UNIVERSITEIT GENT

FACULTEIT WETENSCHAPPEN

The role of photoheterotrophic and chemoautotrophic prokaryotes in the microbial food web in terrestrial Antarctica: a cultivation approach combined with functional analysis

Guillaume Tahon

Promotor

Prof. Dr. Anne Willems

Dissertation submitted in fulfillment of the requirements for the degree of Doctor (Ph.D.) of Science: Biotechnology (Ghent University)

Tahon Guillaume | The role of photoheterotrophic and chemoautotrophic prokaryotes in the microbial food web in terrestrial Antarctica: a cultivation approach combined with functional analysis

ISBN-number: 978-94-6197-523-2

Printed by University Press | http://www.universitypress.be

Ph.D. thesis, Faculty of Sciences, Ghent University, Ghent, Belgium

This Ph.D. work was supported by the Fund for Scientific Research – Flanders (project G.0146.12)

Publically defended in Ghent, Belgium, May 5th, 2017

Examination committee

Prof. Dr. Savvas Savvides (Chairman)
L-Probe: Laboratory for protein Biochemistry and Biomolecular Engineering
Faculty of Sciences, Ghent University, Belgium

VIB Inflammation Research Center
VIB, Ghent, Belgium

Prof. Dr. Anne Willems (Promotor)
LM-UGent: Laboratory of Microbiology
Faculty of Sciences, Ghent University, Belgium

Prof. Dr. Elie Verleyen (Secretary)
Laboratory of Protistology and Aquatic Ecology
Faculty of Sciences, Ghent University, Belgium

Em. Prof. Dr. Paul De Vos
LM-UGent: Laboratory of Microbiology
Faculty of Sciences, Ghent University, Belgium

Dr. Natalie Leys
SCK·CEN: Environment, Health and Safety
Belgian Nuclear Research Centre, Mol, Belgium

Dr. Annick Wilmotte
Cyanobacteria Research group and BCCM/ULC Cyanobacteria Collection
Center for Protein Engineering, Liège University, Belgium

“Antarctica, an inhospitable wasteland, but even here, on the Earth's frozen bottom, we find life. And not just any life.”

Werner Herzog

Dankwoord

Het eerste dat ik dacht was "*Een dankwoord schrijven, hoe begin je daar in godsnaam aan, want het toch is onmogelijk om mijn 'dank' aan zoveel mensen beknopt op papier neer te pennen*.". Na veel schrijven en vooral herschrijven ben ik uiteindelijk tot een dankwoord gekomen. Onderstaande woorden zullen en kunnen nooit alles en iedereen omvatten, maar hopelijk zeggen mijn ietwat poëtische woorden toch een beetje hoe dankbaar ik ben voor jullie steun de afgelopen vijf jaar.

Beginnen kan ik niet anders dan met Prof. Anne Willems, Anne, de persoon door wie dit doctoraatsproject mogelijk was. Het is ondertussen iets meer dan vijf jaar geleden dat u mij uitnodigde voor een gesprek over een Antarcticaproject en mij ook de kans gaf om dit aan te vatten. Ik wil u niet alleen bedanken om mij deze kans te geven, maar natuurlijk ook voor alle hulp bij het opzetten van de experimenten, het analyseren van de data, het schrijven van de artikels en, last but not least, de niet-wetenschappelijke momenten.

I would also like to thank the members of my examination committee for taking the time to review my dissertation and for their helpful comments.

Bjorn, Helen en Pia, waar moet ik beginnen… Ooit zaten we samen op het nulde. Nu, zovele jaren later hebben jullie al alle drie andere oorden opgezocht (en blijf ik hier eenzaam en alleen achter op het labo), maar toch is het contact er niet minder op geworden. Dankjewel voor alles! Al is er ondertussen veel veranderd (Helen is nu een getrouwde vrouw) en staat veel op het punt te veranderen (Pia verwacht een klein boeleke en Bjorn begint als Verschrikkelijke Sneeuwman in Spitsbergen en Antarctica), toch hoop ik dat we nog regelmatig leuke momenten en Delicieuxlunches mogen beleven. En Pia, weet dat ik nog steeds in blijde verwachting ben van je handgemaakte en -geschilderde soepbord.

Bram, Charles (aka Xavier, aka Carlos), Chris, Prem, Frédéric en Pieter-Jan (aka PJ), onze tijd samen was van korte, doch krachtige duur. Merci voor de kwart-voor-vier-lachmomentjes, het gezever en gezwans, en natuurlijk ook de professionele momenten. Onze eigenste Wall of Fame zal de toekomstige werknemers nog vele jaren herinneren (of toch zolang de verf het volhoudt…) aan hoe tof het was toen wij er waren. Velen zullen proberen, maar niemand zal

er in slagen ons te evenaren, want onze doortocht gaat de geschiedenisboeken in als de Gouden Jaren!

Anique, Yihua, Pablo, Arianna, Francesca and Maria-Francesca, thank you all for the nice moments, both in the lab and at various places in and near Ghent. Jessy, binnenkort zoem jij er vandoor op de tonen van Dalida's grootste meesterwerk: Jessy apicoltore. Spreid je vleugels en verken de wereld, maar doe het wel niet te vroeg. September is naar het schijnt een mooie maand om het nest te verlaten.

Ondertussen zal ze misschien al denken dat ik haar vergeten ben, maar niets is minder waar. Het schrijven van mijn ode aan jou lukte gewoon niet vroeger wegens het constant volschieten van mijn gemoed. Zelfs nu heb ik de grootste moeite van de wereld om de tranen niet langs mijn wangen te laten stromen. Liesbeth, mijn lama, er bestaat in feite geen enkel woord in eender welke taal om te zeggen hoe dankbaar ik ben voor al je hulp. Eerst was er, samen met Evie (nog zo'n klassevrouw), de vijfsterren sequencing service. Later volgden dan de experimenten met ons 'beest'. Natuurlijk deed je zoveel meer dan dat. Iedereen die over jou hoort mag terecht jaloers en bedroefd zijn dat ze geen zo'n collega hebben en ooit zullen hebben. Ik ben er dan ook zeker van dat jij de muze was voor de wereldhit "Je te donne" van Jean-Jacques Goldman.

Natuurlijk zijn er nog tal van andere collega's die me de afgelopen jaren hebben bijgestaan met raad en daad en daarom niet mogen ontbreken. Jeannine, merci om altijd alle benodigdheden te bregen en om schoon te maken wat ik vuil maakte. Maarten, merci voor de gezellige momenten en de MALDI-TOF MS service. Kim en Geert H., merci voor de babbels en de hulp bij de experimenten. Mario en Bart H., merci voor technische hulp en de onnozele momenten. Marjan en de twee Sofies, merci om zo'n tijgers te zijn. Timo, Leen, Cindy en Jindrich, merci voor de vele ontspannende, uiteraard zeer hoogstaande gesprekken. Epifania, bedankt voor het helpen bij de vele adinistratie de afgelopen jaren.

