolism	1.23%	1.27%	1.23%	1.36%	1.20%	1.56%
Carbohydrates	13.08%	14.89%	12.83%	12.03%	17.54%	15.10%

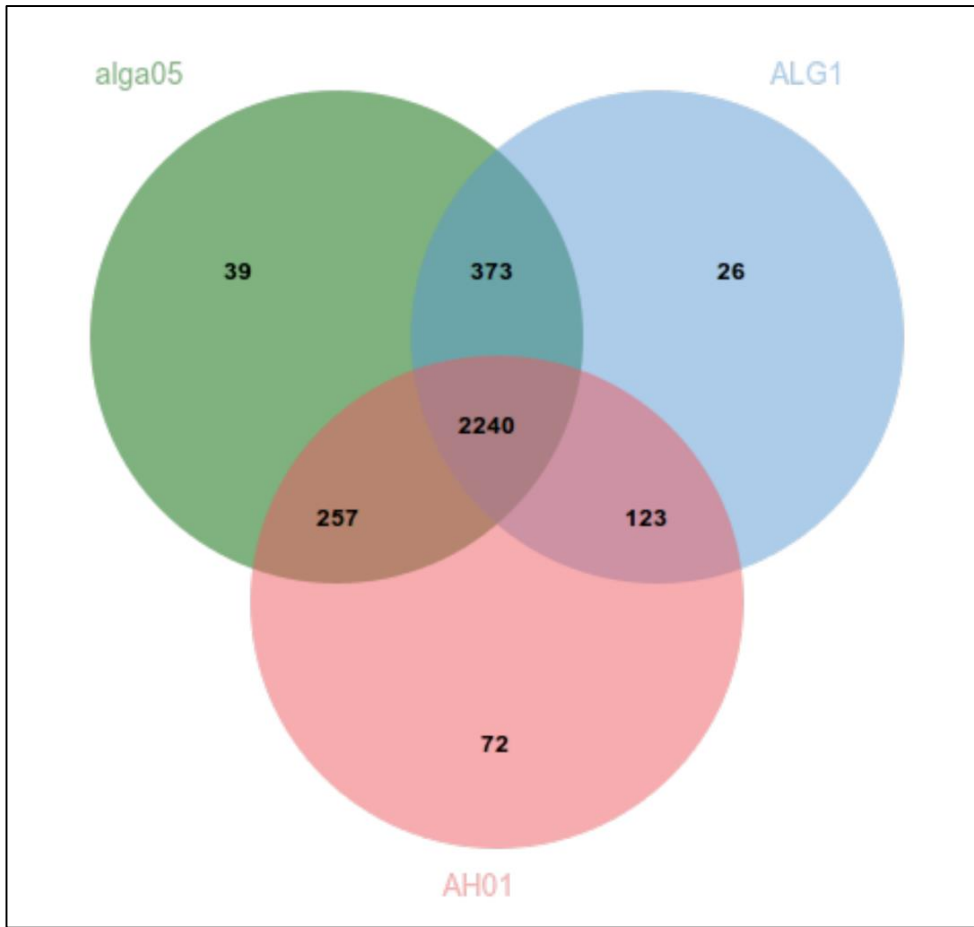
Suppl. Table 2. Numbers of gene clusters in the studied genomes.

	Organism		
	alga05	ALG1	AH01
Nr. of ORFs	3778	3385	3885
Nr. of gene clusters	2903	2762	2690
Nr. of gene singletons	786	571	1054
% of clusters shared by all three strains	77.3	81.2	83.4
% of unique clusters	1.1	0.9	2.7

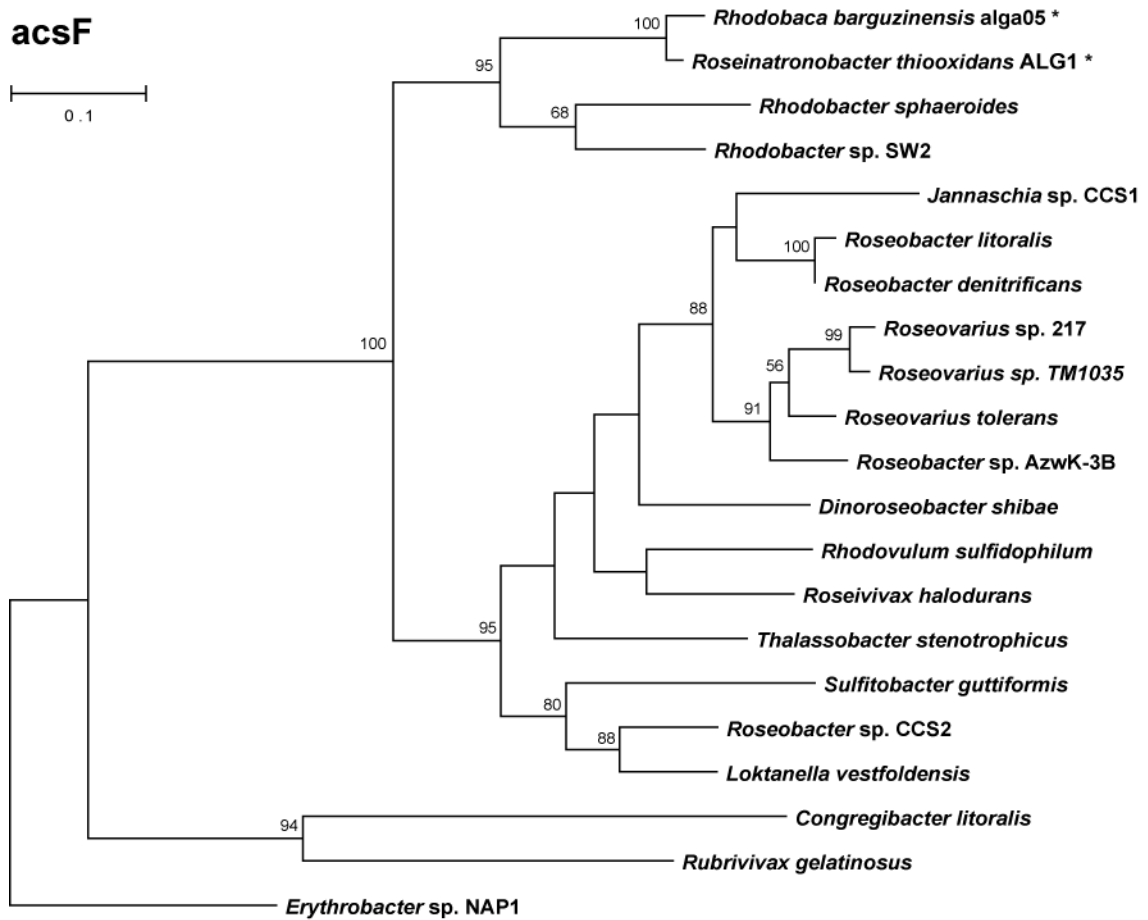
Suppl. Table 3. Presence of selected metabolic pathways in the studied genomes.

	DSM 2351	2.4.1	alga05	ALG1	AH01	PD 1222
Nitrogen Metabolism						
Denitrification	○	○	○	○	○	+
Dissimilatory nitrate reduction	○	○	○	○	+	+
Sulfur Metabolism						
DMSP demethylase (<i>dmdA</i>)	–	–	–	–	+	–
DMSP lyase (<i>ddlL</i>)	–	+	+	+	–	–
DMSP cleavage (<i>dddD</i>)	–	–	–	–	–	–
DMSP cleavage (<i>dddP</i>)	–	–	–	–	–	–
Sulfur oxidation (<i>sox</i>)	+	○	+	○	+	+

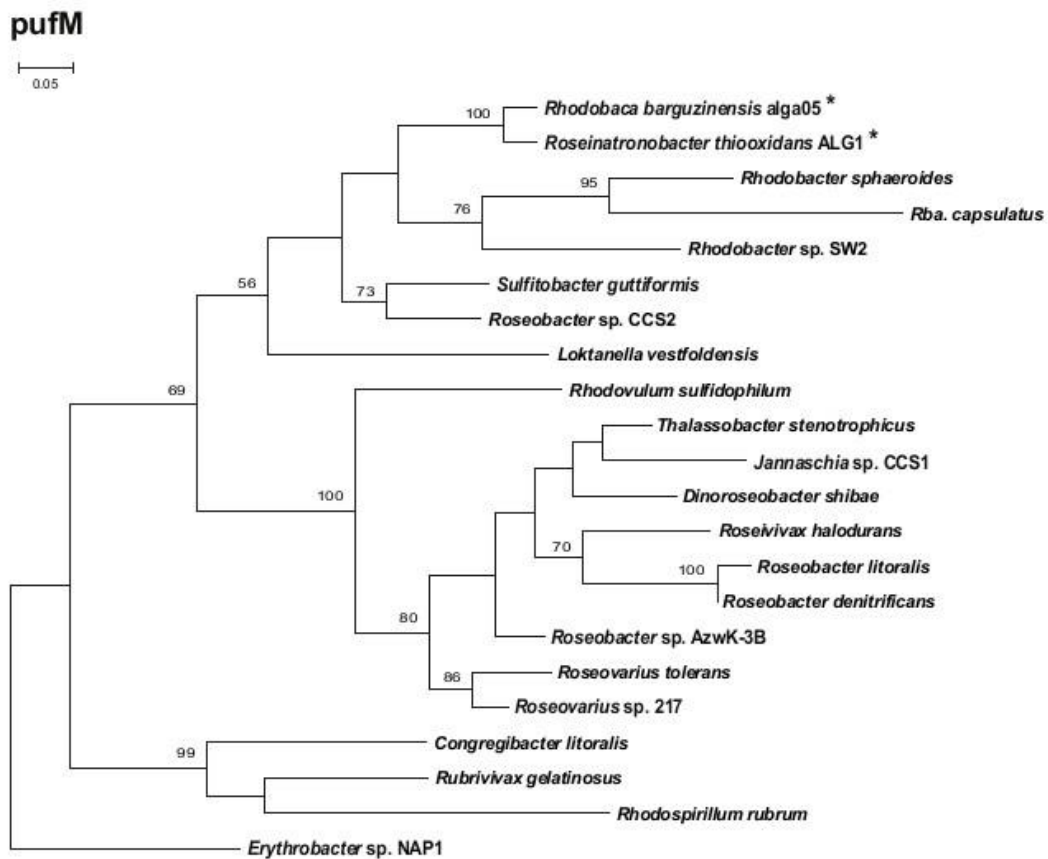
The genes present in the genome and complete metabolic pathways are marked "+"; the incomplete pathways are marked "○"; "–" means gene/enzyme/metabolic pathway absent.



Suppl. Fig. 1. Venn-Diagram showing the distribution of shared gene families among sequenced strains. Each strain is represented by a circle, overlapping regions illustrate gene clusters that are shared between each strain. The cluster number in each region is listed.



Suppl. Fig. 2. Maximum likelihood phylogenetic tree based on protein sequences of the oxygen-dependent (*acsF* gene product, 334 common amino acid positions), bootstrap 500x. *Congregibacter litoralis*, *Rubrivivax gelatinosus* and *Erythro bacter* sp. NAP1 were used as an outgroup organism. Scale bars represent changes per position. Bootstrap values >50% are shown. Studied strains are marked by the asterisk.

