

Evidence of microbial rhodopsins in Antarctic Dry Valley edaphic systems

Leandro D. Guerrero, Surendra Vikram, Thulani P. Makhwanyane and Don A. Cowan*

Centre of Microbial Ecology and Genomics, Department of Genetics, University of Pretoria, Pretoria, South Africa.

* Corresponding Author: Prof. Don A. Cowan,

Centre for Microbial Ecology and Genomics,

University of Pretoria,

Lynwood Road, 0028 Pretoria,

South Africa

Email: don.cowan@up.ac.za

Running title: Microbial rhodopsins in Antarctic Dry Valley soils

Originality-Significance Statement

Hypolithic microbial communities are important contributors to ecosystem services in hot and cold oligotrophic desert soils. Here we present a functional metagenomic study of Antarctic Dry Valley soils and demonstrate, for the first time, the presence of a substantial diversity of microbial rhodopsins in an edaphic system. The proportion of bacterial species with rhodopsins was similar to that found in marine environments, where rhodopsins are the dominant photosynthetic proteins. The assembly of three near-complete microbial genomes harbouring rhodopsins allows us to infer their ecological relevance in this extreme desert environment.

Summary

Microorganisms with an ability to synthesize rhodopsins have a capacity to translocate ions through their membranes, using solar energy to generate a proton motive force. Rhodopsins are the most abundant phototrophic proteins in oceanic surface waters and are key constituents in marine bacterial ecology. However, it remains unclear how rhodopsins are used in most microorganisms. Despite their abundance in marine and fresh-water systems, the presence of functional rhodopsin systems in edaphic habitats has never been reported. Here, we show the presence of several new putative H^+ , Na^+ and Cl^+ pumping rhodopsins identified by metagenomic analysis of Antarctic desert hypolithic communities. Reconstruction of two Proteobacteria genomes harboring xanthorhodopsin-like proteins and one Bacteroidetes genome with a Na-pumping-like rhodopsin indicated that these bacteria were aerobic heterotrophs possessing the apparent capacity for the functional expression of rhodopsins. The existence of these protein systems in hypolithic bacteria expands the known role of rhodopsins to include terrestrial environments and suggests a possible predominant function as heterotrophic energy supply proteins, a feasible microbial adaptation to the harsh conditions prevalent in Antarctic edaphic systems.

Introduction

In some oligotrophic environments such as surface pelagic waters, microbial photoautotrophy is the dominant physiology and microbial rhodopsins are the primary light capture structures (Moran and Miller, 2007). The light-driven ion pump rhodopsins are transmembrane proteins, which absorb light and generate a transmembrane gradient of H^+ , Na^+ or Cl^- (Inoue *et al.*, 2015). The transmembrane gradient is usually exploited to provide biochemical energy (via proton-driven ATPase activity), for secondary transport or to maintain osmotic balance (Fuhrman *et al.*, 2008). Most rhodopsins use a retinal protein as a chromophore to harvest light, with the exception of xanthorhodopsin (XR) that also uses a carotenoid molecule to expand the spectral adsorption range (Balashov *et al.*, 2005). Rhodopsins are usually associated with a response to survive periods of low nutrient availability (Gómez-Consarnau *et al.*, 2010; Kimura *et al.*, 2011; Steindler *et al.*, 2011). However, recent studies suggest that rhodopsin activity may also be related to diverse environmental factors such as salinity (Feng *et al.*, 2013), vitamin auxotrophy (Gómez-Consarnau *et al.*, 2015) or the availability of a particular substrate (Xing *et al.*, 2015). Although the benefits conferred by rhodopsins on microorganisms remain poorly understood, the ability to synthesize these photoactive compounds appears to represent a significant adaptive feature in oligotrophic environments (Gómez-Consarnau *et al.*, 2010; Kimura *et al.*, 2011).

Rhodopsins are widely distributed in oceans and are also present in other habitats where water is abundant (Béjà *et al.*, 2000; Finkel *et al.*, 2013). However, the presence of rhodopsins in non-aquatic environments has not been widely reported, and is restricted to a limited number of studies on isolated edaphic bacterial species (Nakamura *et al.*, 2003; Gushchin *et al.*, 2013) and plant leaf surface communities (Atamna-Ismaeel *et al.*, 2012). It has been concluded that these proteins have evolved in, and are mostly limited to, aquatic environments (Finkel *et al.*, 2013).

In extreme hot and cold desert soil environments, which are largely or completely devoid of higher plants, microbial communities on the ventral surfaces of translucent rocks, known as

hypolithons, are considered to be the primary contributors to ecosystem services (Pointing *et al.*, 2009; Cary *et al.*, 2010). Hypolithons are widely distributed in Antarctic soils, particularly in the McMurdo Dry Valleys of eastern Antarctica (Cary *et al.*, 2010; Cowan *et al.*, 2010), and their microbial community compositions have been well characterized using 16S rRNA gene phylogenetics and microscopy (Pointing *et al.*, 2009; de los Ríos *et al.*, 2014). Photosynthetic cyanobacteria are the dominant primary producers in these niche communities (Pointing *et al.*, 2009; Chan *et al.*, 2012), but heterotrophic phyla (particularly Proteobacteria and Actinobacteria) are also present as major components of these assemblages (Cary *et al.*, 2010; Chan *et al.*, 2013; Makhalanyane *et al.*, 2013). Although several autotrophic and heterotrophic strategies, including pathways for chemoautotrophy and diazotrophy, have been identified in hypolithic taxa (Chan *et al.*, 2013), the functional capacities of the members on this community have not been extensively studied and remain poorly understood. The ability of Antarctic edaphic communities (and hypolithic prokaryotic taxa in particular) to exploit solar energy by means of rhodopsins has, to our knowledge, never been reported. The presence of these systems in Antarctic terrestrial environments could change the understanding of the overall energy flux in hypolithic microbial food webs.

Results

Hypolith Metagenome

We conducted shotgun sequencing of metagenomic DNA extracted from hypolithic microbial biomass recovered from the hyperarid east Antarctic Miers Valley (78.1600 S, 164.1000 E) (Makhalanyane *et al.*, 2013). Hypolithic samples were collected, using standard aseptic techniques, from the upper regions of the valley, some 10km from the coast, as described previously (Cowan *et al.*, 2011). None of the sample sites are impacted by marine mammal activity (Cowan DA, pers. comm.). These sub-lithic communities are thought to be stable

assemblages over decades or centuries (Cowan et al., 2011) and phylogenetic analyses show no significant marine signatures (Khan et al., 2011).

The Antarctic hypolithon metagenomic sequence dataset comprised approximately 158 million paired-end reads of which 48.5% were assembled into 316,555 contigs (≥ 500 bp), representing 480 Mbp of DNA sequence. Taxonomic analysis of reconstructed rRNA small subunit sequence (OTU 97% similarity) and single copy marker genes in the metagenomic contigs were used to estimate the abundance of the bacterial taxa present in the sample. Both methods showed that Actinobacteria, Bacteroidetes and Proteobacteria were present as the major phyla (**Supplementary Figures 1 to 3**). Some differences in the abundance estimates could be attributed to the two different approaches used (i.e., different software and databases), the bias introduced by the rRNA operon copy number (Farrelly *et al.*, 1995) and the high number of unassigned sequences for the single copy marker genes (more than 25% for some genes) (**Supplementary Fig. 3**).