Daarnaast zijn er ook tal van mensen naast het werk die een belangrijke rol gespeeld hebben. Master William, Sir Rudiger en Don Thomas, het begon lang geleden aan het Boerekot. Ondertussen zovele jaren later hebben we al veel meegemaakt. Reeds enkele jaren volgen we de wijze raad van Madonna op. Tussen ons gezegd: *I've got the feeling that if we took a holiday, took some time to celebrate, it would be so nice*! Dus *spread the word*, want

binnenkort is het weer *time for the good times*! Hannes en Jonathan, bedankt voor alle leuke momenten. Dat er nog veel mogen volgen. En natuurlijk ook alle vrienden van KSA Lo voor het bezorgen van zoveel leuke momenten en herinneringen.

Uiteraard was dit alles niet mogelijk geweest zonder mijn familie, en dan in het bijzonder mijn ouders. Papa en mama, bedankt voor alles! Zonder jullie steun, motivatie en blijvend vertrouwen was dit nooit gelukt. Alles wat jullie mij gegeven hebben sinds dag één was het grootste geschenk dat iemand me ooit gegeven heeft. Mathieu, Valentine, opa en oma, jullie mogen natuurlijk ook niet ontbreken. Dikke merci voor alle steun, hulp en veelvuldige babbels de afgelopen jaren en alle jaren daarvoor.

Guillaume Tahon

Gent, 5 mei 2017

A

AA	amino acid
AAP	aerobic anoxygenic photosynthetic bacteria
ADP	adenosine diphosphate
Ala	alanine
anfD/AnfD	iron-dependent nitrogenase α subunit gene / enzyme
anfG/AnfG	iron-dependent nitrogenase δ subunit gene / enzyme
anfH/AnfH	iron-dependent nitrogenase γ subunit gene / enzyme
anfK/AnfK	iron-dependent nitrogenase β subunit gene / enzyme
APB	anoxygenic phototrophic bacteria
Arg	arginine
ATP	adenosine triphosphate

B

bchB/BchB	dark-operative protochlorophyllide oxidoreductase B subunit gene / enzyme
Bchl	bacteriochlorophyll
bchL/BchL	dark-operative protochlorophyllide oxidoreductase L subunit gene / enzyme of anoxygenic phototrophs
bchLNB/BchLNB	dark-operative protochlorophyllide oxidoreductase L, N and B subunits genes / enzymes of anoxygenic phototrophs
bchN/BchN	dark-operative protochlorophyllide oxidoreductase N subunit gene / enzyme
bchX/BchX	chlorophyllide oxidoreductase X subunit gene / enzyme
bchXYZ/BvhXYZ	chlorophyllide oxidoreductase X, Y and Z subunits genes / enzymes
bchY/BchY	chlorophyllide oxidoreductase Y subunit gene / enzyme
bchZ/BchZ	chlorophyllide oxidoreductase Z subunit gene / enzyme
BLAST	basic local alignment tool
bp	base pair
BP	before present
BSA	bovine serum albumin

C

C_2H_2	acetylene
C_2H_4	ethylene
C_2H_6	ethane
CBB	Calvin-Benson-Bassham
cbbL	type I ribulose bisphosphate carboxylase large subunit gene
cbbM	type II ribulose bisphosphate carboxylase large subunit gene
Chl	chlorophyll
Chlide	chlorophyllide
chlL/ChlL	dark-operative protochlorophyllide oxidoreductase L subunit gene / enzyme of oxygenic phototrophs
chlLNB/ChlLNB	dark-operative protochlorophyllide oxidoreductase L, N and B subunits genes / enzymes of oxygenic phototrophs
CN^-	cyanide
CO_2	carbon dioxide
COR	chlorophyllide oxidoreductase

D

D1	photosystem 2 D1 subunit gene / enzyme in Cyanobacteria
D2	photosystem 2 D2 subunit gene / enzyme in Cyanobacteria
DNA	deoxyribonucleic acid
DPOR	light-independent protochlorophyllide oxidoreductase
DPOR	dark-operative protochlorophyllide oxidoreductase

E

e^-	electron

F

Fe	iron
Fe^{2+}	ferrous iron

G

Ga	giga annum
Glu	glutamic acid

H

H^+	hydron
H_2	dihydrogen gas
H_2S	hydrogen sulfide

I

Ile	isoleucine
IMG	integrated microbial genome database
LH	light-harvesting
LPOR	light-dependent protochlorophyllide oxidoreductase

M

MALDI-TOF MS	matrix assisted laser desorption/ionization time-of-flight mass spectrometry
MEGA	molecular evolutionary genetics analysis
MG-RAST	metagenomics-rapid annotation using subsystem technology
ML	maximum likelihood
Mo	molybdenum

N

N_2	dinitrogen gas
$NADP^+$	nicotinamide adenine dinucleotide phosphate
NADPH	reduced nicotinamide adenine dinucleotide phosphate
NCBI	national center for biotechnology information
NH_3	ammonia
NH_4^+	ammonium
nifD/NifD	molybdenum-dependent nitrogenase α subunit gene / enzyme
nifH/NifH	molybdenum-dependent nitrogenase γ subunit gene / enzyme
nifK/NifK	molybdenum-dependent nitrogenase β subunit gene / enzyme

O

O_2	oxygen
OD	optical density
ONU	operational nitrogenase unit
OPU	operational puf unit
ORU	operational RuBisCO unit

OTU	operational taxonomic unit

P

Pchlide	protochlorophyllide
PE	paired-end
PEG	protein encoding gene
PES	Princess Elisabeth Station
Pi	phosphate ion
PR	proteorhodopsin
Pro	proline
PS	photosystem
psaA/PsaA	photosystem 1 A subunit gene / enzyme in Cyanobacteria
psaB/PsaB	photosystem 1 B subunit gene / enzyme in Cyanobacteria
pscA/PscA	type 1 photochemical reaction center gene / enzyme in Chlorobi and Acidobacteria
pshA/PshA	type 1 photochemical reaction center gene / enzyme in Firmicutes
pufL/PufL	photosynthetic reaction center subunit L gene / enzyme
pufLM/PufLM	photosynthetic reaction center subunits L and M gene / enzyme cluster
pufM/PufM	photosynthetic reaction center subunit M gene / enzyme

R

RC	reaction center
RNA	ribonucleic acid
rRNA	ribosomal RNA
RuBisCO	ribulose-1,5-biphosphate carboxylase/oxygenase

S

SRM	Sør Rondane Mountains
SSU	small subunit

T

Thr	threonine
TOC	total organic carbon

U

UT	Utsteinen
UV	ultraviolet

V

V	vanadium
vnfG	vanadium-dependent nitrogenase δ subunit gene / enzyme
vnfH/VnfH	vanadium-dependent nitrogenase α subunit gene / enzyme