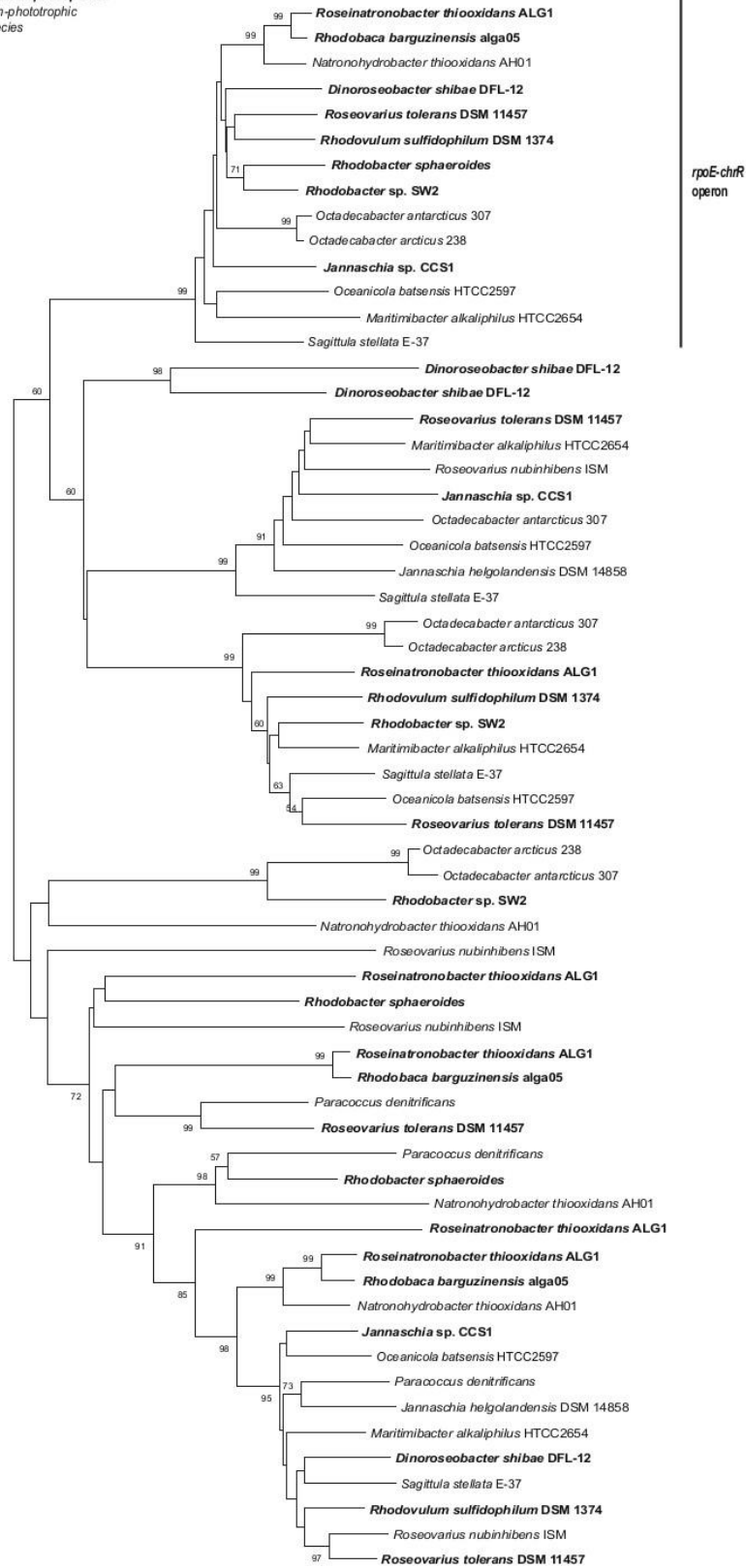


Suppl. Fig. 3. Maximum likelihood phylogenetic tree based on protein sequences of the photosynthetic reaction center subunit M (*pufM* gene product, 289 common amino acid positions), bootstrap 500x. *Erythrobacter* sp. NAP1, *Rhodospirillum rubrum*, *Rubrivivax gelatinosus* and *Congregibacter litoralis* were used as an outgroup organism. Scale bars represent changes per position. Bootstrap values >50% are shown. Studied strains are marked by the asterisk.

rpoE



Phototrophic species
Non-phototrophic species



Suppl. Fig. 4. Phylogenetic tree of sigma-24 factors in *Rhodobacterales*. Species with a photosynthesis gene cluster are in bold. Sigma-24 factors in an operon with *chrR* form a distinct phylogenetic group, marked by a vertical bar. The tree has been constructed using the Neighbour-joining (NJ) method with pairwise deletion of gaps and 500 bootstraps.

2.3 Paper 3.

Clustered core- and pan-genome content on *Rhodobacteraceae* chromosomes.

Replication of the bacterial chromosome starts from a single origin of replication (*oriC*) and continues on both replichores. Physical distance of genes from the *oriC* is one of the most conserved properties of the genome architecture (Eisen *et al.*, 2000; Sobetzko *et al.*, 2012). This parameter can influence gene expression (Rocha 2004; Couturier and Rocha, 2006), gene conservation (Flynn *et al.*, 2010; Sobetzko *et al.*, 2012), as well as evolutionary rates of orthologous genes (Flynn *et al.*, 2010).

In the third study presented we focussed on the genome architecture of 101 fully-sequenced *Rhodobacteraceae* representatives. For each strain we identified highly conserved core genes and estimated their statistical distributed along the chromosome. The vast majority of the species had their conserved genes distributed statistically closer to the origin of replication. Interestingly, the obtained data also showed a statistically significant bias in the clustering of core genes towards the terminus of replication in some *Rhodobacteraceae* representatives. In these few representatives the number of highly conserved genes increased with the distance from the *oriC*. This kind of genome architecture was the most pronounced in the haloalkaliphilic photoheterotrophic bacterium *Rhodobaca barguzinensis* and heterotrophic freshwater bacterium *Loktanella vestfoldensis*. We propose that the observed mosaic architecture of *Rhodobacteraceae* genomes can be partially explained by phage integration and HGT.

Clustered core- and pan-genome content on *Rhodobacteraceae* chromosomes

Karel Kopejtko^{1,2}, Yan Lin^{3,4}, Markéta Jakubovičová⁵, Michal Koblížek^{1,2}, Jürgen Tomasch^{6*}

¹*Laboratory of Anoxygenic Phototrophs, Center Algatech, Institute of Microbiology CAS, Třeboň, Czech Republic*

²*Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic*

³*Department of Physics, School of Science, Tianjin University, Tianjin, China*

⁴*SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering, Tianjin, China*

⁵*Faculty of Information Technology, Czech Technical University in Prague*

⁶*Department of Molecular Bacteriology, Helmholtz Centre for Infection Research, Braunschweig, Germany*

Under review in *Genome Biology and Evolution*

*Author for correspondence: Juergen.Tomasch@helmholtz-hzi.de

ABSTRACT

In Bacteria, chromosome replication starts at a single origin of replication and proceeds on both replichores. Due to its asymmetric nature, replication influences chromosome structure and gene organization, mutation rate and expression. To date, little is known about the distribution of highly conserved genes over the bacterial chromosome. Here, we use a set of 101 fully-sequenced *Rhodobacteraceae* representatives to analyze the relationship between the number of core genes and their distance from the origin of replication. Twenty-two of the analyzed species had their core genes clustered significantly closer to the origin of replication with representatives of the genus *Celeribacter* being the most apparent example. Interestingly, there were also eight species with the opposite organization. In particular *Rhodobaca barguzinensis* and *Loktanella vestfoldensis* showed a significant increase of core genes with distance from the origin of replication. The uneven distribution of low-conserved regions is in particular pronounced for genomes in which the halves of one replichore differ in their core gene content. Phage integration and horizontal gene transfer partially explain the scattered nature of *Rhodobacteraceae* genomes. Our findings lay the foundation for a better understanding of bacterial genome evolution and the role of replication therein.

Key words: genome architecture, genome evolution, origin of replication, *Rhodobacteraceae*.

INTRODUCTION

Replication is assumed to be a key factor in the evolution of genome structure and organization. In contrast to eukaryotes and archaea, where the chromosome replication proceeds simultaneously from multiple sites, replication of bacterial chromosomes starts from a single origin of replication (*oriC*) and continues equally along both replichores (two halves of the chromosome extending from *oriC*) up to the terminus of replication (*ter*). Since cell division is often shorter than the time required for the replication of the chromosome itself, it leads to the occurrence of multiple replication complexes in the cell. In result, the genes located in the early replicating regions near the *oriC* can be present in multiple copies and hence have a higher expression level compared to genes in the late replicating regions. This so-called gene-dosage effect is especially pronounced in fast-growing bacteria, where strongly expressed genes are preferentially concentrated near the *oriC* (Rocha 2004; Couturier and Rocha, 2006). Bacterial chromosome architecture can be also shaped by large-scale interreplichore translocations, as was recently shown using genome sequence comparisons between 262 closely related pairs of bacterial species (Khedkar and Seshasayee, 2016).

Relative distance of genes from the *oriC* is commonly thought to be one of the most conserved properties of genome organization (Eisen *et al.*, 2000; Sobetzko *et al.*, 2012). Results of an extensive analysis comprising a set of 131 gammaproteobacterial genomes showed strong conservation in the relative distance of conserved genes coding for regulatory elements from the *oriC* (Sobetzko *et al.*, 2012).

A replication-biased genome organization was also revealed in archaeal genomes; Flynn *et al.* (2010) selected six *Sulfolobus* genomes each with multiple replication origins, to test the hypothesis that genes situated close to *oriC* tend to be more conserved than genes more distant

from *oriC*. Results of this study clearly demonstrated a bias in the location of conserved orthologous genes (orthologs) towards the *oriC* site. Moreover, the analysis of evolutionary rates of these orthologs revealed their slower evolution when compared with genes more distant from *oriC* (Flynn *et al.*, 2010). Another study of the *Sulfolobus* genome architecture showed a mosaic of recombinant single nucleotide polymorphisms (SNP) along the chromosomes of ten closely related *Sulfolobus islandicus* strains. This comparative genome analysis revealed large genomic regions surrounding all *oriC* sites that show reduced recombination rates (Krause *et al.*, 2014). Based on an analysis of the codon composition of genomes from 59 prokaryotic organisms, it was shown that genes organized close to the *ter* are in many cases A+T-enriched at the third codon position thought to reflect a higher evolutionary rate (Daubin and Perrière, 2003). Also, horizontally transferred DNA was suggested to cluster near the *ter* (Rocha 2004); this was documented in the genome of *Escherichia coli* (Lawrence and Ochmann, 1998), as well as in other prokaryotic genomes (Touchon and Rocha, 2016). Early studies presenting complete genome sequence of *Bacillus subtilis* (Kunst *et al.*, 1997) and *E. coli* (Blattner *et al.*, 1997; Lawrence and Ochmann, 1998) reported a frequent occurrence of prophages around the *ter*. Furthermore, a recent study conducted on a large and diverse sample of bacterial species revealed a positive bias in the occurrence of hot-spots for HGT containing prophages towards the *ter* (Oliveira *et al.*, 2017). Also, it has been speculated that the gene-dosage effect might lead to fixing of typically weakly (or not at all) expressed horizontally transferred genes closer to the *ter* site (Rocha 2004).

To date, there is no comprehensive study on the distribution of conserved genes over the bacterial chromosome focused on one bacterial family. Therefore we decided to analyze the relationship between the degree of gene conservation and its distance from the origin of replication for the

core- and pan-genome of all *Rhodobacteraceae* (Alphaproteobacteria) with closed genomes. All the genes present in a certain (microbial) clade are the pan-genome. The genes present in genomes of all the strains are defined as core genome, whereas the term accessory genome describes partially shared and strain specific genes (Medini *et al.*, 2005; Tettelin *et al.*, 2005). *Rhodobacteraceae* were selected as a family with dynamic evolution, and a large number of sequenced genomes. The dataset comprises 109 species originating from diverse habitats of soil, freshwater, marine and hypersaline environment (Simon *et al.*, 2017). The frequent occurrence of plasmids (Petersen *et al.*, 2013), transposable elements (Vollmers *et al.*, 2013) and gene-transfer agents (Tamarit *et al.*, 2017) within this family suggest that HGT plays an important role in shaping the genomes of this family.

METHODS

Data. Nucleotide genomic sequences and corresponding Genbank FASTA files for 109 fully-sequenced *Rhodobacteraceae* strains (Fig. 1) were obtained from NCBI GenBank. 16S rRNA gene sequences for the same set of strains were obtained either from the SILVA database (Quast *et al.*, 2012) or NCBI GenBank (May, 2018).

Comparative genomic analysis. In order to standardize further analysis we re-annotated all genomes using Prokka (Seeman 2014). Orthologous gene cluster analysis was performed using the Proteinortho dataframe (Lechner *et al.*, 2011). As in previous comparative studies (Kalhöfer *et al.*, 2011; Thole *et al.*, 2012; Vollmers *et al.*, 2013), orthologous protein sequences were identified with three cut-off criteria : (1) the e-value should be $\leq 1e-10$; (2) the alignment should cover at least 70% of the sequences; and (3) the sequence identity should be $\geq 30\%$. Since for highly conserved proteins with $>50\%$ sequence identity the probability of completely incorrect

annotation is very low (<6%) (Sangar *et al.*, 2007), we also involved an additional cut-off value of 60% sequence identity and 80% coverage to determine the number of orthologs with highly conserved sequences.

Phylogenomic analysis. Amino acid sequences for the 85 highly conserved core genome proteins (60% sequence identity, 80% coverage) were individually aligned using ClustalX version 2.1. Sites containing gaps and ambiguously aligned regions were removed from each alignment using Gblocks (Talavera and Castresana, 2007) and finally these alignments were concatenated with Geneious version 8.1.2 (Biomatters Ltd.). The phylogenomic tree was inferred by MEGA 6.0 software using the maximum likelihood (ML) algorithm with LG model and 100 bootstrap replicates. 16S rRNA gene sequences for the same set of strains as in phylogenomic tree were aligned using ClustalX version 2.1, ambiguously aligned regions and gaps were excluded from the alignment using Gblocks. The 16S rRNA tree was constructed by PhyML/MEGA 6.0 software using the ML algorithm with HKY85 nucleotide substitution model and 1,000 bootstrap replicates. The four strains of the deep branching “*Stappia* group” (Pujalte *et al.*, 2014), i.e. *Pannonibacter phragmitetus* 31801, *Labrenzia* sp. VG12, *L.* sp. CP4, and *L. aggregata* RMAR6-6, were used as outgroup organisms to root both trees.

Identification of *oriC*. The origin of replication (*oriC*) of studied strains were identified using Ori-Finder (Gao and Zhang, 2008; Luo *et al.*, 2018), which was developed mainly based on the analysis of nucleotide composition asymmetry using the Z-curve approach and the distribution of DnaA boxes. Three different DnaA box motives (i.e. TTATCCACA, TGTTTCACG, and TGTGGATAT) were used during the search. Typically, the *Escherichia coli* perfect DnaA box (TTATCCACA) is the most used motive for regular prediction (Mackiewicz *et al.*, 2004). When only one unmatched site was allowed, the *oriCs* of a few genomes could not be identified.