In order to identify putative rhodopsins present in the Antarctic hypolith metagenome, a pool of sequences of known microbial rhodopsins was used to search against the 316,555 contigs (≥ 500 bp) of the metagenome. Using this approach, 29 sequences comprising most of the rhodopsin functional residues ($>65\%$) were detected (**Fig. 1 and Supplementary Table 1**). The presence of key conserved amino acids suggested that all rhodopsins were potentially functional (Inoue *et al.*, 2015). The identified rhodopsin sequences could be assigned to four different rhodopsins groups: 1 sequence was classified as proteorhodopsin (PR), 15 sequences were classified as XR-like and 13 as Na^+ or Cl^- pumping rhodopsins (NQ). Of the latter, 9 were classified as Na^+ rhodopsins (NDQ), characterized by the presence of conserved Asn85, Asp89 and Gln96 residues, whereas 4 were classified as Cl^- rhodopsins (NTQ), where a threonine replaces the aspartic acid at position 89 (Yoshizawa *et al.*, 2014) (**Supplementary Table 1**). As an estimate of the abundance of bacteria containing rhodopsin genes, the number of rhodopsins was compared with the number of

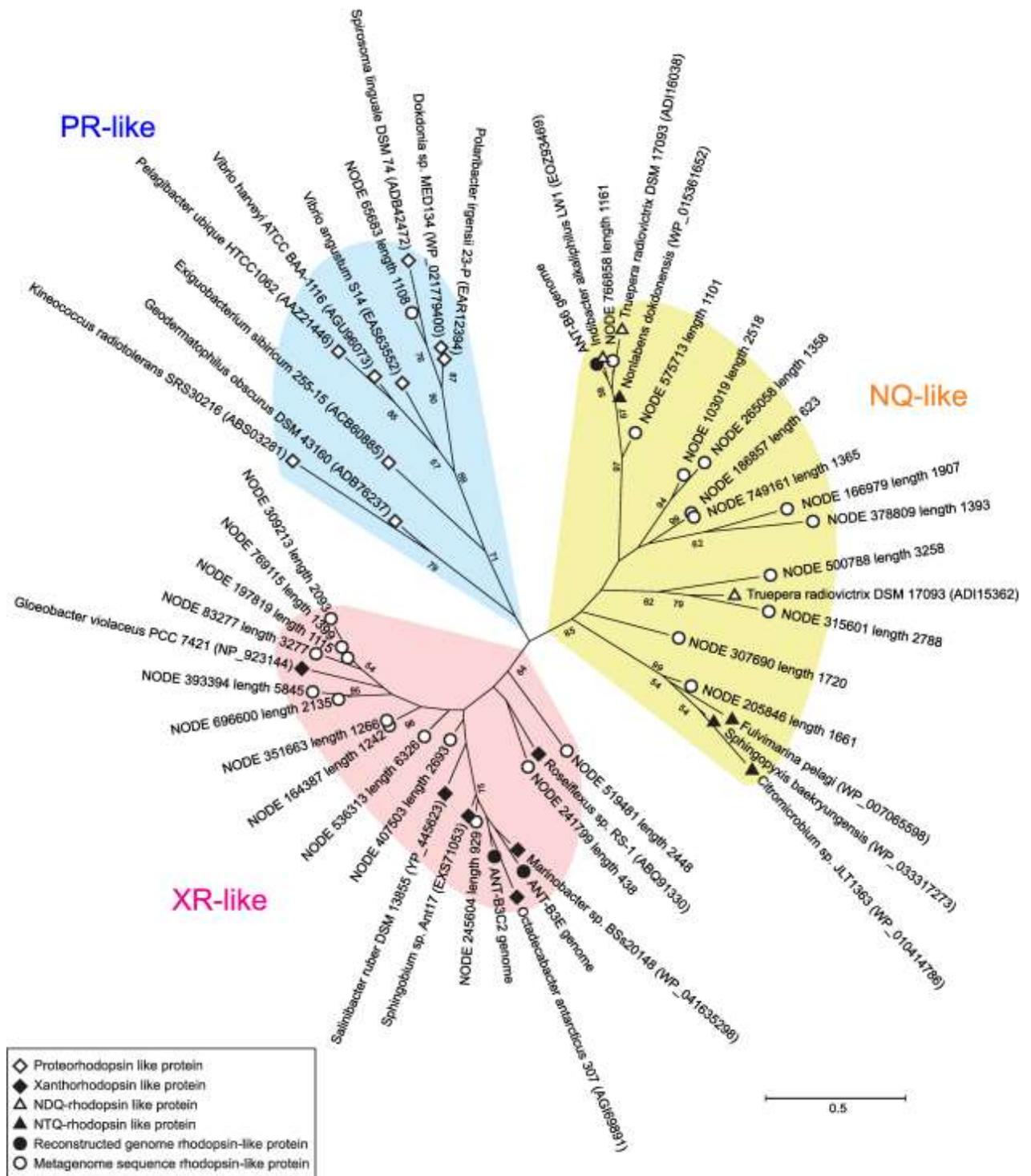


Fig. 1. Maximum likelihood tree of microbial rhodopsins and sequences identified in the Antarctic hypolith metagenome. Rhodopsins-like proteins identified in the ANT-B3C2, ANT-B3E, ANT-B6 genomes and in the hypolith Antarctic metagenome (Identified by contig name and length) are shown. For reference sequences, accession numbers are provided in brackets. Bootstrap support values >50% are shown.

single copy marker genes present in the metagenome (**Supplementary Table 2**). A ratio of approximately 1:4 was calculated, similar to that found in other typical rhodopsin-containing environments (Finkel *et al.*, 2013), whereas the abundance of rhodopsin-containing bacteria in the hypolith metagenome comprised approximately 20% of the total bacterial diversity (**Supplementary Table 2**).

Hypolithic bacterial genome reconstructions

To infer the ecological relevance for rhodopsins in hypolithic communities, a number of bacterial genomes, including several rhodopsin-encoding bacterial genomes, were reconstructed and analysed. Our approach was to first cluster contigs based on total GC content and coverage (**Supplementary Fig. 4**), followed by tetranucleotide composition binning (**Supplementary Fig. 5**), yielding thirteen complete genomes (**Supplementary Table 3**). Of these assembled genomes, two contained a XR-like rhodopsin (ANT-B3C2 and ANT-B3E), and one a NDQ-like rhodopsin (ANT-B6) (**Supplementary Table 1**). Based on the average genome coverage and the coverage of single marker genes present in the metagenome, we were able to estimate the relative abundance of these three genomes. ANT-B3E, ANT-B3C2 and ANT-B6 belong to the 15 most abundant bacterial phylotypes identified in the metagenome, with coverage values of 87.8x, 90.3x and 93.2x (relative abundance 1.83, 1.88 and 1.94%), respectively (**Supplementary Fig. 3**).