Table of Contents

Chapter 1

General introduction

1.1. Antarctica

1.1.1. The Antarctic continent

Stories about the existence of *Terra Australis Incognita*, a landmass in the far south to balance the northern continents, date back to the first century A.D. However, Antarctica, Earth's southernmost continent, remained a hypothetical continent until its discovery in 1820. With a surface area of ~13.6-14 million km², Antarctica represents about 10 % of the Earth's land surface [1, 2]. Throughout most of Antarctica's history, the continent's climate was temperate to sub-tropical. However, a combination of gradually declining atmospheric CO_2 partial pressure and tectonic reconstructions led to the onset of a large-scale glaciation of Antarctica ~33.7 million years ago [3, 4]. Nowaydays, almost entirely enclosed within the Antarctic Polar Circle (66° 33' 46.7" S)[1] (Figure 1), the continent is a barren wasteland which has been isolated from the rest of the world for over 10 million years [5, 6]. Due to historical events, its geographical isolation, and more importantly, its geographical location, the continent is nearly completely covered by ice [7], which accounts for approximately 90 % (25.4 M km³) of the Earth's ice [8, 9]. Permanently ice-free regions, however, are extremely scarce and constitute only 0.32 to 0.4 %, or 44 000 to 56 000 km², of Antarctica's surface (Figure 1) [4].

Antarctica as a whole can be divided into four major geographical regions: West and East Antarctica, the Antarctic Peninsula and the Sub-Antarctic region. The first two (*i.e.* West and East Antarctica) are, however, by far the largest units, separated from each other by one of the longest (~3500 km) mountain ranges on our planet, the Transantarctic Mountains (Figure 1) [10, 11]. West Antarctica, or Lesser Antarctica, is the 'lower' part of the continent (on average 850 m high) that lies in the Western Hemisphere, and is covered by the West-Antarctic ice sheet (~1.97 M km²). East Antarctica, on the other hand, comprises about two-thirds of the total area of the continent (Figure 1) and is covered by the much thicker East-Antarctic ice sheet with a size of ~10.35 M km² and an average altitude of 2300 m [6, 10, 12]. The Antarctic Peninsula, sometimes considered part of West Antarctica and covered by the remains of what once was the Antarctic Peninsula ice sheet, is the only part of the continent that extends a significant way northwards from the main ice sheet, its tip close the 63° S (Figure 1) [10, 13]. With an average width of 70 km and a mean height of 1500 m, this mountainous region is in fact a bedrock archipelago that acts as a barrier between its east and

[1] As of April 28, 2017 (http://neoprogrammics.com/obliquity_of_the_ecliptic/).

west coast atmospheric and oceanic circulations [6, 10]. Finally, the Sub-Antarctic region (45-55° S) encompasses a ring of oceanic islands characterized by milder environmental conditions compared to the more Southern regions [14].

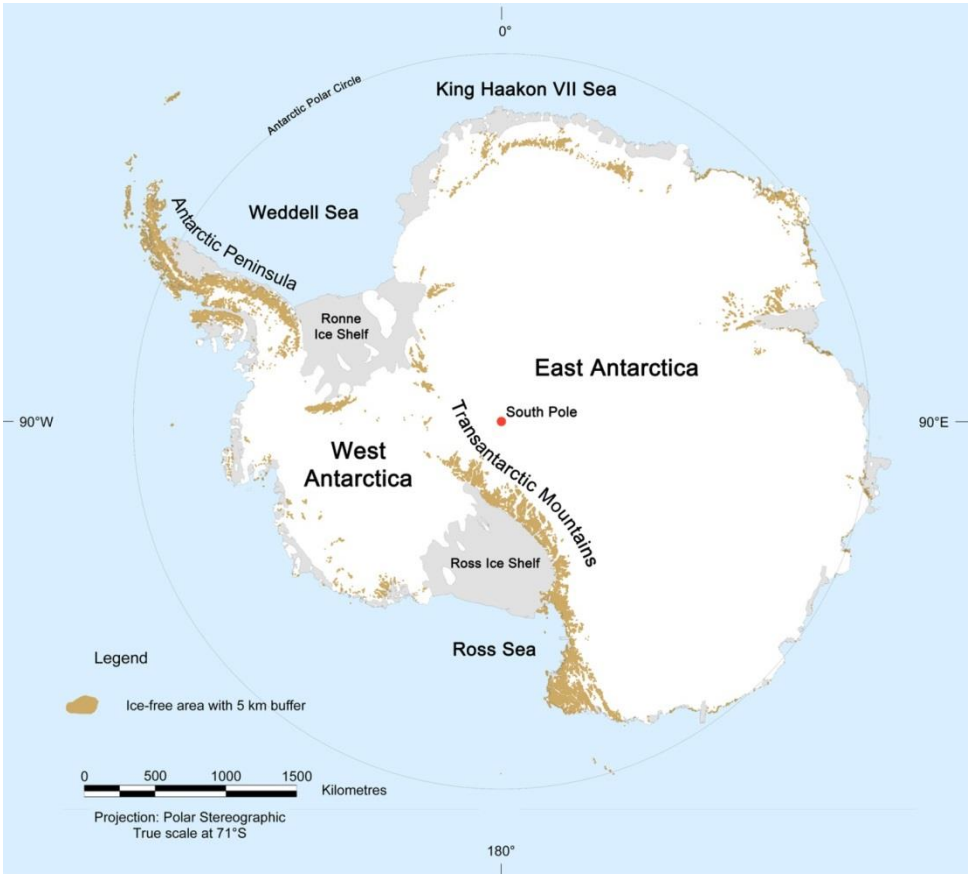

Figure 1 Overview of the Antarctic continent showing the major geographic regions. Ice-free areas are highlighted in brown (modified from Australian Antarctic Data Centre, map no. 13766 [15])

If anything, Antarctica can be denominated a continent of extremes. Unlike any other place on Earth, this frozen continent is constantly dominated by extreme environmental conditions (*e.g.* limited organic nutrients, low humidity, frequent freeze-thaw and wet-dry cycles), resulting from a series of factors [16].

The most commonly made association with Antarctica is its low temperature, caused by, among others, the permanent ice cover creating an albedo effect (*i.e.* reflecting solar radiation), an average altitude of over 2 km, and long periods of (nearly) complete darkness (*i.e.* the austral winter) combined with an aphelion position [6, 12]. Although the all-time record of lowest temperature (-93.2 °C) was recorded on Antarctica in 2010[2], yearly average

[2] Source: http://www.nasa.gov/content/goddard/nasa-usgs-landsat-8-satellite-pinpoints-coldest-spots-on-earth.

temperatures – the Antarctic Peninsula not included – usually range from -25 to -70 °C in winter time and -4 to -30 °C during the summer, the warmest temperatures appearing in the coastal areas [6]. Average annual temperatures on the Peninsula are, due to its lower latitude, much higher, although clear differences can be observed between its west and east coast. While the west coast of the Peninsula is characterized by a mild maritime climate with average annual temperatures of approximately -1.8 °C, temperatures at the east coast are about 7 °C lower as a result of southerly winds and a greater extent of sea ice [6].