Whereas, a number of alternative *oriC*s were predicted when we allowed two unmatched sites. The output of Ori-Finder was manually curated. When a number of alternative *oriC*s was predicted, we decided which one is most likely the right one considering these three criteria: (1) proximity to the GC disparity minimum; (2) location within the local GC minimum; (3) proximity to the *parAB* genes.

Identification of phages and horizontally transferred genes. The web-based PHAge Search Tool – Enhanced Release (PHASTER) was used for predicting prophage sequences or remnants of those (Arndt *et al.*, 2016). For prediction of genomic islands (GIs) we used AlienHunter (Vernikos and Parkhill, 2006) and the web-based tool IslandViewer (Bertelli *et al.* 2017). AlienHunter predicts GIs using Interpolated Variable Order Motifs (IVOMs). This approach exploits compositional biases by determining of variable order motif distributions. IslandViewer integrates three different GI prediction tools: IslandPath-DIMOB (Hsiao *et al.*, 2003), SIGI-HMM (Waack *et al.*, 2006) and IslandPick (Langille *et al.*, 2008). All GIs predicted by at least one method were considered for further analysis.

Statistical analysis. Mean number of orthologs and mean distance to *oriC* was calculated for sliding windows of 20 genes. Linear and quadratic models were fitted to the data. For the linear models slope and corresponding p-value were extracted. The chromosomes were separated into eight equally sized segments and the mean and standard deviation for the number of orthologs was calculated. The number of phage regions and genomic islands were calculated for three parts with increasing distance from *oriC*. The distribution of the number of loci was visualized using boxplots. Analysis of variance (ANOVA) was used to test for significant differences between the eight chromosomal segments as well as the three parts used for Phage and HGT analysis. Tukey's test was used to identify the segments and parts with significant differences in number of

orthologs or phages and HGT regions, respectively.

RESULTS AND DISCUSSION

Phylogenomic analysis. To show the overall picture of phylogenetic relationships between the studied strains we constructed a phylogenomic species tree and 16S rRNA tree as a reference. Analysis of the selected *Rhodobacteraceae* genomes identified a core genome of 85 genes. The core genes were used to construct a robust phylogenomic tree (Figure 1, left). The obtained tree had good statistical support and it also agreed well with other recent phylogenomic studies of this family (Simon *et al.*, 2017; Brinkmann *et al.*, 2018). The 16S rRNA phylogenetic tree (Figure 1, right), which was based on an alignment with 1,260 common nucleotide positions, shows a more mosaic branching pattern with considerably lower statistical support when compared with the phylogenomic tree. The most striking difference between both methods was in the clustering of *Rhodovulum* species. In the phylogenomic tree, these strains clearly clustered with the *Roseobacter* group (Simon *et al.*, 2017), whereas in the 16S rRNA tree they were placed close to the *Rhodobacter/Rhodobaca* (*RR*) group, as we found before (Kopejtko *et al.*, 2017; Kopejtko *et al.*, 2018). The selected strains represent the full spectrum of *Rhodobacteraceae* from various environments (Supplementary Table S1) although marine *Roseobacter* strains and in particular the genus *Phaeobacter* (Freese *et al.*, 2014) are over-represented.

Identification of *oriC* locus. We started the data analysis by assigning to each genome the coordinates of its *oriC* region. After manual curation of the predictions made by Ori-Finder we were able to clearly pinpoint the *oriC* for 101 strains. Due to missing overrepresentation of DnaA boxes and/or lack of distinct differences in nucleotide composition compared with the rest of the genome we could not unambiguously identify the *oriC* of eight strains, which we excluded from further analysis (Supplementary Table S2).

Location of conserved genes in relation to *oriC*. To reveal potential bias in localization of core genes along the bacterial chromosome for each of the 101 genomes with identified *oriC* we compared the mean number of orthologous per gene for sliding windows of 20 genes to the midpoint distance of the sliding window to *oriC*. Next, we fitted linear and quadratic models of the number of orthologs for increasing distances from *oriC* for all these 101 genomes (Supplementary Figures S1 and S2). We identified strains with statistically significant negative (i.e. mean number of orthologs decreasing with distance to *oriC*) or positive (i.e. mean number of orthologs increasing with distance to *oriC*) slope values of the linear model (Figure 2A; Supplementary Table S2). For orthologs identified under less stringent conditions (sequence identity $\geq 30\%$, coverage $\geq 70\%$) we found 34 strains, analysis conducted with a more stringent cut-off criteria (sequence identity $\geq 60\%$, coverage $\geq 80\%$) yielded 21 strains with statistically significant negative or positive slope values (Supplementary Figure S3 and Supplementary Table S2). In some cases – in particular for *Paracoccus* sp. CBA4604 – the quadratic model showed a better fit than the linear model. This indicated that the number of orthologs is on average higher in the middle of the replichores rather than close to the *oriC* or *ter* regions (Supplementary Figures S1 and S2).

The majority of strains (81 out of 101; sequence identity $\geq 30\%$, coverage $\geq 70\%$) had a negative slope. Thus, the genomes showed a tendency towards having highly conserved genes clustered closer to *oriC*. While the negative slope values perfectly followed a normal distribution, the positive slope values were systematically higher than what would be expected from normally distributed data (Figure 2A). We then specifically focused on the strains with the lowest (all from genus *Celeribacter*) and highest (*R. barguzinensis* and *L. vestfoldensis*.) slope values indicating a significant increase (Figure 2B, upper panel) or decline (Figure 2B, lower panel) in mean

number of orthologs with increasing relative distance from *oriC*.

Clustering of core- and pan-genome content in representative *Rhodobacteraceae*. We created chromosome maps for all *Rhodobacteraceae* with identified *oriC* (Supplementary Figure S4) and particularly focused on the five strains representing two different extremes of chromosome architecture. The strain with the most negative slope value, *C. marinus*, showed a conspicuous switch in GC-skew within the right replicore (Supplementary Figure S5). This indicates either a recent genomic inversion event or a misassembled genome. Thus, this genome is not further discussed.

The four further analyzed strains including *L. vestfoldensis* and *R. barguzinensis*, *C. indicus* and *C. manganoxidans* indeed showed highly conserved regions in which the core genes clustered, interrupted by regions of genes with orthologs in only a small number or even no other strains (Figure 3). However, besides the distance to *oriC*, the distribution of core genes also varied between both replicores. We separated the chromosome into eight equally-sized segments and calculated the average number of orthologs in each segment (segment 1 surrounded *oriC* and counting proceeding clockwise). On average the number of conserved genes was higher in two opposing segments surrounding *oriC* and *ter* (Figure 3, center panel). Only segment 1 showed a significant enrichment in orthologs compared to all others, with the exception being segment 5 (Supplementary Figure S6).

L. vestfoldensis and *R. barguzinensis* both showed the core genes concentrated around *ter* and bigger regions of low conservation around *oriC* (Figure 3A). The *L. vestfoldensis* genome showed differences between segments with higher average of conserved genes in the segments 3 to 6, thus around *ter* and on the right replicore. In contrast the genome of *R. barguzinensis* showed a concentration of highly conserved and core genes in the segments 4 to 6 while in particular the

segments 1, 3, 7 and 8 contained large stretches of regions with only few conserved orthologs in the genomes of other strains. Thus, the increase of conserved genes towards *ter* can be attributed to large regions of low conservation on both replichores.

C. indicus and *C. manganoxidans* both showed the core genes in a tendency clustered towards *oriC* (Figure 3B). However, in both cases the distribution of highly conserved core genes followed a more complex pattern. The genome of *C. indicus* showed a pronounced mosaic pattern with alternating regions of core and accessory genes. The core genes were concentrated around *oriC* (segments 1 and 2), *ter* (segment 5), and segments 3 and 6 in the middle of both replichores. A huge region of low conservation is found in segment 4. The *C. manganoxidans* genome shows a concentration of core genes on the left replichore, with low conserved regions concentrated in segments 2 to 5. Thus, for both genomes the decrease of conserved genes towards *ter* was the result of uneven distribution of core and accessory genes between replichores.

Influence of phages and HGT on architecture of *Rhodobacteraceae* genomes. We identified prophage sequences and regions putatively acquired through horizontal gene transfer and compared those to the clustering of core and accessory genes. The strains with the core genes shifted more towards the *ter* region had phages integrated near *oriC* (Figure 3A). The strains with the core genes more shifted towards *oriC* had phages integrated near *ter* (Figure 3B). Regarding all analyzed strains there was a significant enrichment in the absolute number of phages and the proportion of phage DNA near *ter* (Figure 4), confirming previous results (Oliveira *et al.*, 2017). Both methods used for HGT identification found overlapping regions of putative foreign origin. Genomic islands of *R. barguzinensis* and *C. manganoxidans* were found always closer to *oriC* and *ter*, respectively (Figure 3). Genomic islands of *L. vestfoldensis* and *C. indicus* were scattered throughout the genome with no preference towards *oriC* or *ter* (Figure 3). Not all regions with

low conservation were identified as horizontally transferred. In particular, only parts of the huge accessory genome regions in *R. barguzinensis* (segments 3 and 6) and *C. indicus* (segment 4) contained identifiable genomic islands. Interestingly, the number of HGT regions was on average higher closer to *oriC* (Figure 4). However a significant enrichment of HGT regions and the proportion of DNA within those was only found for the AlienHunter but not the IslandViewer results. In summary, phages and other sources of foreign DNA can only explain part of the observed core- and pan-genome clustering.

CONCLUSION

Our comparative genomic analysis revealed an unexpected bias in the clustering of homologous genes along the *oriC* → *terC* replication axis in several *Rhodobacteraceae* representatives. We observed the general trend that the part of the genome closer to *oriC* contains on average a higher density of core genes. This finding is in line with previous publications (Rocha, 2004, Touchon and Rocha, 2016, Oliveira *et al.*, 2017). However, we also identified remarkable exceptions to this trend, namely *L. vestfoldensis* and *R. barguzinensis*. Further investigation of the strains with the distribution of core genes most biased regarding distance to *oriC* revealed complex patterns with core genes often clustered in one half of one replichore. The analysis was restricted to *Rhodobacteraceae* with closed genomes. However, this subset contains represent strains of different genera from various habitats. Therefore it seems unlikely that the observed patterns will change when more genomes are included which would alter the core- and pan-genome content of the dataset.

The forces that may have driven the evolution of the observed pattern in this prokaryotic family remain to be elucidated. Selective gene loss alone cannot explain the huge regions containing

only accessory genes. The genome of the last common ancestor of the *Rhodobacteraceae* must be assumed unrealistically large to contain the pan-genome of this family (Dagan and Martin 2007). Gene loss and gene gain by HGT might both have contributed to the evolution of clustered genomes. The role that replication might have played during evolution has to be investigated in greater detail as it was possible here. However, our data clearly shows, that a model of preferential integration of transferred genes and phages at the terminus of replication, e.g. as compensation for dosage effects might not be generalizable.

ACKNOWLEDGMENTS

The authors thank Jason Dean BSc. for the language revision, and Dr. Yonghui Zeng for help with the phylogenetic analysis. This research has been supported by the GAČR project P501/12/G055, the DAAD project 57155424, and the EC-funded project Algatech Plus (LO1416). JT was supported by the Deutsche Forschungsgemeinschaft (DFG) within the Transregio TR51 Roseobacter.

AUTHOR CONTRIBUTION

JT, KK and MK designed the study. KK, YL, MJ and JT gathered and processed data. KK and JT analyzed the data with contributions from YL and MK. KK and JT wrote the manuscript with the help of MK.

REFERENCES

- Arndt D, et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44(W1):W16-W21.
- Bertelli C, et al. 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45(W1):W30-W35.
- Blattner FR, et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277(5331):1453-1462.
- Brinkmann H, Göker M, Koblížek M, Wagner-Döbler I, Petersen J. 2018. Horizontal operon transfer, plasmids, and the evolution of photosynthesis in Rhodobacteraceae. *ISME J.* 12(8): 1994-2010.
- Couturier E, Rocha EP. 2006. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol Microbiol.* 59(5):1506-1518.
- Dagan T, Martin W. 2007. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc Natl Acad Sci.* 104(3):870-875.
- Daubin V, Perriere G. 2003. G+ C3 structuring along the genome: a common feature in prokaryotes. *Mol Biol Evol.* 20(4):471-483.
- Eisen JA, Heidelberg JF, White O, Salzberg SL. 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* 1(6):research0011.1–0011.9
- Flynn KM, Vohr SH, Hatcher PJ, Cooper VS. 2010. Evolutionary rates and gene dispensability associate with replication timing in the archaeon *Sulfolobus islandicus*. *Genome Biol Evol.* 2:859-869.
- Freese HM, et al. 2017. Trajectories and drivers of genome evolution in surface-associated marine *Phaeobacter*. *Gen Biol Evol.* 9(12):3297-3311.
- Gao F, Zhang CT. 2008. Ori-Finder: a web-based system for finding oriC s in unannotated bacterial genomes. *BMC Bioinformatics* 9(1):79.
- Hsiao W, Wan I, Jones SJ, Brinkman FS. 2003. IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* 19(3):418-420.
- Kalhöfer D, et al. 2011. Comparative genome analysis and genome-guided physiological analysis of *Roseobacter litoralis*. *BMC Genomics* 12(1):324.