Broadly conserved proteins (**Supplementary Table 4**) were identified and used to place these genomes into two phylogenetic trees using known reference genomes (370 Proteobacteria and 85 Bacteroidetes genomes) available at the NCBI database. This classification was supported for ANT-B3C2 and ANT-B3E genomes based on 16S rRNA gene sequence analysis. However, no rRNA genes could be positively linked to the ANT-B6 genome assembly. ANT-B3C2 was classified as a member of the *Sphingomonadaceae* family, closely related to *Altererythrobacter* spp. and *Erythrobacter* spp. (**Fig. 2A**, **Supplementary Fig. 6** and **Supplementary Fig. 7A**). Our

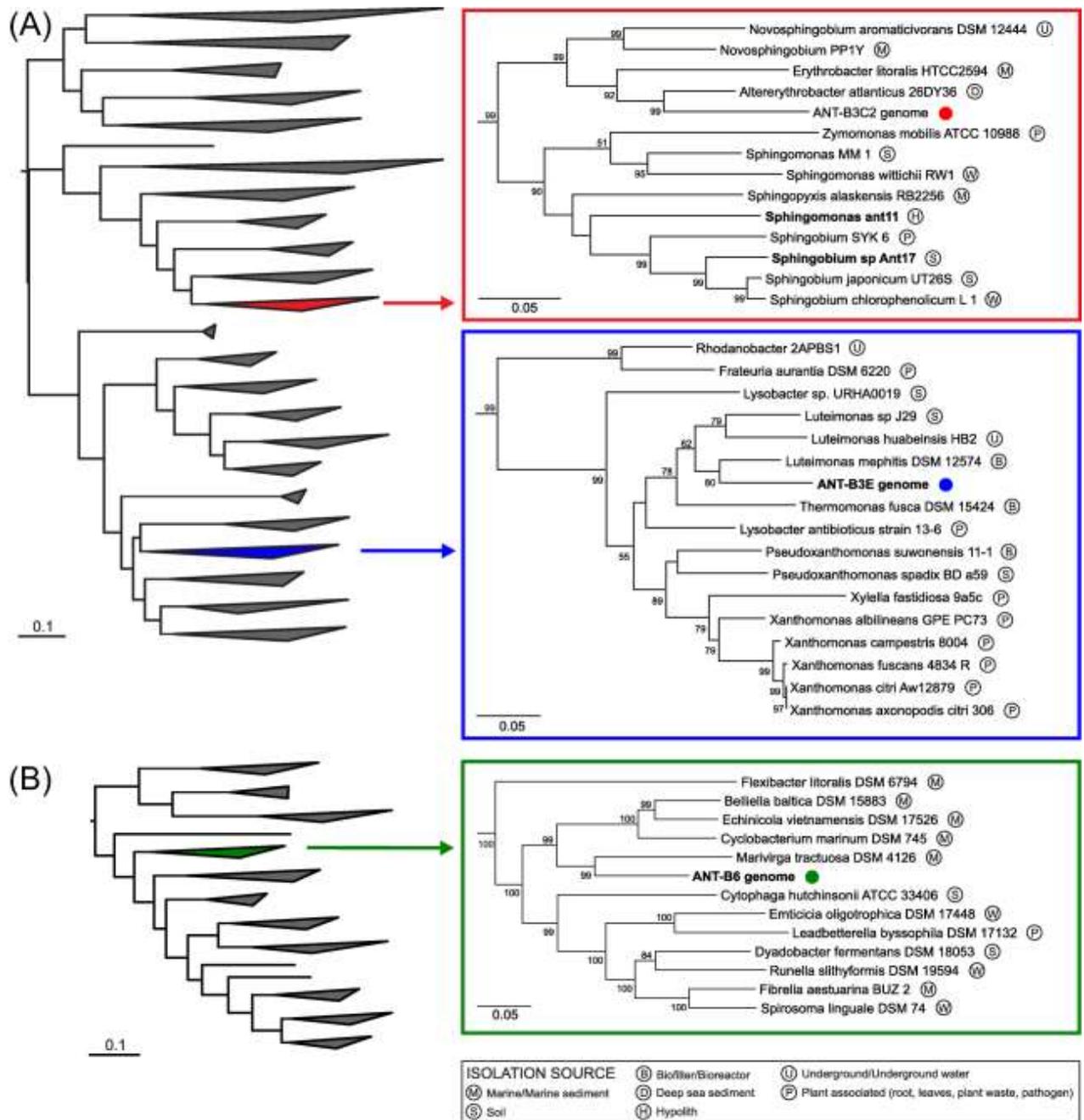


Fig. 2. Maximum likelihood tree of genomes based on concatenated marker proteins alignment. Branches containing ANT genomes were expanded. (a) Proteobacteria: ANT-B3C2 (red) and ANT-B3E (blue). (b) Bacteroidetes: ANT-B6 (green). The isolation source is indicated and Antarctic origin genomes are in bold. Bootstrap values higher than 50% are shown for expanded branches. Expanded phylogenetic trees are provided in Supplementary Figures 5 and 7.

analysis shows that ANT-B3C2 shares several features with their closely related neighbours. *Erythrobacter* spp. are described as aerobic anoxygenic photosynthetic bacteria (AAPB) (Koblížek *et al.*, 2003) with bacteriochlorophyll *a* (Bchl *a*) and a reaction center-light-harvesting

complex I (RC-LH I). These characteristics were also present in the ANT-B3C2 genome assembly. ANT-B3E is probably a member of the genus *Luteimonas* (**Fig. 2A, Supplementary Fig. 6 and Supplementary Fig. 7B**). Members of this genus are characterized as obligate, heterotrophic non-motile aerobes (as is the complete *Xanthomonadaceae* family) (Kerstens and Vancanneyt, 2005). ANT-B6 was classified as closely related to the *Marivirga* genus (Nedashkovskaya *et al.*, 2010)(**Fig. 2B and Supplementary Fig. 8**). This genus belongs to the class Cytophagia (family *Flammeovirgaceae*), whose members exhibit an aerobic chemoorganotrophic metabolism and encode gliding motility genes, both identified in the ANT-B6 genome.

Metabolism of the reconstructed genomes

To further assess whether these bacterial genomes contained the genetic capacity to produce functional rhodopsins, we investigated the presence of other protein-encoding genes necessary for a functional photoautotrophic system. Retinal, the light harvesting protein of the rhodopsin complexes, is synthesized from beta-carotene by a 15,15'-beta-carotene dioxygenase (*blh*) (Martinez *et al.*, 2007). This, and other genes implicated in the synthesis of beta-carotene from geranylgeranyl pyrophosphate, have been reported as operons (McCarren and DeLong, 2007). The all three genomes showed typical operon arrangements containing the genes necessary for the synthesis of retinal (**Supplementary Fig. 9**). These genes were identified from contigs, which were assigned, according to their GC content, coverage and tetranucleotide composition, to the genomes. We excluded all contigs represented as outliers or inaccurately assigned (**Supplementary Fig. 10 and Supplementary Fig. 11**). The ANT-B6 genome seemingly lacked the *blh* gene, although it is possible that the gene was undetected due to gaps in the genome assembly. However, analyses of genomes containing NQ-type rhodopsins suggest that these organisms apparently do not require *blh* and *idi* (isopentenyl diphosphate δ -isomerase) for the

biosynthesis of retinal (Kwon *et al.*, 2013). The XR have been shown to use a carotenoid as antenna for absorbing light energy, and transfer it to retinal (Balashov *et al.*, 2005). Two carotenoids have been identified in members of the *Salinibacter* and *Gloeobacter* genera (salinixanthin and echinenone, respectively), both are characterized by the presence of a 4-keto group which is critical for carotenoid binding to the rhodopsin (Balashov *et al.*, 2010). The addition of a keto group is mediated by carotenoid ketolases (for example, the *crtO* and *crtW* genes) (Moise *et al.*, 2014). The high homology of the ANT-B3C2 and ANT-B3E rhodopsins to XR suggests the possible presence of similar carotenoids in these organisms. Genes homologous to *crtW/O* were identified in the ANT-B3E and ANT-B6 genomes

The presence of complete retinal synthesis pathways suggests that ANT-B3C2, ANT-B3E and ANT-B6 all potentially have the capacity to synthesize functional rhodopsins. Furthermore, the rhodopsin protein gene present in ANT-B3C2 and ANT-B3E may represent a XR-like rhodopsin with a carotenoid antenna protein.