As an immediate consequence of the below-zero temperatures, most precipitation sublimates before it reaches the surface. The remaining precipitation comes in the form of snow, most of which falls from March-May, although significant regional variation across the continent exists [17, 18]. Throughout the continent, the amount of precipitation is, however, extremely low, with up to only 600 mm yr^{-1} water equivalent (*i.e.* 20.7 mm yr^{-1} global sea level), and an average of 6 mm global sea level equivalent each year (*i.e.* 175 mm yr^{-1} water equivalent) (Figure 2) [11, 19, 20]. This makes Antarctica the largest cold desert on Earth, since desert climates are characterized by annual precipitation rates of less than 250 mm water equivalent [21].

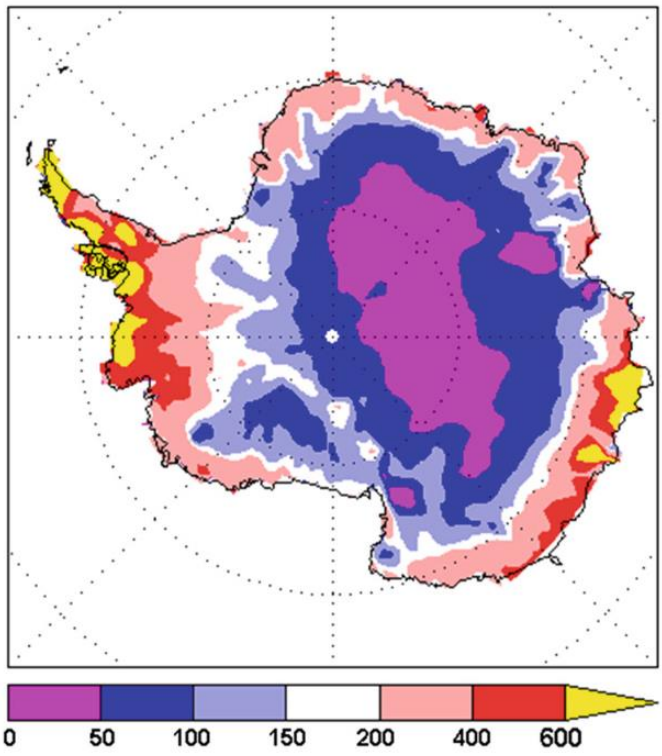

Figure 2 Distribution of mean-annual precipitation in Antarctica (in mm yr^{-1} water equivalent) (taken from Bockheim, 2015 [20])

1.1.2. Life on the frozen continent

Although it was for a long time thought that life was absent from continental Antarctica, in recent years it has become clear that this frozen continent is anything but a sterile environment and harbors a diverse, although depauperate, fauna and flora [22].

Only two species of native vascular plants have been recorded, spread only in the more favorable oases of maritime Antarctica (*i.e.* the Peninsula and several of the sub-Antarctic islands) [23-25]. Here, these Antarctic hair grasses (*Deschampsia antarctica*) and Antarctic pearlworts (*Colobanthus quitensis*) occupy all favorable places for growing including rocks, hollows, beaches and corniches [25]. More recently, their populations started to increase markedly in size and they started to slowly expand their habitat more southwards due to increasing temperatures [10, 26]. In the more severe inland surroundings, the flora is, however, predominantly cryptogamic, and dominated by algae, mosses (> 100 species) and lichens (> 200 species), which can be found up to latitudes of 86° S [22, 25, 27, 28].

The continental Antarctic fauna harbors no native mammals, amphibians or reptiles, but consists entirely of invertebrates, and then with only two higher insect species present: *Parochlus steinenii* and *Belgica antarctica*, both belonging to Diptera [24]. Instead, these communities are dominated by microinvertabrates (nematodes, rotifers, tardigrades), protozoans and microscopic arthropods (mites, springtails) [24, 28].

The bulk of Antarctica's continental diversity lies in the less visible groups, represented by the microbiota. In the early years of Antarctic research, culture-dependent studies provided only a glimpse of the actual resident microbial diversity. More recently, with the rise of metagenomics and next-generation sequencing techniques, a better understanding of its structure and functioning has been provided. Nonetheless the Antarctic bacterial diversity, dominated by members of the Acidobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria, Firmicutes and Proteobacteria, is considerably lower than in many temperate systems [28, 29]. Investigation of the individual species composition, however, suggests selection for a specialized, highly adapted microbial community, especially in the representation of nutrient limitation stress-response, thermal and osmotic pathways, and primary metabolic functions including diazotrophy and autotrophy [28, 30-32].

1.1.3. Terrestrial Antarctic systems

Ice-free regions are extremely scarce throughout the Antarctic continent, comprising only 0.32 to 0.4 % of the total surface area (Figure 1) [4]. The largest exposed regions constitute of mountain ranges such as the Transantarctic Mountains (24 200 km²), the Ellsworth Mountains (~2100 km²), and the Pensacola Mountains (~1500 km²), which are sometimes considered part of the Transantarctic Mountains. Other large ice-free areas are mainly situated in coastal sites of the Peninsula (10 000 km²) and East Antarctica (~11 800 km²) [20].

For a long time it was uncertain whether the loose, sandy material devoid of plants and of an organic layer could be called soil [33]. However, the definition of soil[3] broadly encompasses the diverse range of Antarctic "soil" habitats, despite the complete absence of higher plants in all areas other than the Pensinsula [34, 35]. Furthermore, even the most extreme and apparently depauperate ice-free Antarctic areas can be considered as soil habitats, given that they all contain detectable – although sometimes very low – levels of organic carbon and microbial populations [34]. Although many different types of Antarctic soils exist, in general, they comprise a surface pavement (*i.e.* a layer of gravel, stones or boulders formed largely by weathering and the removal of fine materials mainly by wind action) and a seasonally thawed active layer over permafrost. The active layer, often loose and unconsolidated, ranges from only a couple of centimeters up to over one meter [36].

In the hostile, poor and barren Antarctic wasteland, nutrient-rich soils are often found near the coastal areas, where birds play an important role in soil modification, considered the major factor influencing soil organic carbon and nutrient levels. The effects of high levels of carbon (> 4 %), nitrogen (> 2 %) and phosphorus (> 1 %) in these ornithogenic soils is, however, not restricted to areas with direct seabird manure inputs [20, 37]. Bird trampling and unfavorable chemical conditions (*e.g.* low pH, high conductivity) result in rookeries that are almost entirely devoid of vegetation, leaving nutrient-rich surface soils susceptible to wind and water erosion. Due to these two transport mechanisms, nutrients can be transported over a wider region [20].