- Khedkar S, Seshasayee ASN. 2016. Comparative genomics of interreplichore translocations in bacteria: a measure of chromosome topology? *G3* 6(6):1597-1606.
- Kopejtka K, et al. 2017. Genomic analysis of the evolution of phototrophy among haloalkaliphilic *Rhodobacterales*. *Genome Biol Evol.* 9(7):1950-1962.
- Kopejtka K, et al. 2018. The complete genome sequence of *Rhodobaca barguzinensis* alga05 (DSM 19920) documents its adaptation for life in soda lakes. *Extremophiles* 22(6):839-849.
- Krause DJ, Didelot X, Cadillo-Quiroz H, Whitaker RJ. 2014. Recombination shapes genome architecture in an organism from the archaeal domain. *Genome Biol Evol.* 6(1):170-178.
- Kunst F, et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390(6657):249-256.
- Langille MG, Hsiao WW, Brinkman FS. 2008. Evaluation of genomic island predictors using a comparative genomics approach. *BMC Bioinformatics* 9(1):329.
- Lawrence JG, Ochman H. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci.* 95(16):9413-9417.
- Lechner M, et al. 2011. Proteinortho: detection of (co-) orthologs in large-scale analysis. *BMC Bioinformatics* 12(1):124.
- Luo H, Quan CL, Peng C, Gao F. 2018. Recent development of Ori-Finder system and DoriC database for microbial replication origins. *Brief Bioinform.* <https://doi.org/10.1093/bib/bbx174>
- Mackiewicz P, Zakrzewska-Czerwińska J, Zawilak A, Dudek MR, Cebrat S. 2004. Where does bacterial replication start? Rules for predicting the oriC region. *Nucleic Acids Res.* 32(13):3781-3791.
- Medini D, Donati C, Tettelin H, Massignani V, Rappuoli R. 2005. The microbial pan-genome. *Curr Opin Genet Dev.* 15(6):589-594.
- Oliveira PH, Touchon M, Cury J, Rocha EP. 2017. The chromosomal organization of horizontal gene transfer in bacteria. *Nat Commun.* 8(1):841.
- Petersen J, Frank O, Göker M, Pradella S. 2013. Extrachromosomal, extraordinary and essential—the plasmids of the *Roseobacter* clade. *Appl Microbiol Biotechnol.* 97(7):2805-2815.
- Pujalte MJ, Lucena T, Ruvira MA, Arahall DR, Macián MC. 2014. The Family *Rhodobacteraceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F

- (eds). The Prokaryotes. Springer, Berlin, Heidelberg. p. 439-512.
- Quast C, et al. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41(D1):D590-D596.
- Rocha EP. 2004. The replication-related organization of bacterial genomes. *Microbiology* 150(6):1609-1627.
- Sangar V, Blankenberg DJ, Altman N, Lesk AM. 2007. Quantitative sequence-function relationships in proteins based on gene ontology. *BMC Bioinformatics* 8(1):294.
- Sangar V, Blankenberg D J, Altman N, Lesk AM. 2007. Quantitative sequence-function relationships in proteins based on gene ontology. *BMC Bioinformatics* 8(1):294.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068-2069.
- Simon M, et al. 2017. Phylogenomics of *Rhodobacteraceae* reveals evolutionary adaptation to marine and non-marine habitats. *ISME J.* 11(6):1483.
- Sobetzko P, Travers A, Muskhelishvili G. 2012. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc Natl Acad Sci.* 109(2):E42-E50.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 56(4):564-577.
- Tamarit D, Neuvonen MM, Engel P, Guy L, Andersson SG. 2017. Origin and evolution of the *Bartonella* gene transfer agent. *Mol Biol Evol.* 35(2):451-464.
- Tettelin H, et al. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc Natl Acad Sci.* 102(39):13950-13955.
- Thole S, et al. 2012. *Phaeobacter gallaeciensis* genomes from globally opposite locations reveal high similarity of adaptation to surface life. *ISME J.* 6(12):2229.
- Touchon M, Rocha EP. 2016. Coevolution of the organization and structure of prokaryotic genomes. *Cold Spring Harb Perspect Biol.* 8(1):a018168.
- Vernikos GS, Parkhill J. 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* 22(18):2196-2203.
- Vollmers J, et al. 2013. Poles apart: Arctic and Antarctic *Octadecabacter* strains share high

genome plasticity and a new type of xanthorhodopsin. *PLoS One* 8(5):e63422.

Waack S, et al. 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* 7(1):142.

Figures

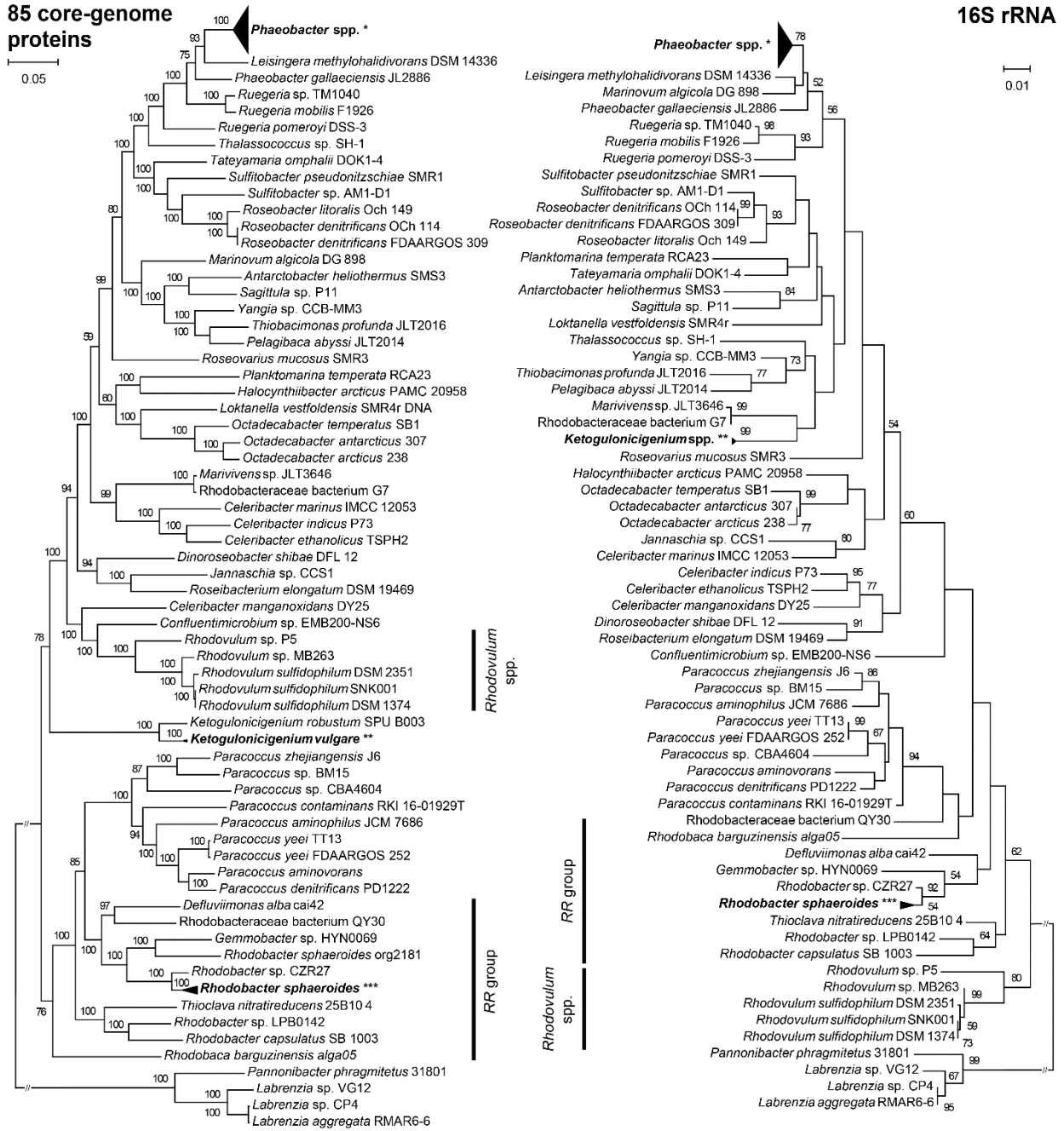


Figure 1 - Comparison of phylogenomic and 16S rRNA trees. Both trees comprise the same set of 109 *Rhodobacteraceae* strains. *Pannonibacter phragmitetus* 31801, *Labrenzia* sp. VG12, *Labrenzia* sp. CP4, and *Labrenzia aggregata* RMAR6-6 were used to root the trees as outgroup species. Scale bars represent changes per position. Bootstrap values >50% are shown. Bold vertical bars refer to different clustering patterns of the *Rhodovulum* spp. and *RR* (*Rhodobacter-*

Rhodobaca) group inside both trees. * collapsed *Phaeobacter* (*P.*) branches involve species *P. gallaciensis* (strains DSM 26640, P11, P63, P73, P75, P128, and P129), *P. inhibens* (strains 2.10, DOK1-1, DSM 17395, P10, P24, P30, P48, P51, P54, P57, P59, P66, P70, P72, P74, P78, P80, P83, P88, and P92), *P. piscinae* (strains P13, P14, P18, P23, P36, P42, and P71), and *P. porticola* P97; ** collapsed *Ketogulonicigenium vulgare* branches involve strains Hbe602, SKV, SPU B805, WSH-001, and Y25; *** collapsed *Rhodobacter sphaeroides* branches involve strains ATCC 17025, ATCC 17029, MBTLJ-8, MBTLJ-13, MBTLJ-20, and KD131 in both trees with additional strain org2181 in the 16S rRNA tree. Maximum-likelihood (ML) tree (left panel) based on concatenated alignments of amino acid sequences of the 85 most conserved core-genome proteins (27,668 common amino acid positions). The ML tree was calculated with 100 bootstrap replicates. 16S rRNA phylogenetic tree (right panel). Nucleotide sequences were aligned using ClustalX version 2.1 resulting in alignment with 1,260 common nucleotide positions after applying G-blocks. The phylogenetic tree was inferred using the ML algorithm with the HKY85 nucleotide substitution model and 1,000 bootstrap replicates. When possible, the strains were listed in the same order as in the phylogenomic tree.

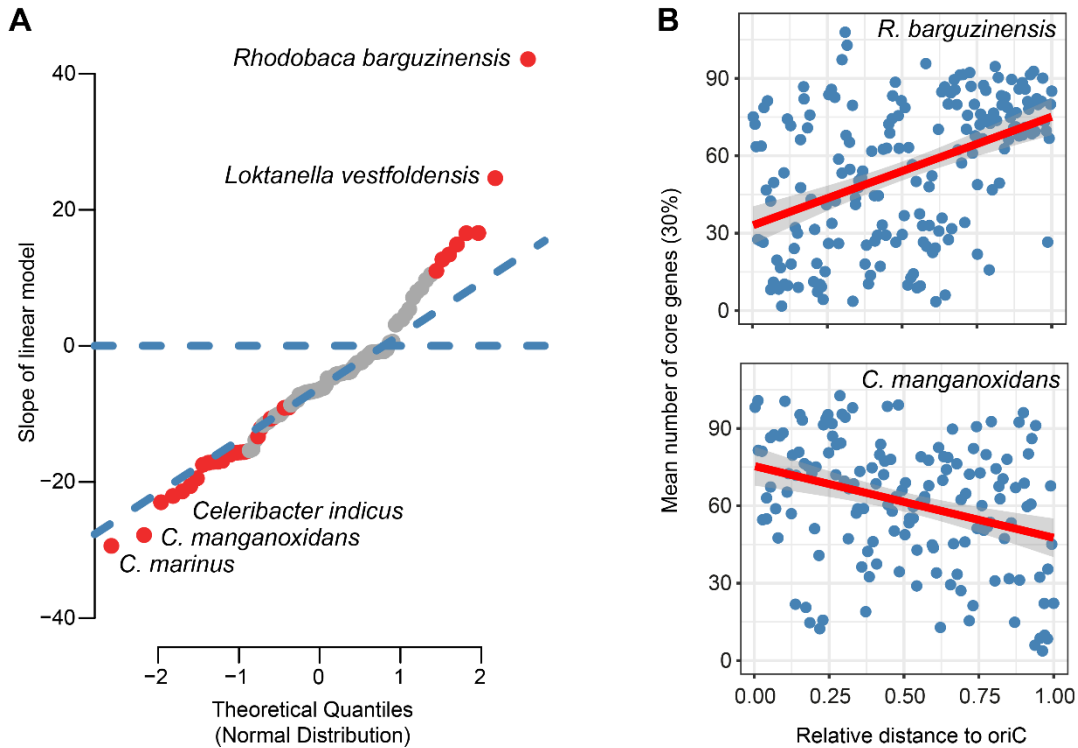


Figure 2 - Gradient in number of core genes with increasing distance from *oriC*. The average number of core genes and distance to the origin of replication was calculated for sliding windows of 20 genes and a linear model was fitted. **(A)** Quantil-quantil plot comparing the slope values extracted from the linear model of each genome to a theoretical normal distribution. Increasing slope values reflect the increase in mean number of core genes with increasing distance from *oriC*. Deviations from the normal distribution are indicated by increasing distance from the sloped blue line. The horizontal blue dashed line highlights the coordinate on the y-axis where the slope value is equal to 0. Red dots represent strains with slope values significantly different from 0 ($p < 0.05$). Names of two strains with the highest negative and positive slope values are shown. These two strains represent groups with different genome architecture. **(B)** Mean number of core genes compared to distance from origin of replication for *R. barguzinensis* (upper panel) and *C. manganoxidans* (lower panel). The linear function (red line) fitted to the data showed a significant increase (upper panel) or decline (lower panel) in mean number of core genes with increasing relative distance from *oriC*. Core genes were identified under less stringent conditions (sequence identity $\geq 30\%$, coverage $\geq 70\%$).