In silico metabolic reconstruction of the genomes was used to infer their metabolic capacities. All three genomes possessed typical features of aerobic metabolism, characterized by the presence of a complete electron transport chain (ETC) for oxidative phosphorylation. The central metabolic pathways included complete or near complete glycolysis, TCA cycle, pentose phosphate and Entner-Duodoroff pathways (**Fig. 3**). Additionally, in ANT-B3E the complete glyoxylate cycle pathway was identified. Moreover, several transporters that facilitate the uptake of carbon compounds were identified in the proteobacterial genomes ANT-B3C2 and ANT-B3E; including the bacterial phosphotransferase system (PTS) which translocates and phosphorylates sugars (Kotrba *et al.*, 2001). The capacity to secrete proteins across the membrane, mediated by the Sec-dependent secretion pathway and the twin-arginine transport system, was identified in all three genomes. Evidence for the presence of type II and IV secretion systems was identified in ANT-

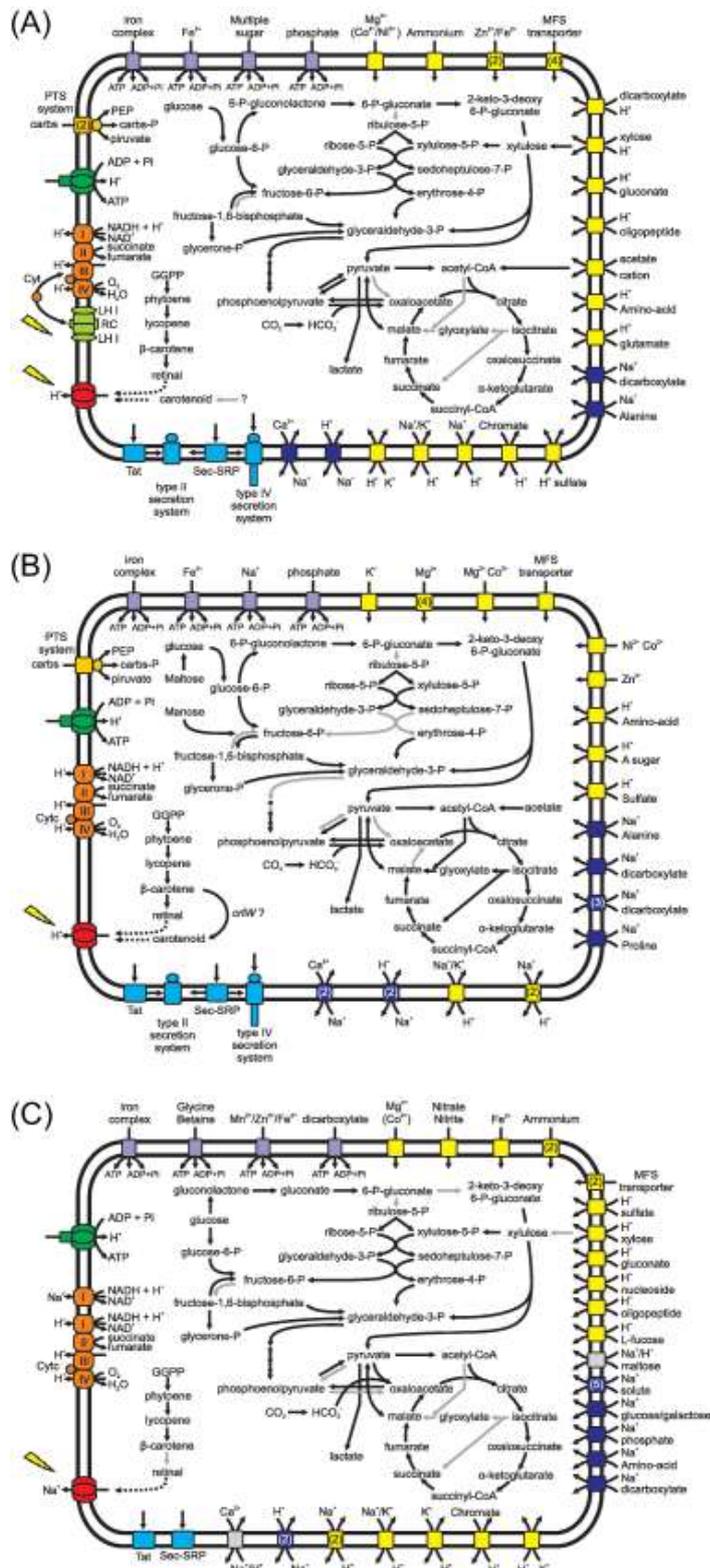


Fig. 3. Schematic overview of the three draft genomes indicating important metabolism pathways. (A) ANT-B3C2, (B) ANT-B3E and (C) ANT-B6 genome. Green arrows represent reactions identified in the genome, while red arrows indicate absent or unidentified reactions. The numbers in brackets (inside transporters) represents the number of copies identified for that transporter in the genome reconstruction.

B3C2, whereas type II and III secretion systems genes were identified in ANT-B3E (**Fig. 3A and 3B**). Secreted proteins and lipoproteins may participate in surface adhesion and degradation of polysaccharides and peptides for acquiring both carbon and energy (González *et al.*, 2011). Additionally, a number of glycoside hydrolases and other CAZymes (Cantarel *et al.*, 2009) were identified (**Table 1**). The presence of several peptidases, mostly of the serine (S) and metallo (M) MEROPS families (Rawlings *et al.*, 2014), also suggests the capacity to degrade (and uptake) extracellular proteins, using amino acids as a possible carbon and nitrogen source (Xing *et al.*, 2015). This is supported by the presence of specific transporters for peptides and amino acids (**Fig. 3**), and the presence of TonB receptors (**Table 1**) which have been associated with rhodopsins in the transport of nutrients (Morris *et al.*, 2010). In all three genomes, cotransporters able to use concentration gradients of H⁺ and Na⁺ were identified, mostly associated with the acquisition of carbon and nitrogen compounds. The presence of antiporters that exchange both Na⁺ and H⁺ suggests that these proteins may be used with the rhodopsins to modify the concentration of both ions based on the requirement of other solute cotransporters (Häse *et al.*, 2001).

The ANT-B6 genome encoded Na⁺ translocating (rather than H⁺-translocating) NADH:quinine reductase (NQR) subunits, indicating an affinity of this bacterium for Na⁺ (Kimura *et al.*, 2011). In all three genomes, genes coding for a carbonic anhydrase and anaplerotic CO₂ fixation enzymes were identified (**Table 1 and Fig. 3**). The enzyme PEP carboxylase was present in ANT-B3C2 and ANT-B3E genomes, whereas pyruvate carboxylase was present in ANT-B6. These enzymes replenish the intermediate compounds in the TCA cycle when they are used for biosynthesis (Moran and Miller, 2007).

Functional inferences based on phylogeny and metabolic reconstruction suggest that the three assembled genomes are typical of heterotrophic aerobic bacteria. These genomes have the

Table 1. General characteristics of reconstructed genomes.