The majority of all Antarctic soils is, however, extremely poor, lacking high amounts of water (< 2 %) and nutrients [17, 38-40]. Centuries, or even millennia of the aforementioned extreme

[3] The Soil Science Society of America (2008) [35] defines a soil as "the unconsolidated mineral or organic matter on the surface of the Earth that has been subjected to and shows effects of genetic and environmental factors of: climate (including water and temperature effects), and macro- and microorganisms, conditioned by relief, acting on parent material over a period of time" or "the unconsolidated mineral or organic material on the immediate surface of the Earth that serves as a natural medium for the growth of plants".

environmental conditions, in combination with the absence of vascular plants, have depleted soil substrates, hence reducing the soil's capacity to retain water and nutrients. Furthermore, the absence of large numbers of primary producers has resulted in soils containing only very low amounts of organic matter. This in turn enforces a vicious circle in the soils, since absence of (liquid) water and nutrients, especially carbon and nitrogen, further subjects the soil's biota to starvation, desiccation, high salt concentrations and rapid temperature changes [17, 40]. Nevertheless, sudden environmental changes may, temporarily, restore the impoverished soils. For example, wind erosion transports allochthonous nutrients, whereas during summer, meltwater runoff wets soils [20, 41, 42]. Additionally, but less frequent and very localized, animal carcasses can create a nutrient hotspot, thus supporting the soil's biota [17, 43, 44].

1.1.4. Study area: the Sør Rondane Mountains

Throughout this dissertation, the (functional) soil bacterial diversity is reported from several exposed soil samples from the Sør Rondane Mountains, eastern Dronning Maud Land, East Antarctica. This ~220 km long mountain chain (71° 30' – 72° 40' S, 22 – 28° E), consisting of larger mountain ranges and isolated nunataks, is wedge-shaped and widens to the east (Figure 3). It is located 200 km inland from the King Haakon VII Sea and has mountains rising up to 1500 m above the ice sheet surface, its highest point being Mount Gjelsvikfjella (3000 m above sea level) [38, 45]. Together with two other mountain chains, the Mühlig-Hoffman and Wohlthat Mountains, the Sør Rondane Mountains are the predominant ice-free area in Queen Maud Land, comprising 900 km² of exposed surface [38].

In the western part of the mountain range, just north of the Sør Rondane Mountains, the zero-emission Belgian Princess Elisabeth Station was built during the Antarctic summers of 2006-2008 (Figure 3) [46, 47]. The research station is situated on a small, relatively flat, granite ridge (*i.e.* Utsteinen ridge, 71° 57' S 23° 20' E). The ridge, with an elevation of 1390 m above sea level, is 700 m long, approximately 16 m wide, oriented in a north-south direction and protrudes 20 m above the surrounding snow surface [47]. Further south, approximately 1 km away from the ridge, lies Utsteinen nunatak, a granite peak with a highest point at 1564 m above sea level [47]. Utsteinen is characterized by relatively mild climate conditions, compared to other nearby sites. Summer and winter average air temeratures are about -8 °C and -23 °C, respectively [47]. Nevertheless, the yearly temperature variation curve is similar to those of more continental locations. Because of its location in the north-west of the

mountain chain, Utsteinen ridge is protected, by the mountains, of the high katabatic wind[4] speeds [48]. Mean summer and winter wind speeds are 4.5 and 6 m s^{-1}, respectively, while the major wind component is from east to south-east [47].

In Utsteinen, near the research station, the most frequently observed life are snow petrel birds (*Pagodroma nivea*), who's rookeries all lie in the western part of the mountain chain. Towards the east, however, the number of birds seen diminishes rapidly [49]. Other life in the area is predominantly microscopic, although lichens are found, again mainly in the western part of the Sør Rondane Mountains [50]. To monitor microscopic life in the proximity of the station, as well as the impact of humans and the research station itself, environmental samples have been collected prior to, and after the construction of the Princess Elisabeth Station, during several field sampling studies. Few biological studies have been reported from the Sør Rondane Mountains, focusing either on the invertebrates [49, 51, 52], lichens [50], the cultivable bacterial diversity [46, 53], or the total bacterial community composition [54, 55], hence leaving the functional potential of these ice-free areas an unexplored topic.

The research conducted in this thesis involves four environmental samples from Utsteinen. Although these samples are not representative of the entire region, their analysis will allow the first insights in functional potential and a preliminary reconstruction of the bacterial food web inhabiting the investigated sites. Given the objectives of the thesis, and the absence of many environmental and geochemical parameters of the samples at the start of the experiments, sample selection was based on the following criteria: i) availability of enough sample mass, ii) sample exposedness, iii) visual absence of cyanobacteria and algae and iv) concentration of extracted DNA. Based on these criteria, samples KP2, KP15, KP43 and KP53 were selected. All four samples were collected during the Antarctic summer of 2008-2009 (31 Jan 2009 to 03 Feb 2009), originate from Utsteinen and consist of weathered granite parent material. Photographs of the samples can be found in Chapter 2.

[4] Katabatic wind is the common name for downslope winds flowing from high elevations of mountains, plateaus, and glacier down their slopes to the valleys or planes underneath(Source: Kumar 2014 [48]).