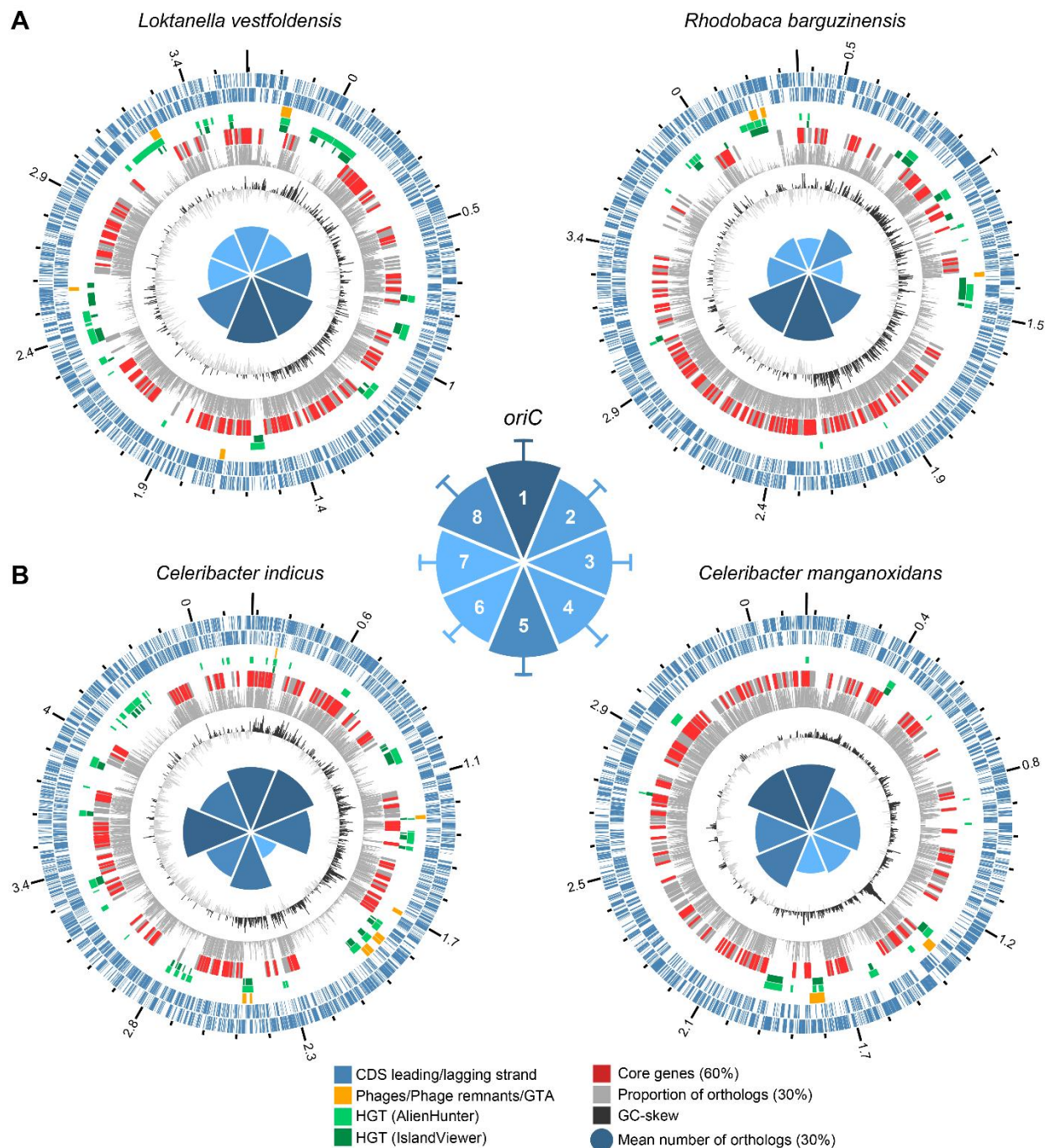


Figure 3 - Chromosome plots of four strains representing two different kinds of chromosome architecture in *Rhodobacteraceae*. (A) Two representatives from the group of strains for which the number of orthologs increases with the distance from the origin of replication. (B) Two representatives from the major group of strains in their genomes the number of orthologs decreases with the distance from the origin of replication. The outer to inner rings represent: scale of genome size in Mb and position of *oriC*; position of ORFs encoded on the plus strand; position of ORFs encoded on the minus strand; groups of HT genes as defined

in the graphical legend below; position of genes with orthologs in all 108 *Rhodobacteraceae* strains (red bars – highly conserved core genes (sequence identity $\geq 60\%$, coverage $\geq 80\%$); grey bars – less conserved core genes (sequence identity $\geq 30\%$, coverage $\geq 70\%$)); barchart displaying the proportion of strains in which orthologs of each representative's genes have been found (sequence identity $\geq 30\%$, coverage $\geq 70\%$); GC-skew; polar plot showing an average number of less conserved core genes in each quadrant (less stringent conditions). Polar plot in the middle: number of orthologs in each segment calculated as an average for all strains; the darker blue the higher number (sequence identity $\geq 30\%$, coverage $\geq 70\%$). See Supplementary Figure S for Tukey's HSD test for the eight segments.

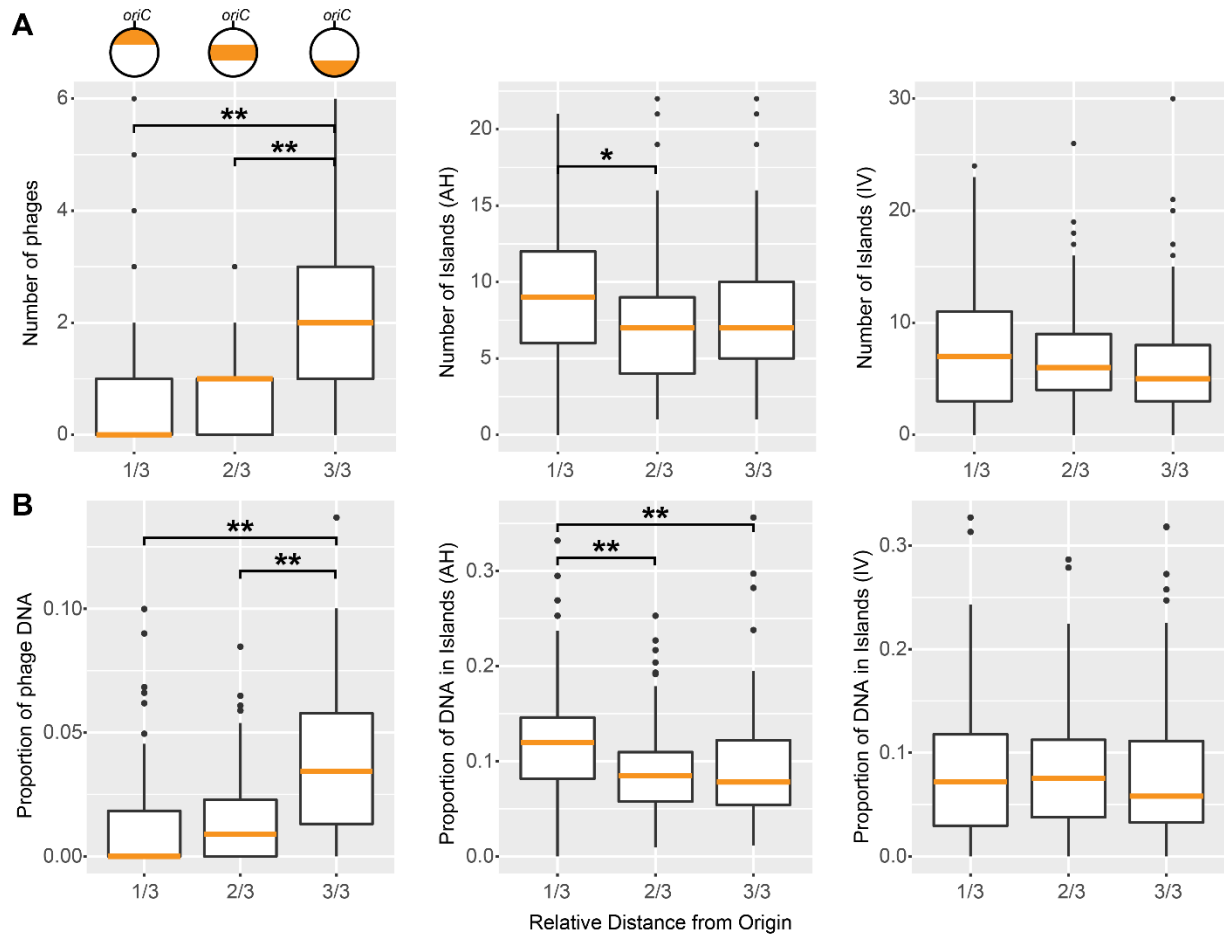


Figure 4 – Distribution of HT genes along the chromosome in 101 *Rhodobacteraceae*. (A) Mean numbers of phage regions identified by Phaster (phages, left panel), Genomic Islands identified by AlienHunter (AH, middle panel), and IslandViewer (IV, right panel) were calculated for each third of the chromosome. **(B)** Proportion of DNA found in phages or genomic islands, panel order as in (A). The orange horizontal lines represent median values. ANOVA was used to test for significant differences between the three parts of the chromosome. Asterisks indicate significant differences between comparisons identified using Tukey’s HSD test (* $p < 0.05$, ** $p < 0.01$).

Supplementary information

Clustered core- and pan-genome content on *Rhodobacteraceae* chromosomes

Karel Kopejtká^{1,2}, Yan Lin^{3,4}, Markéta Jakubovičová⁵, Michal Koblížek^{1,2}, Jürgen Tomasch^{6*}

¹*Laboratory of Anoxygenic Phototrophs, Center Algatech, Institute of Microbiology CAS,*

Třeboň, Czech Republic

²*Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic*

³*Department of Physics, School of Science, Tianjin University, Tianjin, China*

⁴*SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering, Tianjin, China*

⁵*Faculty of Information Technology, Czech Technical University in Prague*

⁶*Department of Molecular Bacteriology, Helmholtz Centre for Infection Research, Braunschweig, Germany*

**Author for correspondence: Juergen.Tomasch@helmholtz-hzi.de*

This file contains:

Legends for supplementary Figures S1, S2 and S4 and Supplementary Tables S1 and S2

Supplementary Figures S3, S5 and S6

Supplementary Figure S1, S2 and S4 are as separate PDF files on the included CD.

Supplementary Tables S1 and S2 are as separate Excel files on the included CD.

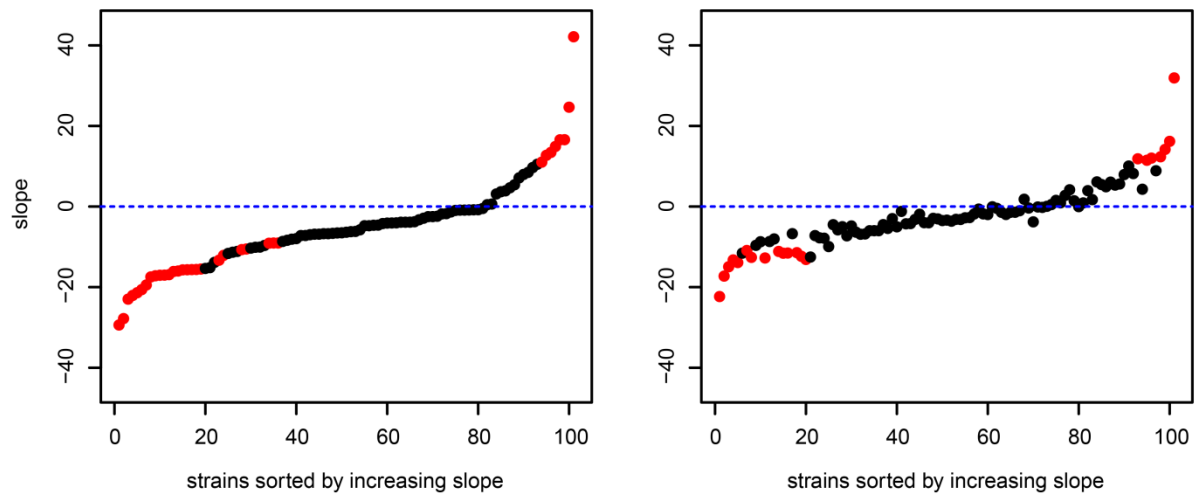
Supplementary Figure S1 – Number of core genes with increasing distance from *oriC* for all strains under less stringent conditions. The average number of core genes and distance to the origin of replication was calculated for sliding windows of 20 genes and a linear and a quadratic model were fitted. Confidence intervals of both fitted curves are shown. Core genes were identified under less stringent conditions (sequence identity $\geq 30\%$, coverage $\geq 70\%$).