	ANT-B3C2	ANT-B3E	ANT-B6
Contigs	42	57	255
Assembly size (Mb)	2.7	2.8	3.6
GC%	66.5	70.1	41.3
Coding sequences	2669	2652	3360
Essential genes	103	70	96
Completeness (Contamination) ^a	91.45% (0.37)	92.86% (0.0)	85.27% (0.15)
tRNAs	36	37	22
Rhodopsin-like protein (ion)	XR (H ⁺)	XR (H ⁺)	NDQ (Na ⁺)
Bacteriochlorophyll a (Bchl a)	+	–	–
Carbon fixation (Anaplerotic)	PEP carboxylase	PEP carboxylase	Pyruvate carboxylase
Motility	not detected	not detected	Gliding
Polysaccharide lyases ^b	2	1	2
Glycoside hydrolases ^b	90	60	134
Carbohydrate esterases ^b	32	27	28
Auxiliary Activities ^b	11	7	8
Carbohydrates-Binding Modules ^b	44	45	97
Peptidases	63	65	73
TonB receptors	28	18	39

PEP: Phosphoenolpyruvate.

a. Calculated with CheckM.

b. Identified in the CAZymes database.

potential to use a broad range of organic molecules, including proteins and polymeric compounds, as carbon and nitrogen sources.

Discussion

In hypolithons, primary production is performed almost exclusively by photosynthetic bacteria, with previous studies reporting Cyanobacteria as the phenotypically dominant phylum (Chan *et al.*, 2012; de los Ríos *et al.*, 2014). Based on the estimation of community composition using the 16S rRNA and other marker genes, we identified the phyla Actinobacteria, Bacteroidetes and Proteobacteria as the most abundant and diverse in hypolithons. However, the Cyanobacteria abundance in the metagenome is still high and the bacterial diversity reported here was broadly similar to that previously reported from studies using ‘classical’ 16S rRNA gene amplicon sequence analyses (Makhalanyane *et al.*, 2013; Van Goethem *et al.*, 2016). This observation supports the importance of phyla other than Cyanobacteria as key components in hypolithic communities and, at the same time, raises the question how these bacteria support their energetic requirements.

Here we describe the presence of a rhodopsin-like system in hypolithons, the first such observation for soil environments. Given the presence of different types of rhodopsins in approximately 20% of the bacteria in the hypolithic metagenome and, in the absence of any evidence of contamination with other rhodopsin-carrying bacterial sources, we suggest that these proteins may be significant components of these bacteria-dominated communities. The abundance of rhodopsins in the hypolithic community was similar to that of other environments, such as marine systems, where rhodopsins are widely observed and appear to play an important ecological role (Finkel *et al.*, 2013). Of the different types of rhodopsins identified in this study, the conserved amino acid residues and sequence homologies have shown that these proteins are putative light-driven ion pumps (H^+ , Na^+ or Cl^-), with most classified as XR-like and NQ-like

proteins. The presence of NQ-like rhodopsins, which are generally associated with microbial communities on high salt concentration environments (Kwon *et al.*, 2013), is intriguing in hypolithons, which inhabit a relatively low salinity niche (hypolith Cl^- concentration, 10 mg/kg) (Makhalanyane *et al.*, 2013). However, we note that bacterial cells in hypolithic communities are often embedded in substantial EPS matrices (de los Ríos *et al.*, 2014), in which the local ion concentrations may be relatively high.

Among the 13 identified hypolithic microbial genomes, three harbouring rhodopsin genes. These genomes, identified as the most abundant in the metagenome, suggest that they may be ubiquitous species in hypolithons and well adapted to this environment. The three assembled genomes showed metabolic capacities typical of heterotrophs and encoded all the genes necessary to synthesize functional rhodopsins. The presence of potentially functional rhodopsins, coupled with carbonic anhydrases and anaplerotic enzymes (**Table 1 and Fig. 3**) which facilitate the incorporation of CO_2 into TCA intermediates, may represent an important physiological benefit to ANT-B3C2, ANT-B3E and ANT-B6. The use of solar energy to drive CO_2 fixation may complement the consumption of fixed carbon for the generation of biomass (Hauruseu and Koblížek, 2012; Palovaara *et al.*, 2014). Energy from rhodopsin activity may also enhance the uptake of nutrients, without depleting limited carbon resources. The coupling of rhodopsin and transport activities is most likely in ANT-B6, where the NDQ-like rhodopsin can use light energy to generate a transmembrane Na^+ gradient capable of driving Na^+ -dependent transporters (Kimura *et al.*, 2011). This observation is supported by the higher number of Na^+ -cotransporters, compared to H^+ -cotransporters in ANT-B6. In the ANT-B3C2 and ANT-B3E genomes, which contain XR-like rhodopsin genes, Na^+ -cotransporters are less common (**Fig. 3**).

A second system that uses light as an energy source, the AAPB system, was identified in the ANT-B3C2 genome. AAPB use Bchl *a* and a RC-LH I complex for ATP synthesis without fixing CO_2 (Yurkov and Beatty, 1998). This is a common feature in marine and fresh water microbiota

(Koblížek, 2015). This system was described in *Erythrobacter* (Koblížek *et al.*, 2003), one of the closest genera to the ANT-B3C2 genome. However, to our knowledge bacterium with a Bchl *a* metabolism and rhodopsin has never been described. In this sense, we only can hypothesize that the presence of both systems may be advantageous to exploit a wide range of the radiation spectrum. The *in vivo* absorption of Bchl *a* is at 800-870 nm (Yurkov and Beatty, 1998), whereas the blue and green rhodopsins absorb at 490 and 525 nm respectively (Bamann *et al.*, 2014). For XR, the protein with the highest homology to the rhodopsin in ANT-B3C2, the maximum absorbance is at 560 nm, which is extended to 487 and 521 nm for the action of the carotenoid (salinixanthin) (Balashov *et al.*, 2005).

It is widely accepted that rhodopsins play a significant role in marine microbiota, providing alternative energy generation and metabolic strategies under oligotrophic conditions. The diversity and abundance of rhodopsin-encoding sequences found in an Antarctic soil hypolithic metagenome suggests that these proteins may play similar roles in this cold edaphic environment. The quantitative contribution of this process to the energy budget of the hypolithic microbial community remains unknown, but it is tempting to speculate that bacterial rhodopsin-mediated photoheterotrophy may play a significant role in energy homeostasis.

Experimental Procedures

Sample collection and DNA extraction

A total of 50 samples were collected and stored in sterile Whirl-Pak bags (Nasco International, Fort Atkinson, WI, USA). Samples were maintained at -20°C in the field and transported to the University of Pretoria and stored at -80°C before further analysis. We extracted DNA from 0.5g of each sample (n=50) using the PowerSoil DNA isolation kit (MO BIO, Carlsbad, CA, USA) as detailed previously (Makhalanyane *et al* 2013). All samples yielded, high molecular weight, intact DNA. The DNA was pooled, sheered into fragments of approximately 300 bases and

retrieved from agarose gels. Before performing bridge amplification and sequencing, adapters were ligated to the ends of the DNA fragments.

Shotgun Metagenomic sequencing

Samples were combined and sequencing was carried out using Illumina HiSeq-2000 paired-end technology (2 X 101 bp) as described previously (Vikram *et al.*, 2015). A library of 334 bp inserts length was constructed and the quality was checked as detailed previously (Vikram *et al.*, 2015). Reads with an average quality value below 25 or containing ambiguous bases (N) were eliminated, using in house python scripts.