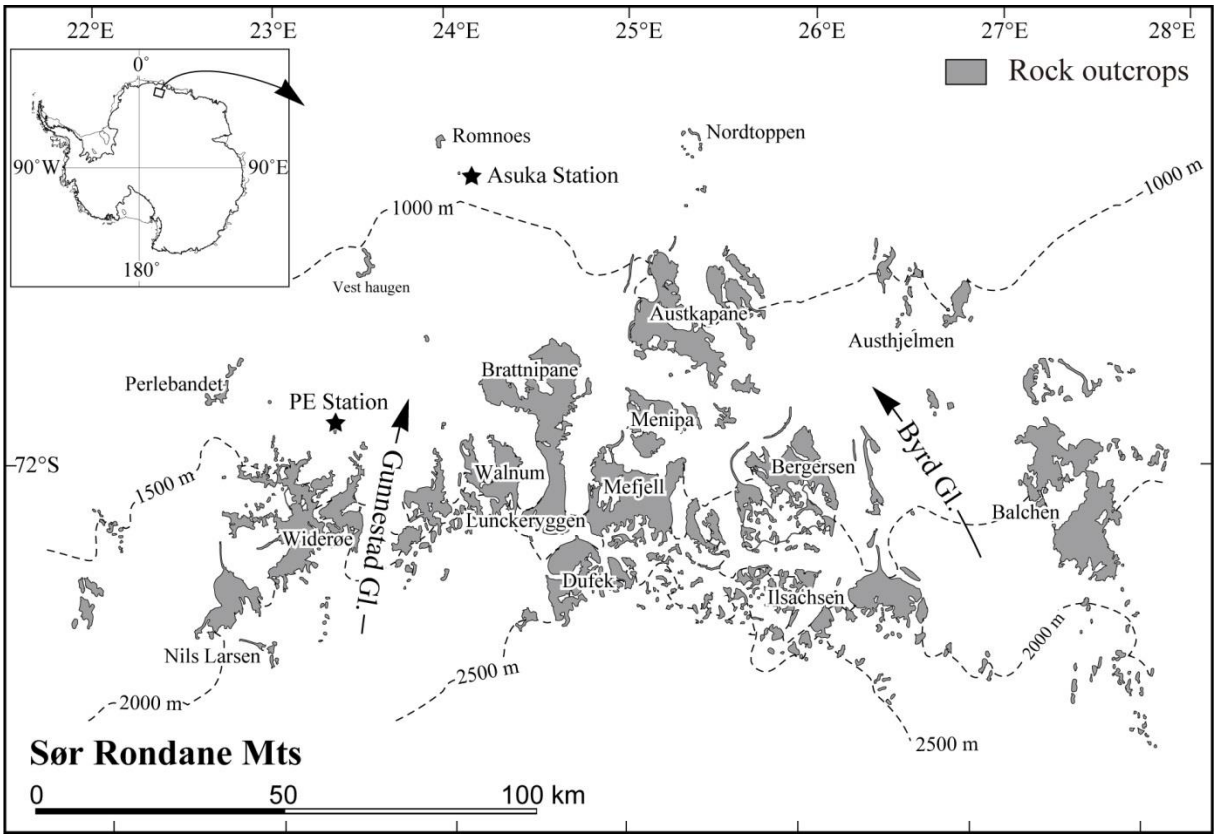

Figure 3 Location and setting of the Sør Rondane Mountains. PE = Princess Elisabeth Station (taken from Suganuma et al., 2014 [56])

Box 1. Metabolic diversity

All cells require an energy source and a metabolic strategy for conserving energy from it to drive energy-consuming processes. Energy can be obtained from two natural sources: light and chemicals (Figure 4). Organisms that conserve energy from light are called phototrophs, whereas those relying on chemicals are called chemotrophs [57].

In addition to energy, all cells require carbon for growth. Depending on the carbon source used, organisms can be considered autotrophs or heterotrophs. The first require carbon dioxide (CO_2) as their carbon souce, whereas the latter use organic compounds as their carbon source. Phototrophs can be thus be further subdivided in to photoautotrophs and photoheterotrophs, whereas chemotrophs can be subdivided in to chemoheterotrophs and chemoautotrophs [57].

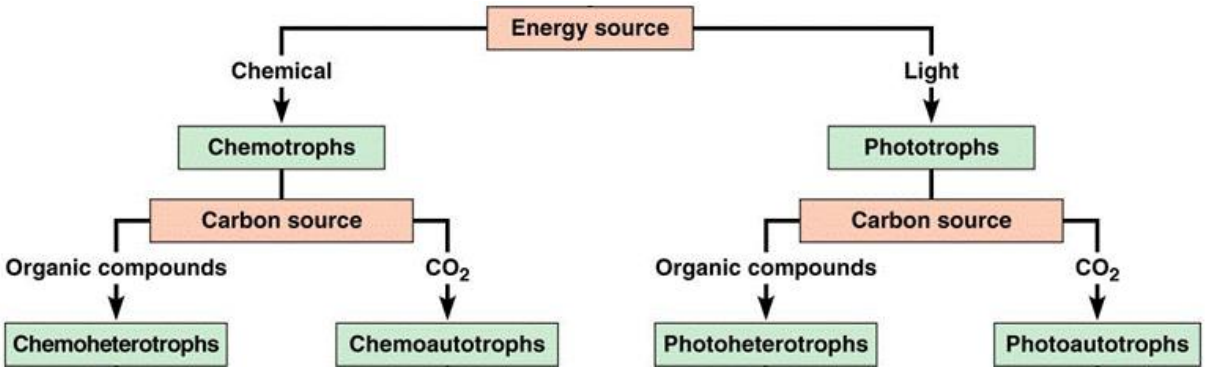

Figure 4 Metabolic options for conserving energy (Source: http://logyofbio.blogspot.be/2015/11/)

1.2. Autotrophy

1.2.1. Carbon fixation on Earth

Ultimately, all ecosystems are based on the input of carbon and energy by autotrophic organisms, making autotrophy without a doubt one of the, or maybe even the most important biogeochemical process in the world [58, 59], influencing our planet since the very beginning of life. Although different hypotheses about the evolution of the earliest life forms exist [60-63], it is thought that cells resembling prokaryotes appeared approximately 3.8 billion years (Ga) ago, during the Archean era [64]. These organisms are suggested to have been chemoautotrophs [64], microorganisms gaining the energy for CO_2 fixation from the oxidation of inorganic compounds [65]. Later on (~3.5 Ga ago), a second autotrophic process, photoautotrophy, evolved, enabling organisms to gain energy from light [66]. In time, autotrophic processes became spread among different taxa of bacteria, archaea, algae and plants, enabling these organisms to synthesize all of the necessary building block of life through the bioconversion of carbon in its most oxidized form (CO_2), to more reduced organic carbon compounds required for their metabolism [67]. This way, 104-161 petagram of carbon dioxide is yearly reduced to organic compounds, although extreme climate events can greatly influence this amount [68].

1.2.2. Diversity of autotrophic pathways

1.2.2.1. The Calvin-Benson-Bassham cycle

Anno 2017, six CO_2 fixation pathways that serve autotrophic growth are documented. Taken together, these pathways are distributed within different groups of organisms in all domains of Life [69]. The Calvin-Benson-Bassham (CBB) cycle (Figure 5), however, is responsible for sequestrating the majority of all CO_2 fixed by these six pathways [70]. This scheme, described in 1954 [71], was long thought to be the only pathway through which carbon dioxide could be fixed. It is, nonetheless, considered the most important cycle, due to its high abundance in plants, algae, archaea and bacteria, hence receiving the majority of scientific attention [72, 73].

The enzyme responsible for the actual fixation of CO_2 in the CBB cycle is ribulose-1, 5-biphosphate carboxylase/oxygenase (RuBisCO), by far the most abundant enzyme in the world (Figure 5) [74, 75]. RuBisCO, encoded by highly conserved *cbb* genes, combines one molecule of carbon dioxide with ribulose 1,5-biphosphate, yielding an unstable six-carbon intermediate which immediately decays in to two molecules of 3-phosphoglycerate [76]. Eventually, fructose-6-phosphate is formed, which can be assimilated into biomass [69]. All other enzymes of the CBB cycle only serve the regeneration of ribulose 1,5-biphosphate [76]. The Calvin cycle, however, is energetically very expensive, since, for the formation of one molecule of triose phosphate from three CO_2 molecules, it expends a total of nine ATP and six NADH molecules [77].