Supplementary Figure S2 – Number of core genes with increasing distance from *oriC* for all strains under more stringent conditions. The average number of core genes and distance to the origin of replication was calculated for sliding windows of 20 genes and a linear and a quadratic model were fitted. Confidence intervals of both fitted curves are shown. Core genes were identified under more stringent conditions (sequence identity $\geq 60\%$, coverage $\geq 80\%$).

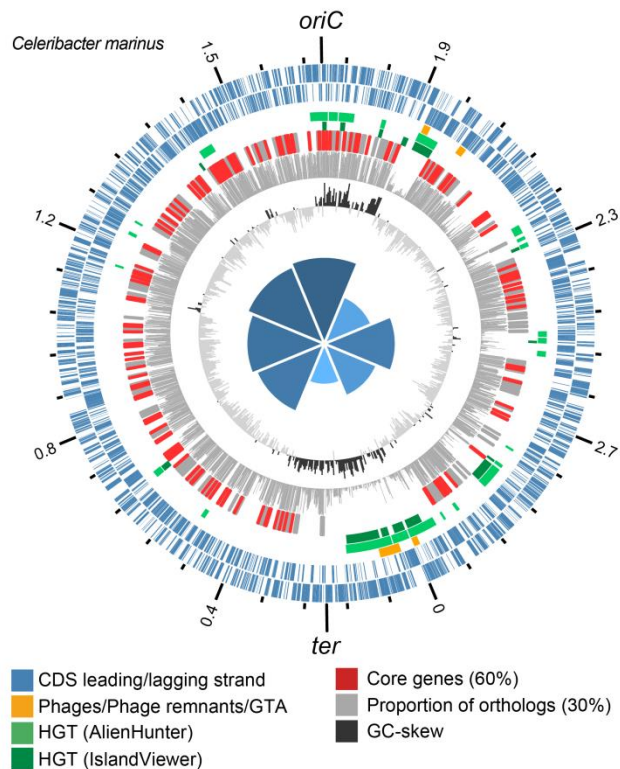
Supplementary Figure S4 – Circular representation of all 101 *Rhodobacteraceae* chromosomes with identified *oriC*. The outer to inner rings represent: scale of genome size in Mb and origin of replication (*oriC*); position of ORFs encoded on the plus and minus strand; groups of HT genes as defined in Figure 2; position of genes with orthologs in all 108 *Rhodobacteraceae* strains; barchart displaying the proportion of strains in which orthologs of *C. marinus* genes have been found (sequence identity $\geq 30\%$, coverage $\geq 70\%$); GC-skew. Note that chromosomes have not been reoriented towards the marked *oriC*.

Supplementary Table S1 – Accession numbers and characteristics of the 109 strains used for core- and pangenome analysis. Habitats are abbreviated: water (w), soil (s), clinical (c), freshwater (f) and marine (m).

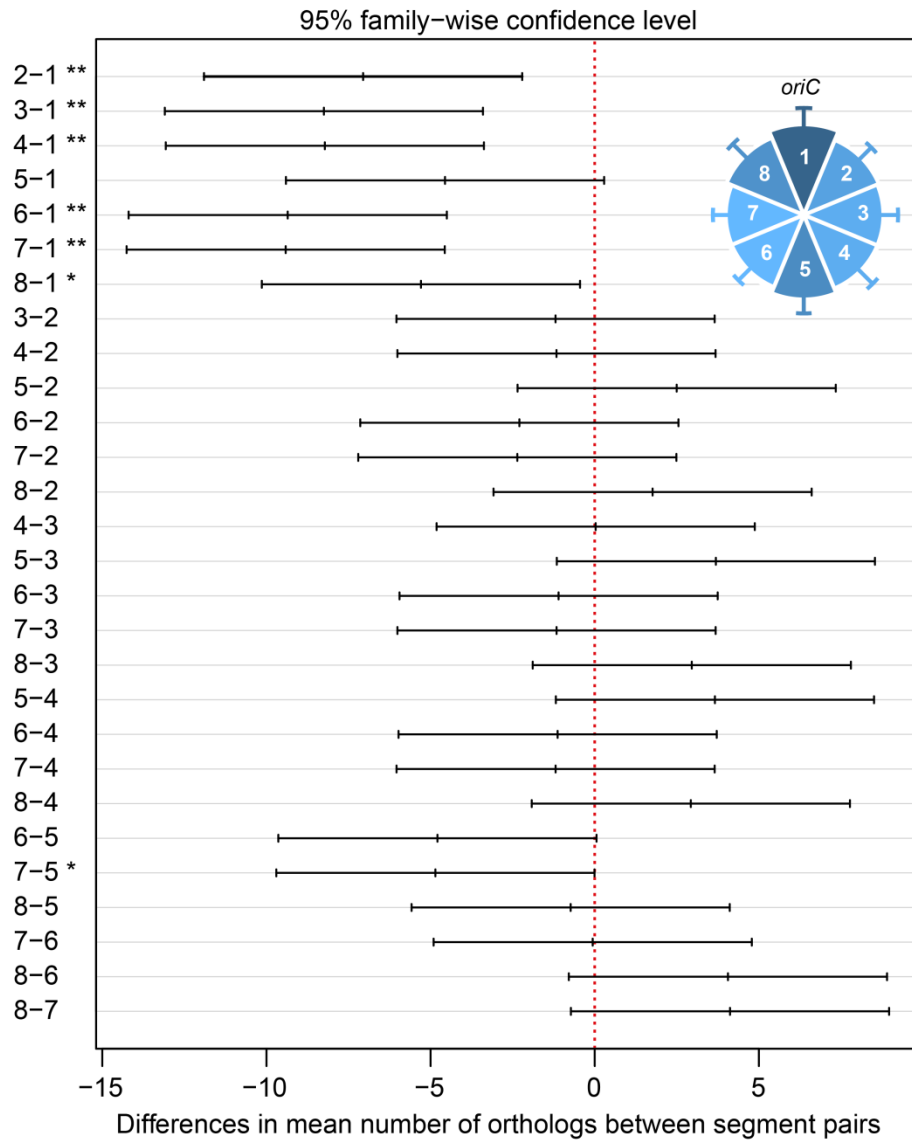
Supplementary Table S2 – Location of *oriC* and results of linear models for the 101 strains with identified *oriC*. Results of modelling are shown for using less stringent (lm30) and more stringent (lm60) identification of core genome as input.



Supplementary Figure S3 – Number of core genes with increasing distance from *oriC* for all strains under less stringent conditions. Core genes were identified under less stringent (left panel) and more stringent (right panel) conditions. The average number of core genes and distance to the origin of replication was calculated for sliding windows of 20 genes and a linear model was fitted. In both panels, strains are sorted by increasing slope for the less stringent conditions. Red dots represent strains with slope values significantly different from 0 ($p < 0.05$).



Supplementary Figure S5 – Circular representation of the *Celeribacter marinus* chromosome. The outer to inner rings represent: scale of genome size in Mb and origin of replication(*oriC*); position of ORFs encoded on the plus strand; position of ORFs encoded on the minus strand; groups of HT genes as defined in the graphical legend below; position of genes with orthologs in all 108 *Rhodobacteraceae* strains (red bars – highly conserved core genes (sequence identity $\geq 60\%$, coverage $\geq 80\%$); grey bars – less conserved core genes (sequence identity $\geq 30\%$, coverage $\geq 70\%$)); barchart displaying the proportion of strains in which orthologs of *C. marinus* genes have been found (sequence identity $\geq 30\%$, coverage $\geq 70\%$); GC-skew.



Supplementary Figure S6 – Result of Tukey’s HSD test comparing all pairs of segments.

Mean and 95% confidence interval of the differences in number of orthologs (30 % identity and 70% coverage) are plotted. Asterisks indicate significant differences between the compared segment pair (* $p < 0.05$, ** $p < 0.01$).

3. Conclusions

This PhD thesis elucidates the evolution of phototrophy among haloalkaliphilic representatives of the order *Rhodobacterales*. With the opportunity to sequence three closely related haloalkaliphilic bacteria from the order *Rhodobacterales* we specifically focussed on this subgroup of *Alphaproteobacteria*. We used the phylogenomic approach to study the evolution of phototrophy in this clade. Our original hypothesis was, that the wide metabolic diversity between closely related *Rhodobacterales* representatives is caused by differences in genes coding for regulatory proteins. Finally we did not find such differences in studied genomes and our data indicates that the evolution of phototrophy among the haloalkaliphilic members of the *Rhodobacter-Rhodobaca* group inside the *Rhodobacterales* order was regressive. Further, our research revealed, that the strain, *Rhodobaca (Rca.) barguzinensis*, which we used as a model organisms for our studies, is interesting not only phylogenetically, but also in terms of its unusual genome architecture. Phylogenetically interesting bacterium *Rca. barguzinensis* was used as a model organismus for this clade. The thesis consists of two first-author publications and one first-author manuscript under revision.

The main conclusions are as follows:

- *Rca. barguzinensis* is a metabolically flexible organism adopted to life in soda lake
- *mrpB* gene coding for the B subunit of the MRP Na⁺/H⁺ antiporter is crucial for adaptation to high pH
- *Rca. barguzinensis* represents a phylogenetically transition species between anaerobic and aerobic anoxygenic phototrophic bacteria
- phototrophy is the ancestral trophic mode among the haloalkaliphilic representatives of the *Rhodobacter-Rhodobaca* group inside the *Rhodobacterales* clade
- *Rca. barguzinensis* shows a remarkable bias in the clustering of homologous genes on its chromosome
- phage integration and horizontal gene transfer can to certain extent explain the scattered nature of *Rhodobacteraceae* genomes

4 References

- Alarico S, Rainey FA, Empadinhas N, Schumann P, Nobre MF, da Costa MS. 2002. *Rubritepida flocculans* gen. nov., sp. nov., a New Slightly Thermophilic Member of the α -1 Subclass of the Proteobacteria. *Syst Appl Microbiol.* 25(2):198-206.
- Androga DD, Özgür E, Eroglu I, Yücel M, Gündüz U. 2012. Photofermentative hydrogen production in outdoor conditions. Croatia: INTECH Open Access Publisher.
- Antony CP, Murrell JC, Shouche YS. 2012. Molecular diversity of methanogens and identification of *Methanolobus* sp. as active methylotrophic Archaea in Lonar crater lake sediments. *FEMS Microbiol Ecol.* 81(1):43-51.
- Archibald JM. 2009. The puzzle of plastid evolution. *Curr Biol.* 19(2):R81-R88.
- Awramik SM. 1992. The oldest records of photosynthesis. *Photosynth Res.* 33(2):75-89.
- Battistuzzi FU, Feijao A, Hedges SB. 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol.* 4(1):44.
- Beatty JT. 2002. On the natural selection and evolution of the aerobic phototrophic bacteria. *Photosynth Res.* 73(1-3):109-114.
- Bekker A, Holland HD, Wang PL, Rumble D, Stein HJ, Hannah JL, Coetsee LL, Beukes NJ. 2004. Dating the rise of atmospheric oxygen. *Nature* 427(6970):117-120.
- Biebl H, Allgaier M, Tindall BJ, Koblížek M, Lünsdorf H, Pukall R, Wagner-Döbler I. 2005. *Dinoroseobacter shibae* gen. nov., sp. nov., a new aerobic phototrophic bacterium isolated from dinoflagellates. *Int J Syst Evol Microbiol.* 55(3):1089-1096.
- Blankenship RE. 1992. Origin and early evolution of photosynthesis. *Photosynth Res.* 33(2):91-111.
- Blankenship RE. 2010. Early evolution of photosynthesis. *Plant Physiol.* 154(2):434-438.
- Boldareva EN, Akimov VN, Boychenko VA, Stadnichuk IN, Moskalenko AA, Makhneva ZK, Gorlenko VM. 2008. *Rhodobaca barguzinensis* sp. nov., a new alkaliphilic purple nonsulfur bacterium isolated from a soda lake of the Barguzin Valley (Buryat Republic, Eastern Siberia). *Microbiology* 77:206-218.
- Boscaro V, Felletti M, Vannini C, Ackerman MS, Chain PSG, Malfatti S, Vergez LM, Shin M, Doak TG, Lynch M, Petroni G. 2013. Polynucleobacter Necessarius, a Model for