Ribosomal RNA reconstruction

The filtered reads were used to reconstruct small (SSU) and large (LSU) rRNA subunit genes, as well as to infer the proportion of different taxa in the sample. Reads corresponding to SSU and LSU were identified by Metaxa2 (Bengtsson-Palme *et al.*, 2015) and retrieved from the original filtered reads pool. SSU and LSU sequences were reconstructed based on EMIRGE instructions (Miller *et al.*, 2011) and the respective SILVA reference databases (Quast *et al.*, 2013).

For phylogenetic classification and estimation of the proportion of each OTU identified in the sample, bacterial SSU sequences were reconstructed using a 97% similarity cut-off. The relative proportion of each sequence was estimated by EMIRGE (**Supplementary Fig. 1 and 2**). The resultant sequences were classified using the SINA alignment service (SILVA) (Pruesse *et al.*, 2012). The two most similar sequences (> 70% similarity), from each query sequence, were retrieved and used for comparison. All the sequences were aligned using MEGA 6 (Tamura *et al.*, 2013) and the resulting alignment was used to generate a maximum likelihood tree with 500 bootstrapped replicates (**Supplementary Fig. 1**).

Metagenome assembly

The initial metagenome assembly was conducted using Velvet v1.2.10 (Zerbino and Birney, 2008), with parameters of $kmer=71$ in *velveth*. Protein coding sequences were predicted from metagenomic contigs (≥ 500 bp) using MetaGeneMark (Zhu *et al.*, 2010). The presence of 111 essential genes (Albertsen *et al.*, 2013) were predicted using HMMER3 (Eddy, 2011) in order to identify contigs containing at least 1 essential gene. The taxonomic affiliation of the predicted essential genes was determined with MEGAN5 (Huson and Mitra, 2012) using the results of BLAST (blastp) against the NCBI refseq_protein database.

To obtain abundance estimates and taxonomic profiles of the metagenome, independent of the 16S rRNA gene, we identified marker genes and predicted coding sequences (contigs ≥ 500 bp) using the AMPHORA2 software (Wu and Scott, 2012). Only genes shorter than 150 aa (which can be completely codified in ≥ 500 bp contigs) identified by AMPHORA2 (perl MarkerScanner.pl script) were selected (**Supplementary table 2**). Abundance was estimated using the average coverage of the contigs where marker genes were identified. The same contigs were annotated using KRAKEN software (Wood and Salzberg, 2014) to get the taxonomic profiles of the metagenomics data (**Supplementary Fig. 3**).

Identification of microbial rhodopsins

Representative microbial rhodopsins from GenBank were blast searched (tblastn) against the Antarctic hypolith metagenome contigs with a minimum length of 500 bp to identify possible rhodopsins present in the metagenome. The identified amino acid sequences were then aligned with 21 representative rhodopsins in order to identify conserved residues (**Supplementary Table 1**). A maximum likelihood tree was inferred and evaluated using 500 bootstrap replicates (**Fig. 1**). The proportion of bacteria with functional rhodopsins was estimated by dividing the number of identified functional rhodopsins to the number of single copy marker genes (average number of

hits for each gene) (**Supplementary Table 2**). Whereas the abundance of rhodopsins was estimated in the same way by calculating the total coverage instead of the number of hits.

Genomes reconstruction from the metagenome.

The coverage (Ln of coverage) was plotted against the GC content for all contigs longer than 2000 bp. Additionally, contigs containing essential genes (classified as previously explained) were highlighted on the graph using different colours according to the taxonomic affiliation (**Supplementary Fig. 4**). Essential genes were also used for preliminary evaluation of genome completeness, since roughly 100 essential genes are expected in all bacterial species (Albertsen *et al.*, 2013).

Clusters of contigs were manually delimited on the graph and the contigs sequences into these clusters extracted using R (Team, 2014) (**Supplementary Fig. 4**). To isolate contigs belonging to individual genomes, tetranucleotide frequency binning was used to resolve the defined clusters. Contigs assigned to clusters as defined in Supplementary Fig. 3, were first marked and combined for analysis using the Databionic ESOM Tools (Ultsch and Mörchen, 2005) (**Supplementary Fig. 5**) as previously reported by Dick *et al.* (Dick *et al.*, 2009). The procedure was repeated several times combining different clusters. The resulting groups of contigs (bins), which consistently appeared in the different ESOM runs, were considered single genomes or a mixture of multiple closely related genomes (for example species or subspecies) (**Supplementary Fig. 5 and Supplementary Table 3**).

To identify contigs which may be related with a particular bin, but which were not identified in the clustering and binning round, paired-end connections between contigs were tracked. For this purpose paired-end reads were mapped against all contigs using Bowtie2 (Langmead and Salzberg, 2012) and subsequently a perl script (Albertsen *et al.*, 2013) was used to identify the connections between contigs. The connections between contigs were visualized using Cytoscape

(Shannon *et al.*, 2003). Contigs linked by >5 paired-end connections to contigs in a particular bin were added to that bin.

To improve the assembly of each genome in the bins, original paired-end reads for each bin were retrieved using an in-house python script and reassembled using Velvet software. Each new assembly was treated as a single genome where the k-mer length was selected to optimize the N50, number of contigs and total length assembly. Pair-end connectivity of the resulting contigs was analysed and the unconnected contigs shorter than 500 bp were eliminated. Final contigs joining (when possible) was performed manually using Gap5 from the Staden Package (Bonfield and Whitwham, 2010). Correspondence analysis (CA) using tetranucleotide frequencies were performed over the final contigs to check the presence of outliers that could be contaminants (**Supplementary Fig. 10**). The completeness and contamination of the final genome assemblies was estimated using CheckM (Parks *et al.*, 2015).

Genomes rRNA reconstruction

Complete rRNA sequences could not be reconstructed during the de novo assembly. Instead, both subunits genes were reconstructed using EMIRGE to identify the SSU and LSU rRNA genes. The reconstruction was done as described in the Ribosomal RNA reconstruction section above, using a 100% similarity instead of 97% cut-off value in order to minimize the presence of mismatches. For each genome, the corresponding rRNA-reconstructed sequences were identified by paired-end read connections with the genome contigs, which were also joined where possible.

The identity of rRNA sequences was confirmed using Blastn. For both genomes (ANT-B3C2 and ANT-B3E) with rRNA genes detected, phylogenetic analysis of the SSU using highly curated reference sequences from the LTP SILVA database were conducted. For the ANT-B3C2 genome 65 sequences were retrieved corresponding to the *Sphingomonadales* group. Whereas, for the ANT-B3E genome 98 sequences were retrieved corresponding to the *Lysobacterales* group. The

sequences were aligned using MEGA6 (Tamura *et al.*, 2013) and the resulting alignments were used to generate maximum likelihood trees with 500 bootstrapped replicates (**Supplementary Fig. 7**).

Genomes annotation

Genomic protein coding sequences were predicted using GeneMarkS (Besemer *et al.*, 2001). The final assembly of each genome was annotated using RAST (Aziz *et al.*, 2008), KAAS (Moriya *et al.*, 2007) and Blastp (ref_seq proteins database). Pathways, operons, transporters, and pathway holes were predicted using the Pathway Tools software suite (Karp *et al.*, 2010).

CAZymes (Cantarel *et al.*, 2009) were identified using Blastp and the CAZymes data base (year 2014), not including Glycosyl transferases. Some enzymes could be annotated in more than one category according to the database classification. Query sequences were filter by e-value and coverage: Blastp cutoff e-value: = 1.00E-10 and coverage \geq 60 %. Peptidases were identified and classified using blast MEROPS search (Rawlings *et al.*, 2014) and filter according to an e-value \geq 1.00E-10 (**Table 1**).