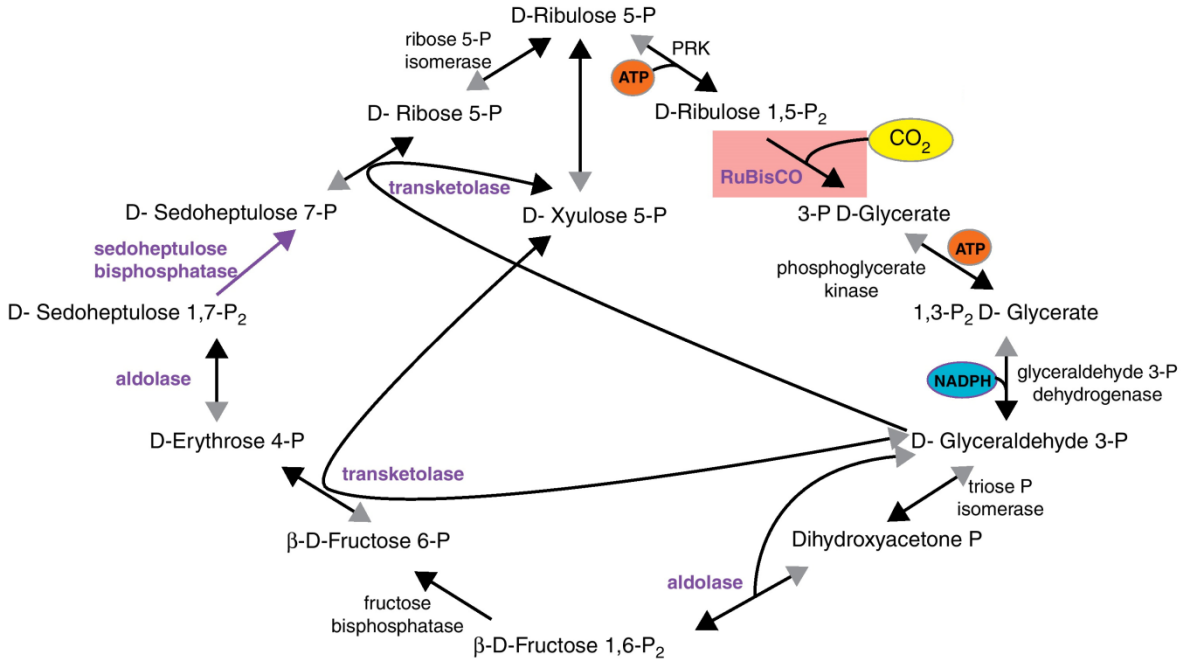

Figure 5 The Calvin-Benson-Bassham cycle. The carboxylating step involving RuBisCO is highlighted in red (modified from Ducat and Silver, 2012 [69])

1.2.2.2. RuBisCO diversity and structure

Four different natural types (I-IV) of RuBisCO are known, all of which differ in O_2 sensitivity, structure and catalytic properties (Table 1) [78]. The dominant hexadecamer type I shows considerable diversification and can be further subdivided into four distinct types, IA-ID, each consisting of eight large and eight small subunits (L_8S_8), encoded by *cbbL* and *cbbS* genes respectively [72]. In addition, type IA and IB enzymes can even be further subdivided into two distinct types. RuBisCO type IA consists of type IAc and IAq, based on the gene structure arrangements in the genomes. The major difference is that type IAc genes are associated with α-carboxysome genes, whereas the IAq genes are not. Instead, they are associated with the *cbbQ1* and *cbbO1* genes. Similarly, type IB enzymes can be subdivided in to IB and IBc to indicate the association of type IBc with carboxysomes in Cyanobacteria [72]. To date, RuBisCO type IA has been predominantly found in Alpha-, Beta and Gammaproteobacteria (types IAc and IAq) and to lesser extent in Cyanobacteria (type IAc) (Table 1). The different type IB enzymes, on the other hand, have been found in Cyanobacteria (type IBc), and green algae and plants (type IB) [72]. The distribution of RuBisCO type IC enzymes is a special case. To date, they have been detected in various bacteria, mainly belonging to Alpha-, Beta- and Gammaproteobacteria [72, 76]. More recently, however, RuBisCO IC enzymes were found in Actinobacteria [79, 80], Thermi [81], Verrucomicrobia [82] and Firmicutes [83]. Although several authors have suggested to group these enzymes in to a novel type I clade, RuBisCO type IE, it remains, to date, a putative type [79, 80, 82]. Finally, type ID RuBisCO enzymes are found in different non-green algae [72].

Often found in combination with RuBisCO type I, the type II enzymes, detected in dinoflagellates, and various – often anaerobic – chemolithotrophic and phototrophic Proteobacteria [70, 84, 85], consist only of large subunits (L_n), encoded by *cbbM* genes (Table 1). The number of subunits, however, varies from two to eight, depending on the organism. Although these enzymes share only ~25-30 % protein similarity with RuBisCO type I, they share several highly conserved regions [86].

Together with types I and II, type III enzymes have a ribulose 1,5-biphosphate dependent carbon fixing ability. They consist of a homo-10-unit polymer (L_{10}) and have up till now only been found in a few Archaea (Table 1) [87]. Finally, the last type, RuBisCO type IV, is considered not to be involved in the CBB cycle due to the lack of several of the required amino acid residues for the catalytic activity of RuBisCO. Therefore type IV is designated as a nonbonafide RuBisCO or RuBisCO-like enzyme [72, 88].

Table 1 RuBisCO enzyme forms and their phylogenetic distribution

RuBisCO Type	RuBisCO subtype	Enzyme structure	Distribution
Type I	IAc	L_8S_8	Alpha-, Beta-, Gammaproteobacteria, Cyanobacteria
	IAq	L_8S_8	Alpha-, Beta-, Gammaproteobacteria, Actinobacteria
	IBc	L_8S_8	Cyanobacteria
	IB	L_8S_8	Green algae, Plants
	IC	L_8S_8	Alpha-, Beta-, Gammaproteobacteria
	ID	L_8S_8	Non-green algae
	IE	L_8S_8	Verrucomicrobia, Thermi, Actinobacteria, Firmicutes
Type II		L_n	Alpha-, Beta-, Gamma-, Zetaproteobacteria, Dinoflagellates
Type III		L_{10}	Archaea
Type IV		Undefined	Bacteria, Archaea

Although extremely widespread, the CBB cycle, and more importantly RuBisCO, is one of the most notoriously inefficient catalysts. In addition to its ability to carboxylate, it can also oxygenate[5] [89], compromising the carbon assimilation and hence resulting in a setback of the carbon fixation process. Furthermore, RuBisCO suffers from an exceedingly low specificity for CO_2 and O_2, and, in case of carboxylation, an extremely low turnover rate of only three to five carbon dioxide molecules per second. Remarkable, its inefficiency is what makes RuBisCO the most abundant protein on Earth, as large amounts are needed to compensate its ineffectiveness [90-92].