- Genome Reduction in Both Free-Living and Symbiotic Bacteria. *Proc Natl Acad Sci.* 110.46(2013):18590–18595.
- Brinkmann H, Göker M, Koblížek M, Wagner-Döbler I, Petersen J. 2018. Horizontal operon transfer, plasmids, and the evolution of photosynthesis in *Rhodobacteraceae*. *ISME J.* 12:1994-2010.
- Brown JR, Doolittle WF. 1997. Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev.* 61(4):456-502.
- Brune DC. 1989. Sulfur oxidation by phototrophic bacteria. *Biochim. Biophys. Acta, Bioenerg.* 975(2):189-221.
- Brune DC. 1995. Sulfur compounds as photosynthetic electron donors. In: Blankenship RE, Madigan MT, Bauer CE (eds). *Anoxygenic photosynthetic bacteria. Advances in Photosynthesis and Respiration*, vol 2. Springer, Dordrecht. p. 847-870.
- Bryant DA, Frigaard NU. 2006. Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol.* 14(11):488-496.
- Bryant DA, Costas AMG, Maresca JA, Chew AGM, Klatt CG, Bateson MM, Tallon LJ, Hostetler J, Nelson WC, Heidelberg JF, Ward DM. 2007. *Candidatus Chloracidobacterium thermophilum*: an aerobic phototrophic acidobacterium. *Science* 317(5837):523-526.
- Carini SA, Joye SB. 2008. Nitrification in Mono Lake, California: Activity and community composition during contrasting hydrological regimes. *Limnol Oceanogr.* 53(6):2546-2557.
- Casjens S. 1998. The diverse and dynamic structure of bacterial genomes. *Ann Rev Genet.* 32(1):339-377.
- Chaconas G, Chen CW. 2005. Replication of linear bacterial chromosomes: no longer going around in circles. In: Higgins N (ed). *The Bacterial Chromosome*. ASM Press, Washington, DC. p. 525-540.
- Chisholm SW, Olson RJ, Zettler ER, Goericke R, Waterbury JB, Welschmeyer NA. 1988. A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* 334(6180):340-343.
- Cogdell RJ, Fyfe PK, Barrett ST, Prince SM, Freer AA, Isaacs NW, McGlynn P, Hunter CN. 1996. The purple bacterial photosynthetic unit. *Photosynth Res.* 48(1-2):55-63.
- Couturier E, Rocha EP. 2006. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol Microbiol.* 59(5):1506-1518.

- Dalton H. 2005. The Leeuwenhoek Lecture 2000 the natural and unnatural history of methane-oxidizing bacteria. *Philos Trans R Soc Lond B Biol Sci.* 360(1458):1207-1222.
- Di Giulio M. 2003. The universal ancestor and the ancestor of bacteria were hyperthermophiles. *J Mol Evol.* 57(6):721-730.
- Eisen JA, Heidelberg JF, White O, Salzberg SL. 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* 1(6):research0011.1–0011.9
- Ekblom R, Wolf JB. 2014. A field guide to whole-genome sequencing, assembly and annotation. *Evol Appl.* 7(9):1026-1042.
- Eme L, Spang A, Lombard J, Stairs CW, Ettema TJ. 2017. Archaea and the origin of eukaryotes. *Nat Rev Microbiol.* 15(12):711.
- Flannery DT, Walter MR. 2012. Archean tufted microbial mats and the Great Oxidation Event: new insights into an ancient problem. *Aust J Earth Sci.* 59(1):1-11.
- Flynn KM, Vohr SH, Hatcher PJ, Cooper VS. 2010. Evolutionary rates and gene dispensability associate with replication timing in the archaeon *Sulfolobus islandicus*. *Genome Biol Evol.* 2:859-869.
- Fraser NJ, Hashimoto H, Cogdell RJ. 2001. Carotenoids and bacterial photosynthesis: The story so far. *Photosynth Res.* 70(3):249-256.
- Fuchs BM, Spring S, Teeling H, Quast C, Wulf J, Schattenhofer M, Yan S, Ferriera S, Johnson J, Glöckner FO, Amann R. 2007. Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis. *Proc Natl Acad Sci.* 104(8):2891-2896.
- Garrity GM, Bell JA, Lilburn T. 2005. Order III. *Rhodobacterales* ord. nov. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM (eds). *Bergey's Manual® of Systematic Bacteriology*. New York: Springer US. p. 270-323.
- Gest H, Favinger JL. 1983. *Heliobacterium chlorum*, an anoxygenic brownish-green photosynthetic bacterium containing a “new” form of bacteriochlorophyll. *Arch Microbiol.* 136(1):11-16.
- Grant WD. 2006. Alkaline environments and biodiversity. Eolss Publishers Oxford, UK. p. 1-19.
- Gray MW. 1992. The endosymbiont hypothesis revisited. In: Wolstenholme DR, Jeon KW (eds). *International review of cytology*. Academic Press. p. 233-357.
- Gray MW. 1993. Origin and evolution of organelle genomes. *Curr Opin Genet Dev.* 3(6):884-890.
- Guzman MI, Martin ST. 2010. Photo-production of lactate from glyoxylate: how minerals can

- facilitate energy storage in a prebiotic world. *Chem Commun.* 46(13):2265-2267.
- Hauruseu D, Koblížek M. 2012. The influence of light on carbon utilization in aerobic anoxygenic phototrophs. *Appl Environ Microbiol.* 78(20):7414-7419.
- Hedges SB, Chen H, Kumar S, Wang DYC, Thompson AS, Watanabe H. 2001. A genomic timescale for the origin of eukaryotes. *BMC Evol Biol.* 1(1):4.
- Hellingwerf KJ, Crielgaard W, Hoff WD, Matthijs HCP, Mur LR, van Rotterdam BJ. 1994. Photobiology of bacteria. *Antonie van Leeuwenhoek* 65(4):331-347.
- Hoffman PF, Kaufman AJ, Halverson GP, Schrag DP. 1998. A Neoproterozoic snowball earth. *Science* 281(5381):1342-1346.
- Holland HD. 2006. The oxygenation of the atmosphere and oceans. *Philos Trans R Soc Lond B Biol Sci.* 361(1470):903-915.
- Jiang H, Dong H, Yu B, Lv G, Deng S, Wu Y, Dai M, Jiao N. 2009. Abundance and diversity of aerobic anoxygenic phototrophic bacteria in saline lakes on the Tibetan plateau. *FEMS Microbiol Ecol.* 67:268–278.
- Kado CI. 2015. Historical events that spawned the field of plasmid biology. In: Tomalsky M, Alonso J (eds). *Plasmids: Biology and Impact in Biotechnology and Discovery*. ASM Press, Washington, DC. p. 3-11.
- Karl D, Michaels A, Bergman B, Capone D, Carpenter E, Letelier R, Lipschultz F, Paerl H, Sigman D, Stal L. 2002. Dinitrogen fixation in the world's oceans. In: Boyer EW, Howarth RW (eds). *The Nitrogen Cycle at Regional to Global Scales*. Springer, Dordrecht. p. 47-98.
- Kasalický V, Zeng Y, Piwosz K, Šimek K, Kratochvílová H, Koblížek M. 2018. Aerobic Anoxygenic Photosynthesis Is Commonly Present within the Genus *Limnohabitans*. *Appl Environ Microbiol.* 84(1):e02116-17.
- Keppen OI, Krasil'nikova EN, Lebedeva NV, Ivanovskii RN. 2013. Comparative study of metabolism of the purple photosynthetic bacteria grown in the light and in the dark under anaerobic and aerobic conditions. *Microbiology* 82(5):547-553.
- Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, Beeby M, Yeates TO. 2005. Protein structures forming the shell of primitive bacterial organelles. *Science* 309(5736):936-938.
- Koblížek M, Bèjà O, Bidigare RR, Christensen S, Benitez-Nelson B, Vetriani C, Kolber MK, Falkowski PG, Kolber ZS. 2003. Isolation and characterization of *Erythrobacter* sp. strains from the upper ocean. *Arch Microbiol.* 180(5):327-338.
- Koblížek M, Mašín M, Ras J, Poulton AJ, Prášil O. 2007. Rapid growth rates of aerobic

- anoxygenic phototrophs in the ocean. *Environ Microbiol.* 9(10):2401-2406.
- Koblížek M, Zeng Y, Horák A, Oborník M. 2013. Regressive Evolution of Photosynthesis in the *Roseobacter* Clade. *Adv Biol Res.* 66:385-405.
- Koonin EV, Wolf YI. 2008. Genomics of Bacteria and Archaea: The Emerging Dynamic View of the Prokaryotic World. *Nucleic Acids Res.* 36(21):6688–6719.
- Kopejtko K, Tomasch J, Zeng Y, Tichý M, Sorokin DY, Koblížek M. 2017. Genomic analysis of the evolution of phototrophy among haloalkaliphilic *Rhodobacterales*. *Genome Biol Evol.* 9(7):1950-1962.
- Lancaster CRD, Michel H. 1996. Three-dimensional structures of photosynthetic reaction centers. *Photosynth Res.* 48(1-2):65-74.
- Land M, Hauser L, Jun S, Nookaew I, Leuze MR, Ahn T, Karpinets T, Lund O, Kora G, Wassenaar T, Poudel S, Ussery DW. 2015. Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* 15(2):141-161.
- Lang AS, Westbye AB, Beatty JT. 2017. The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. *Annu Rev Virol.* 4:87-104.
- Lanzén A, Simachew A, Gessesse A, Chmolowska D, Jonassen I, Øvreås L. 2013. Surprising prokaryotic and eukaryotic diversity, community structure and biogeography of Ethiopian soda lakes. *PLoS One.* 8(8):e72577.
- Liotenberg S, Steunou A, Picaud M, Reiss-Husson F, Astier C, Ouchane S. 2008. Organization and expression of photosynthesis genes and operons in anoxygenic photosynthetic proteobacteria. *Environ Microbiol.* 10(9):2267-2276.
- Macián M, Arahall DR, Garay E, Ludwig W, Schleifer KH, Pujalte MJ. 2005. *Thalassobacter stenotrophicus* gen. nov., sp. nov., a novel marine α -proteobacterium isolated from Mediterranean sea water. *Int J Syst Evol Microbiol.* 55(1): 105-110.
- Mašín M, Nedoma J, Pechar L, Koblížek M. 2008. Distribution of aerobic anoxygenic phototrophs in temperate freshwater systems. *Environ Microbiol.* 10(8):1988-1996.
- McFadden GI, van Dooren GG. 2004. Evolution: red algal genome affirms a common origin of all plastids. *Curr Biol.* 14(13):R514-R516.
- Morden CW, Delwiche CF, Kuhsel M, Palmer JD. 1992. Gene phylogenies and the endosymbiotic origin of plastids. *Biosystems* 28(1–3):75-90.
- Mulkidjanian A, Galperin M. 2009. On the origin of life in the zinc world. 2. Validation of the hypothesis on the photosynthesizing zinc sulfide edifices as cradles of life on Earth. *Biol Direct* 4(1):27.

- Mulkiđjanian AY, Bychkov AY, Dibrova DV, Galperin MY, Koonin EV. 2012. Origin of first cells at terrestrial, anoxic geothermal fields. *Proc Natl Acad Sci.* 109(14):E821-E830.
- Olson JM, Blankenship RE. 2004. Thinking about the evolution of photosynthesis. *Photosynth Res.* 80(1-3):373-386.
- Oren A. 1999. Bioenergetic aspects of halophilism. *Microbiol Mol Biol Rev.* 63(2):334-348.
- Parte AC. 2018. LPSN-List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *Int J Syst Evol Microbiol.* 68(6):1825-1829.
- Pierson BK, Castenholz RW. 1974. A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. *Arch Microbiol.* 100(1):5-24.
- Poole AM, Penny D. 2007. Evaluating hypotheses for the origin of eukaryotes. *Bioessays* 29(1):74-84.
- Rainey FA, Silva J, Nobre MF, Silva MT, da Costa MS. 2003. *Porphyrobacter cryptus* sp. nov., a novel slightly thermophilic, aerobic, bacteriochlorophyll a-containing species. *Int J Syst Evol Microbiol.* 53(1):35-41.
- Rappé MS, Giovannoni SJ. 2003. The uncultured microbial majority. *Annu Rev Microbiol.* 57(1):369-394.
- Rathgeber C, Beatty JT, Yurkov V. 2004. Aerobic phototrophic bacteria: new evidence for the diversity, ecological importance and applied potential of this previously overlooked group. *Photosynth Res.* 81(2):113-128.
- Raymond J, Zhaxybayeva O, Gogarten JP, Gerdes SY, Blankenship RE. 2002. Whole-genome analysis of photosynthetic prokaryotes. *Science* 298(5598):1616-1620.
- Rocha EP. 2004. The replication-related organization of bacterial genomes. *Microbiology* 150(6):1609-1627.
- Rosengarten R, Citti C, Glew M, Lischewski A, Droebe M, Much P, Winner F, Brank M, Spargser J. 2000. Host-pathogen interactions in mycoplasma pathogenesis: virulence and survival strategies of minimalist prokaryotes. *Int J Med Microbiol.* 290(1):15-25.
- Saitoh SI, Inagake D, Sasaoka K, Ishizaka J, Nakame Y, Saino T. 1998. Satellite and ship observations of Kuroshio warm-core ring 93A off Sanriku, northwestern North Pacific, in spring 1997. *J Oceanogr.* 54(5):495-508.
- Sánchez-Baracaldo P, Raven JA, Pisani D, Knoll AH. 2017. Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proc Natl Acad Sci.* 114(37): E7737-E7745.
- Shiba T, Simidu U, Taga N. 1979. Distribution of aerobic bacteria which contain bacteriochlorophyll a. *Appl Environ Microbiol.* 38(1):43-45.
- Shields-Zhou G, Och L. 2011. The case for a neoproterozoic oxygenation event: geochemical