Phylogenetic comparison of genomes

The identified genomes were compared against a series of representative genomes available on the NCBI GeneBank database. Translated proteins of each genome were explored using AMPHORA2 software (Wu and Scott, 2012) to identified 31 phylogenetic marker genes (**Supplementary Table 4**). Only the proteins identified in all the genomes were used to generate alignments as detailed in the AMPHORA2 instructions manual. The individual alignments were concatenated together to generate a unique alignment including all the proteins. This procedure was applied in order to generate one concatenated alignments for Proteobacteria and one for Bacteroidetes. The final alignment for Proteobacteria included 366 reference genomes, one

isolated *Sphingobium* sp. from Antarctic soil (Adriaenssens *et al.*, 2014), one previously isolated *Sphingomonas* sp. from Antarctic hypoliths (Gunnigle *et al.*, 2015) and the two query genomes (ANT-B3C2 and ANT-B3E) (**Supplementary Fig. 6**). The Bacteroidetes alignment included 84 reference genomes and the ANT-B6 query genome (**Supplementary Fig. 8**). Each alignment was used to generate phylogenetic trees using the Maximum-likelihood method with 500 bootstrap replications in MEGA6 (Tamura *et al.*, 2013). The aspect of the trees was modified with iTOL (Letunic and Bork, 2007) (**Fig. 2**).

Accession Numbers

These Whole Genome Shotgun projects have been deposited at DDBJ/ENA/GenBank under the accession numbers LXQI00000000 (ANT-B3C2), LXQJ00000000 (ANT-B3E) and LXQK00000000 (ANT-B6). The versions described in this paper are the first versions, LXQI01000000 (ANT-B3C2), LXQJ01000000 (ANT-B3E) and LXQK01000000 (ANT-B6). Unprocessed reads have been deposited under the accession number SRX1726658.

Acknowledgments

We gratefully acknowledge financial support from the South African National Research Foundation SANAP (93074) and Blue Skies (81693) funding programs, and from the University of Pretoria Genomics Research Institute. The South African Centre for High Performance Computing and the University of Pretoria Centre for Bioinformatics and Computational Biology, for computer cluster access and technical support. L.D.G. was supported by the South African National Research Foundation. S.V. was funded by the Claude Leon Foundation postdoctoral Fellowship program.

The authors declare no conflict of interest.

References

- Adriaenssens, E.M., Guerrero, L.D., Makhalyane, T.P., Aislabie, J.M., and Cowan, D.A. (2014) Draft Genome Sequence of the Aromatic Hydrocarbon-Degrading Bacterium *Sphingobium* sp. Strain Ant17, Isolated from Antarctic Soil. *Genome Announc.* **2**: e00212-14.
- Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., and Nielsen, P.H. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* **31**: 533–8.
- Atamna-Ismaeel, N., Finkel, O.M., Glaser, F., Sharon, I., Schneider, R., Post, A.F., et al. (2012) Microbial rhodopsins on leaf surfaces of terrestrial plants. *Environ. Microbiol.* **14**: 140–6.
- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., et al. (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.
- Balashov, S.P., Imasheva, E.S., Boichenko, V.A., Antón, J., Wang, J.M., and Lanyi, J.K. (2005) Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. *Science* **309**: 2061–4.
- Balashov, S.P., Imasheva, E.S., Choi, A.R., Jung, K.-H., Liaaen-Jensen, S., and Lanyi, J.K. (2010) Reconstitution of gloeobacter rhodopsin with echinenone: role of the 4-keto group. *Biochemistry* **49**: 9792–9.
- Bamann, C., Bamberg, E., Wachtveitl, J., and Glaubitz, C. (2014) Proteorhodopsin. *Biochim. Biophys. Acta* **1837**: 614–25.
- Béjà, O., Aravind, L., Koonin, E. V., Suzuki, M.T., Hadd, A., Nguyen, L.P., et al. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–6.
- Bengtsson-Palme, J., Hartmann, M., Eriksson, K.M., Pal, C., Thorell, K., Larsson, D.G.J., and Nilsson, R.H. (2015) metaxa2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol. Ecol. Resour.*
- Besemer, J., Lomsadze, A., and Borodovsky, M. (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**: 2607–18.
- Bonfield, J.K. and Whitwham, A. (2010) Gap5--editing the billion fragment sequence assembly. *Bioinformatics* **26**: 1699–703.
- Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37**: D233-8.
- Cary, S.C., McDonald, I.R., Barrett, J.E., and Cowan, D.A. (2010) On the rocks: the microbiology of Antarctic Dry Valley soils. *Nat. Rev. Microbiol.* **8**: 129–38.
- Chan, Y., Lacap, D.C., Lau, M.C.Y., Ha, K.Y., Warren-Rhodes, K.A., Cockell, C.S., et al. (2012) Hypolithic microbial communities: between a rock and a hard place. *Environ. Microbiol.* **14**: 2272–82.
- Chan, Y., Van Nostrand, J.D., Zhou, J., Pointing, S.B., and Farrell, R.L. (2013) Functional ecology of an Antarctic Dry Valley. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 8990–5.
- Cowan, D.A., Khan, N., Pointing, S.B., and Cary, S.C. (2010) Diverse hypolithic refuge communities in the McMurdo Dry Valleys. *Antarct. Sci.* **22**: 714–720.

- Dick, G.J., Andersson, A.F., Baker, B.J., Simmons, S.L., Thomas, B.C., Yelton, A.P., and Banfield, J.F. (2009) Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10**: R85.
- Eddy, S.R. (2011) Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**: e1002195.
- Farrelly, V., Rainey, F.A., and Stackebrandt, E. (1995) Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Appl. Environ. Microbiol.* **61**: 2798–801.
- Feng, S., Powell, S.M., Wilson, R., and Bowman, J.P. (2013) Light-stimulated growth of proteorhodopsin-bearing sea-ice psychrophile *Psychroflexus torquus* is salinity dependent. *ISME J.* **7**: 2206–13.
- Finkel, O.M., Béjà, O., and Belkin, S. (2013) Global abundance of microbial rhodopsins. *ISME J.* **7**: 448–51.
- Fuhrman, J.A., Schwalbach, M.S., and Stingl, U. (2008) Proteorhodopsins: an array of physiological roles? *Nat. Rev. Microbiol.* **6**: 488–94.
- Van Goethem, M.W., Makhalanyane, T.P., Valverde, A., Cary, S.C., and Cowan, D.A. (2016) Characterization of bacterial communities in lithobionts and soil niches from Victoria Valley, Antarctica. *FEMS Microbiol. Ecol.* **92**: fiw051.
- Gómez-Consarnau, L., Akram, N., Lindell, K., Pedersen, A., Neutze, R., Milton, D.L., et al. (2010) Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biol.* **8**: e1000358.
- Gómez-Consarnau, L., González, J.M., Riedel, T., Jaenicke, S., Wagner-Döbler, I., Sañudo-Wilhelmy, S. a, and Fuhrman, J. a (2015) Proteorhodopsin light-enhanced growth linked to vitamin-B1 acquisition in marine Flavobacteria. *ISME J.* 1–11.
- González, J.M., Pinhassi, J., Fernández-Gómez, B., Coll-Lladó, M., González-Velázquez, M., Puigbò, P., et al. (2011) Genomics of the proteorhodopsin-containing marine flavobacterium *Dokdonia* sp. strain MED134. *Appl. Environ. Microbiol.* **77**: 8676–86.
- Gunnigle, E., Ramond, J.-B., Guerrero, L.D., Makhalanyane, T.P., and Cowan, D.A. (2015) Draft genomic DNA sequence of the multi-resistant *Sphingomonas* sp. strain AntH11 isolated from an Antarctic hypolith. *FEMS Microbiol. Lett.* **362**: fnv037.
- Gushchin, I., Chervakov, P., Kuzmichev, P., Popov, A.N., Round, E., Borshchevskiy, V., et al. (2013) Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria. *Proc. Natl. Acad. Sci.* **110**: 12631–12636.
- Häse, C.C., Fedorova, N.D., Galperin, M.Y., and Dibrov, P. a (2001) Sodium ion cycle in bacterial pathogens: evidence from cross-genome comparisons. *Microbiol. Mol. Biol. Rev.* **65**: 353–70, table of contents.
- Hauruseu, D. and Koblížek, M. (2012) Influence of light on carbon utilization in aerobic anoxygenic phototrophs. *Appl. Environ. Microbiol.* **78**: 7414–9.
- Huson, D.H. and Mitra, S. (2012) Introduction to the analysis of environmental sequences: metagenomics with MEGAN. *Methods Mol. Biol.* **856**: 415–29.
- Inoue, K., Kato, Y., and Kandori, H. (2015) Light-driven ion-translocating rhodopsins in marine bacteria. *Trends Microbiol.* **23**: 91–8.
- Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T.J., et al. (2010) Pathway Tools version 13.0: integrated software for pathway/genome informatics and