- evidence and biological consequences. *GSA Today* 21(3):4-11.
- Simon M, Scheuner C, Meier-Kolthoff JP, Brinkhoff T, Wagner-Döbler I, Ulbrich M, Klenk HP, Schomburg D, Petersen J, Göker M. 2017. Phylogenomics of *Rhodobacteraceae* reveals evolutionary adaptation to marine and non-marine habitats. *ISME J.* 11(6):1483.
- Sleep NH. 2010. The Hadean-Archaeon Environment. *Cold Spring Harb Perspect Biol.* 2(6):a002527.
- Sobetzko P, Travers A, Muskhelishvili G. 2012. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc Natl Acad Sci.* 109(2):E42-E50.
- Sorokin DY. 1998. Occurrence of nitrification in extremely alkaline natural habitats. *Microbiology* 67:404-408.
- Sorokin DY, Berben T, Melton ED, Overmars L, Vavourakis CD, Muyzer G. 2014. Microbial diversity and biogeochemical cycling in soda lakes. *Extremophiles* 18(5):791-809.
- Söhngen C, Bunk B, Podstawka A, Gleim D, Overmann J. 2013. BacDive - the Bacterial Diversity Metadatabase. *Nucleic Acids Res.* 42(D1):D592-D599.
- Stevens RJ, Laughlin RJ, Malone JP. 1998. Soil pH affects the processes reducing nitrate to nitrous oxide and di-nitrogen. *Soil Biol Biochem.* 30(8):1119-1126.
- Surakasi VP, Antony CP, Sharma S, Patole MS, Shouche YS. 2010. Temporal bacterial diversity and detection of putative methanotrophs in surface mats of Lonar crater lake. *J Basic Microbiol.* 50(5):465-474.
- Suwanto A, Kaplan S. 1992. Chromosome transfer in *Rhodobacter sphaeroides*: Hfr formation and genetic evidence for two unique circular chromosomes. *J Bacteriol.* 174(4):1135-1145.
- Suyama T, Shigematsu T, Suzuki T, Tokiwa Y, Kanagawa T, Nagashima KVP, Hanada S. 2002. Photosynthetic apparatus in *Roseateles depolymerans* 61A is transcriptionally induced by carbon limitation. *Appl Environ Microbiol.* 68(4):1665-1673.
- Suzuki T, Muroga Y, Takahama M, Nishimura Y. 1999. *Roseivivax halodurans* gen. nov., sp. nov. and *Roseivivax halotolerans* sp. nov., aerobic bacteriochlorophyll-containing bacteria isolated from a saline lake. *Int J Syst Bacteriol.* 49(2):629-634.
- Swingley WD, Blankenship RE, Raymond J. 2008. Evolutionary relationships among purple photosynthetic bacteria and the origin of proteobacterial photosynthetic systems. In: Hunter CN, Daldal F, Thurnauer MC, Beatty JT (eds). *The Purple Phototrophic Bacteria. Advances in Photosynthesis and Respiration*, vol 28. Springer, Dordrecht. p. 17-29.

- Tomasch J, Wang H, Hall ATK, Patzelt D, Preusse M, Petersen J, Brinkmann H, Bunk B, Bhujji S, Jarek M, Geffers R, Lang AS, Wagner-Döbler I. 2018. Packaging of *Dinoroseobacter shibae* DNA into Gene Transfer Agent particles is not random. *Genome Biol Evol.* 10(1):359-369.
- Toussaint A, Chandler M. 2012. Prokaryote genome fluidity: toward a system approach of the mobilome. In: van Helden J, Toussaint A, Thieffry D (eds). *Bacterial Molecular Networks. Methods in Molecular Biology (Methods and Protocols)*, vol 804. New York, Springer US. p. 57-80.
- Treangen TJ, Rocha EP. 2011. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* 7(1):e1001284.
- Vincent AT, Derome N, Boyle B, Culley AI, Charette SJ. 2017. Next-generation sequencing (NGS) in the microbiological world: how to make the most of your money. *J Microbiol Methods* 138:60-71.
- Volff JN, Altenbuchner J. 2000. A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution. *FEMS Microbiol Lett.* 186(2):143-150.
- Waidner LA, Kirchman DL. 2005. Aerobic anoxygenic photosynthesis genes and operons in uncultured bacteria in the Delaware River. *Environ Microbiol.* 7(12):1896-1908.
- Wakao N, Yokoi N, Isoyama N, Hiraishi A, Shimada K, Kobayashi M, Kise H, Iwaki M, Itoh S, Takaichi S, Sakurai Y. 1996. Discovery of natural photosynthesis using Zn-containing bacteriochlorophyll in an aerobic bacterium *Acidiphilium rubrum*. *Plant Cell Physiol.* 37(6):889-893.
- Wang J, Yang D, Zhang Y, Shen J, van der Gast C, Hahn MW, Wu Q. 2011. Do patterns of bacterial diversity along salinity gradients differ from those observed for macroorganisms? *PLoS One.* 6(11):e27597.
- Waterbury JB, Watson SW, Guillard RR, Brand LE. 1979. Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* 277:293-294.
- Wilmes P, Simmons SL, Deneff VJ, Banfield JF. 2008. The dynamic genetic repertoire of microbial communities. *FEMS Microbiol Rev.* 33(1):109-132.
- Woese CR. 1987. Bacterial evolution. *Microbiol Rev.* 51(2):221.
- Xiong J, Fischer WM, Inoue K, Nakahara M, Bauer CE. 2000. Molecular evidence for the early evolution of photosynthesis. *Science* 289(5485):1724-1730.
- Xiong J, Bauer CE. 2002. Complex evolution of photosynthesis. *Annu Rev Plant Biol.* 53(1):503-521.
- Xiong J, Liu Y, Lin X, Zhang H, Zeng J, Hou J, Yang Y, Yao T, Knight R, Chu H. 2012.

- Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Environ Microbiol.* 14(9):2457-2466.
- Yan S, Wu G. 2016. Analysis on evolutionary relationship of amylases from archaea, bacteria and eukaryota. *World J Microbiol Biotechnol.* 32(2):24.
- Yarza P, Richter M, Peplies J, Euzéby J, Amann R, Schleifer K, Ludwig W, Glöckner FO, Rosselló-Móra R. 2008. The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol.* 31(4):241-250.
- Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol.* 12(9):635.
- Yurkov VV, Gorlenko VM, Kompantseva EI. 1992. A new type of freshwater aerobic orange-colored bacterium *Erythromicrobium* gen. nov., containing bacteriochlorophyll a. *Microbiology* 61(2):169-172.
- Yurkov VV, van Gemerden H. 1993. Impact of light/dark regimen on growth rate, biomass formation and bacteriochlorophyll synthesis in *Erythromicrobium hydrolyticum*. *Arch Microbiol.* 159(1):84-89.
- Yurkov VV, Beatty JT. 1998. Aerobic anoxygenic phototrophic bacteria. *Microbiol Mol Biol Rev.* 62(3):695-724.
- Yurkov VV, Krieger S, Stackebrandt E, Beatty JT. 1999. *Citromicrobium bathyomarinum*, a novel aerobic bacterium isolated from deep-sea hydrothermal vent plume waters that contains photosynthetic pigment-protein complexes. *J Bacteriol.* 181(15):4517-4525.
- Yurkov V, Csotonyi JT. 2003. Aerobic anoxygenic phototrophs and heavy metalloid reducers from extreme environments. *Recent Res Devel Bacteriol.* 1:247-300.
- Yurkov V. 2006. Aerobic phototrophic proteobacteria. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds). *The Prokaryotes*. Springer US, New York. p. 562-584.
- Yurkov V, Csotonyi JT. 2009. New light on aerobic anoxygenic phototrophs. In: Hunter CN, Daldal F, Thurnauer MC, Beatty JT (eds). *The purple phototrophic bacteria. Advances in photosynthesis and respiration*, vol 28. Springer, Dordrecht. p. 31-55.
- Zehr JP, Waterbury JB, Turner PJ, Montoya JP, Omoregie E, Steward GF, Hansen A, Karl DM. 2001. Unicellular cyanobacteria fix N₂ in the subtropical North Pacific Ocean. *Nature* 412(6847):635.
- Zeng Y, Feng F, Medová H, Dean J, Koblížek M. 2014. Functional type 2 photosynthetic

- reaction centers found in the rare bacterial phylum Gemmatimonadetes. *Proc Natl Acad Sci.* 111(21):7795-7800.
- Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol.* 18(6):292-298.
- Zhang XV, Ellery SP, Friend CM, Holland HD, Michel FM, Schoonen MAA, Martin ST. 2007. Photodriven reduction and oxidation reactions on colloidal semiconductor particles: Implications for prebiotic synthesis. *J Photochem Photobiol, A.* 185(2): 301-311.
- Zheng Q, Zhang R, Koblížek M, Boldareva EN, Yurkov V, Yan S, Jiao N. 2011. Diverse arrangement of photosynthetic gene clusters in aerobic anoxygenic phototrophic bacteria. *PloS One* 6(9):e25050.

Curriculum vitae - MSc. Karel Kopejtko

Born	December 30th, 1984 in Pelhřimov, Czech Republic
Education	since 2013, Ph.D. studies in Molecular and Cell Biology and Genetics, Faculty of Science, University of South Bohemia in České Budějovice, Czech Republic 2011-2013, Master studies in Clinical Biology, Faculty of Science, University of South Bohemia in České Budějovice, Czech Republic; master thesis: “The role of horizontal gene transfer in disseminating tetracycline resistance among bacteria in farm soil“; supervisor: Mgr. Martina Kyselková, PhD. 2008-2011, Bachelor studies in Biomedical Laboratory Techniques, Faculty of Science, University of South Bohemia in České Budějovice, Czech Republic; bachelor thesis: “Assesment of molecular methods for study of social organisation in ant colonies“; supervisor: Mgr. Milan Janda, PhD.
Experience	since 2013, Laboratory of Anoxygenic Phototrophs, Institute of Microbiology CAS, Centre Algatech, Třeboň, Czech Republic 2011-2013, Department of soil microbiology and soil chemistry, Institute of Soil Biology, Biology Centre AS CR v.v.i., České Budějovice, Czech Republic 2009-2011, Laboratory of Ecology and Evolution of Social Insects, Department of Ecology, Institute of Entomology, Biology Centre AS CR v.v.i., České Budějovice, Czech Republic
Fields of interest	evolution of bacterial photosynthesis, horizontal gene transfer, microbial genome evolution, aquatic microbial ecology
Languages	Czech (mother tongue), English (fluent), German (B1 level)

Publications

- Kopejtko K, Lin Y, Jakubovičová M, Koblížek M, Tomasch J. Clustered core- and pan-genome content on *Rhodobacteraceae* chromosomes. *Genome Biol Evol.*, under review.
- Kaftan D, Medová H, Selyanin V, Kopejtko K, Koblížek M. Extended temperature optimum of photosynthetic reaction centers in *Rhodobacterales*. *Photosynthetica*, in press.
- Strunecký O, Kopejtko K, Goecke F, Tomasch J, Lukavský J, Neori A, Kahl S, Pieper DH, Pilarsky P, Kaftan D, Koblížek M. 2019. High diversity of thermophilic cyanobacteria in Rupite hot spring identified by microscopy, cultivation, single-cell PCR and amplicon sequencing. *Extremophiles* 23(1):35-48.
- Kopejtko K, Tomasch J, Bunk B, Spröer C, Wagner-Döbler I, Koblížek M. 2018. The complete genome sequence of *Rhodobaca barguzinensis* alga05 (DSM 19920) documents its adaptation for the life in soda lakes. *Extremophiles* 22(6):839-849.
- Piwosz K, Shabarova T, Tomasch J, Šimek K, Kopejtko K, Kahl S, Pieper DH, Koblížek M. 2018. Determining lineage-specific bacterial growth curves with a novel approach based on amplicon reads normalization using internal standard (ARNIS). *ISME J.* 12(11):2640-2654.
- Kopejtko K, Tomasch J, Zeng Y, Tichý M, Sorokin DY, Koblížek M. 2017. Genomic analysis of the evolution of phototrophy among haloalkaliphilic *Rhodobacterales*. *Genome Biol Evol.* 9(7):1950-1962.
- Kopejtko K, Koblížek M. 2016. Bojovník s epidemiemi. *Vesmír* 95(7):440. (science popularization article in Czech)