- systems biology. *Brief. Bioinform.* **11**: 40–79.
- Kersters, K. and Vancanneyt, M. (2005) Bergey's Manual of Systematic Bacteriology.
- Kimura, H., Young, C.R., Martinez, A., and DeLong, E.F. (2011) Light-induced transcriptional responses associated with proteorhodopsin-enhanced growth in a marine flavobacterium. *ISME J.* **5**: 1641–51.
- Koblížek, M. (2015) Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol. Rev.* **39**: 854–70.
- Koblížek, M., Bèjà, O., Bidigare, R.R., Christensen, S., Benitez-Nelson, B., Vetriani, C., et al. (2003) Isolation and characterization of *Erythrobacter* sp. strains from the upper ocean. *Arch. Microbiol.* **180**: 327–38.
- Kotrba, P., Inui, M., and Yukawa, H. (2001) Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism. *J. Biosci. Bioeng.* **92**: 502–17.
- Kwon, S.K., Kim, B.K., Song, J.Y., Kwak, M.J., Lee, C.H., Yoon, J.H., et al. (2013) Genomic makeup of the marine flavobacterium *Nonlabens* (*Donghaeana*) *dokdonensis* and identification of a novel class of rhodopsins. *Genome Biol. Evol.* **5**: 187–99.
- Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**: 357–9.
- Letunic, I. and Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127–8.
- de los Ríos, A., Cary, C., and Cowan, D. (2014) The spatial structures of hypolithic communities in the Dry Valleys of East Antarctica. *Polar Biol.* **37**: 1823–1833.
- Makhalanyane, T.P., Valverde, A., Birkeland, N.-K., Cary, S.C., Tuffin, I.M., and Cowan, D.A. (2013) Evidence for successional development in Antarctic hypolithic bacterial communities. *ISME J.* **7**: 2080–90.
- Martinez, A., Bradley, A.S., Waldbauer, J.R., Summons, R.E., and DeLong, E.F. (2007) Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 5590–5.
- McCarren, J. and DeLong, E.F. (2007) Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environ. Microbiol.* **9**: 846–58.
- Miller, C.S., Baker, B.J., Thomas, B.C., Singer, S.W., and Banfield, J.F. (2011) EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol.* **12**: R44.
- Moise, A.R., Al-Babili, S., and Wurtzel, E.T. (2014) Mechanistic aspects of carotenoid biosynthesis. *Chem. Rev.* **114**: 164–93.
- Moran, M.A. and Miller, W.L. (2007) Resourceful heterotrophs make the most of light in the coastal ocean. *Nat. Rev. Microbiol.* **5**: 792–800.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**: W182–5.
- Morris, R.M., Nunn, B.L., Frazar, C., Goodlett, D.R., Ting, Y.S., and Rocap, G. (2010) Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J.* **4**: 673–85.

- Nakamura, Y., Kaneko, T., Sato, S., Mimuro, M., Miyashita, H., Tsuchiya, T., et al. (2003) Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. *DNA Res.* **10**: 137–45.
- Nedashkovskaya, O.I., Vancanneyt, M., Kim, S.B., and Bae, K.S. (2010) Reclassification of *Flexibacter tractuosus* (Lewin 1969) Leadbetter 1974 and “*Microscilla sericea*” Lewin 1969 in the genus *Marivirga* gen. nov. as *Marivirga tractuosa* comb. nov. and *Marivirga sericea* nom. rev., comb. nov. *Int. J. Syst. Evol. Microbiol.* **60**: 1858–63.
- Palovaara, J., Akram, N., Baltar, F., Bunse, C., Forsberg, J., Pedrós-Alió, C., et al. (2014) Stimulation of growth by proteorhodopsin phototrophy involves regulation of central metabolic pathways in marine planktonic bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **111**: E3650-8.
- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**: 1043–55.
- Pointing, S.B., Chan, Y., Lacap, D.C., Lau, M.C.Y., Jurgens, J.A., and Farrell, R.L. (2009) Highly specialized microbial diversity in hyper-arid polar desert. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 19964–9.
- Pruesse, E., Peplies, J., and Glöckner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–9.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**: D590-6.
- Rawlings, N.D., Waller, M., Barrett, A.J., and Bateman, A. (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **42**: D503-9.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**: 2498–504.
- Steindler, L., Schwalbach, M.S., Smith, D.P., Chan, F., and Giovannoni, S.J. (2011) Energy starved *Candidatus Pelagibacter ubique* substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS One* **6**: e19725.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**: 2725–9.
- Team, R.C. (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- Ultsch, A. and Mörchen, F. (2005) ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM. In, *Technical Report Dept. of Mathematics and Computer Science, University of Marburg, Germany, No. 46*.
- Vikram, S., Guerrero, L.D., Makhalanyane, T.P., Le, P.T., Seely, M., and Cowan, D.A. (2015) Metagenomic analysis provides insights into functional capacity in a hyperarid desert soil niche community. *Environ. Microbiol.*
- Wood, D.E. and Salzberg, S.L. (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**:
- Wu, M. and Scott, A.J. (2012) Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* **28**: 1033–4.

- Xing, P., Hahnke, R.L., Unfried, F., Markert, S., Huang, S., Barbeyron, T., et al. (2015) Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. *ISME J.* **9**: 1410–22.
- Yoshizawa, S., Kumagai, Y., Kim, H., Ogura, Y., Hayashi, T., Iwasaki, W., et al. (2014) Functional characterization of flavobacteria rhodopsins reveals a unique class of light-driven chloride pump in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **111**: 6732–7.
- Yurkov, V. V and Beatty, J.T. (1998) Aerobic anoxygenic phototrophic bacteria. *Microbiol. Mol. Biol. Rev.* **62**: 695–724.
- Zerbino, D.R. and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**: 821–9.
- Zhu, W., Lomsadze, A., and Borodovsky, M. (2010) Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **38**: e132.