An explanation for the poor kinetic properties of RuBisCO lies in the atmospheric changes that have occurred in the past ~3.5 Ga. Where the early atmosphere contained hardly any oxygen, nowadays this gas makes up ~21 % of the composition of the air. Carbon dioxide levels on the other hand, have declined. These changes have caused the evolution of different RuBisCO types, all having unique kinetic properties; the driving force being the need to optimize these kinetic properties to match the O_2 levels and the CO_2 substrate availability in the catalytic environments in which the enzymes operate [72].

[5] At atmospheric concentrations of CO_2 and O_2, RuBisCO fixes one mole of O_2 for every three moles of CO_2 (source: Furbank and Taylor, 1995 [89]).

1.2.2.3. Other autotrophic pathways

In addition to the CBB cycle, five alternative pathways that serve autotrophic growth have been elucidated, four of which are cyclic: the reductive tricarboxylic acid cycle, the 3-hydroxypropionate cycle, the 3-hydroxypropionate/4-hydroxybutyrate cycle, the dicarboxylate/4-hydroxybutyrate cycle and the non-cyclic reductive acetyl CoA pathway [93]. Based on our current knowledge, they are, however, less widespread and less important. The cyclic cycles share two common features. In addition to utilizing acetyl-CoA/succinyl-CoA to fix carbon, they all incorporate inorganic carbon in to existing carbon backbones. Furthermore, a partial overlap exists between the enzymes used in these cycles [69].

The reductive tricarboxylic acid cycle is basically the citric acid cycle in reverse, although three enzymes have been replaced. However, due to the oxygen sensitivity of several of its enzymes, the reductive tricarboxylic acid cycle seems restricted to anaerobic or microaerophilic bacteria, mostly belonging to the Chlorobiales, Aquificales and Alpha-, Gamma-, Delta and Epsilonproteobacteria [93].

Originally discovered in *Chloroflexus aurantiacus*, the 3-hydroxypropionate fixation pathway is actually a bicycle, characterized by the formation of 3-hydroxypropionate, hence the name of the cycle. Although it seems restricted to members of the Chloroflexaceae, several, but not all, genes of the pathway have been detected in various Alpha- and Gammaproteobacteria [93].

Although Kandler and Stetter suggested an undefined carbon fixation pathway in the crenarchaeal *Sulfolobus brierleyi*, in 1981 [94], it took until 2007 before the outline of the 3-hydroxypropionate/4-hydroxybutyrate cycle was fully described [95]. Since then, the pathway, showing considerable differences with the 3-hydroxypropionate cycle, has been found in several crenarchaeal species [93].

The dicarboxylate/4-hydroxybutyrate pathway was the latest cyclic CO_2 fixation pathway to be elucidated [96]. First discovered in the hyperthermophilic crenarchaeon *Ignicoccus hospitalis*, it has, to date, also been found in several other taxa of Crenarchaeota [93].

Finally, the multi-functional non-cyclic reductive acetyl-CoA pathway, discovered in the beginning of the 1980s by Wood et al. [97], is employed both autotrophically and heterotrophically. Interestingly, acetyl-CoA, the product of the pathway, can either be directed to pyruvate and biomass production, or to acetate and energy conservation, making it the only autotrophic carbon fixation pathway that can be used for energy conservation [98]. The reductive acetyl-CoA pathway, however, uses two extremely oxygen sensitive enzymes. As a result, it is only found in anaerobic bacteria and archaea [93].

1.2.3. RuBisCO in Antarctica

RuBisCO enzymes are omnipresent on our planet, especially due to their presence in a variety of plants. About the distribution and ecology of bacterial RuBisCOs, especially in polar regions, a lot less is known. In addition, on Antarctica, most research groups have focused on aquatic environments (*i.e.* lakes), mostly located in or very close to the McMurdo Dry Valleys. The majority of RuBisCO research in that area was conducted by the group of Rachel Morgan-Kiss, who investigated the autotrophic communities of several Antarctic lakes. Their results suggested that a variety of photo- and chemoautotrophic bacteria, harboring RuBisCO types I and II, inhabited these environments, although large seasonal variations greatly influenced the communities [99-101]. Tebo et al. (2015), on the other hand, screened several ice cave systems in Mount Erebus for potential autotrophy [102]. In addition to screening for RuBisCO type I and II enzymes (*cbbL* and *cbbM*), they also investigated their sediment samples for presence of ATP citrate lyase (*adB* gene), a key enzyme in the reductive tricarboxylic acid cycle. Although PCR amplification of *cbbM* and *adB* was negative for all samples, the retrieved *cbbL* sequences showed clear differences between the different sample locations. Type IA sequences all grouped with that of *Bradyrhizobium*, whereas the majority of all sequences was distributed among the type IC (or IC and IE) phylogeny [102]. Finally, Niederberger et al. investigated arid and wet soils of the McMurdo Dry Valleys for type I RuBisCO. Similar to the finding of Tebo et al. [102], type IC was the most dominantly retrieved, whereas type IA remained undetected [103]. Apart from this relatively well-studied Antarctic region, little attention has been given to the diversity of RuBisCO in the rest of the continent. However, Nakai et al. studied *cbb* genes in a moss pillar originating from Lake Hotoke (Queen Maud Land). Their analyses, however, did not detect type IC, ID or III enzymes. Surprisingly, their type IA sequences were closely related to sequences obtained from deep-sea environments. In addition, one of their OTUs grouped closely with a sequence found in Lake Bonney (McMurdo Dry Valleys), indicating that RuBisCO genes from geographically distant Antarctic lakes can be phylogenetically related [87].

1.3. Diazotrophy

1.3.1. History of nitrogen fixation

Nitrogen is an essential component for all living organims. On the early Earth, atmospheric nitrogen may have been fixed by several abiotic processes (*e.g.* lightning discharge), albeit in very limited amounts, hence restraining the size of the primordial biosphere. At some point, however, these low concentrations of fixed nitrogen would have become insufficient to further expand the microbial biomass. The evolution of nitrogenase enzymes, reducing the relatively inert N_2 to the bioavailable NH_4^+, represented a breakthrough in the radiation of life. Although its exact date of origin remains unknown, isotopic evidence suggests that nitrogenase has been present since at least ~3.2 Ga ago [104, 105].

1.3.2. Biological nitrogen fixation

This N_2 to NH_4^+ reduction process, known as biological nitrogen fixation or diazotrophy, is a microbially mediated keystone process in many ecosystems, providing the members of the community with bioavailable forms of fixed nitrogen. In total, about 61 % (~250 Tg) of the total nitrogen, yearly added to the biosphere, is produced this way. However, only a very limited set of prokaryotes, the diazotrophs, can access the gigantic atmospheric N_2 reserves and reduce it to bioavailable ammonia [106]. Although, to date, no genes encoding for nitrogen fixation have been detected in Eukaryota, diazotrophs are widely distributed, in a paraphyletical way, across the bacterial and archaeal domains, suggesting nitrogenase emerged after the diversification of these two groups [105, 107]. Of all currently sequenced prokaryotic genomes, ~15 % contain nitrogenase [104]. Its distribution among Bacteria is however far broader compared to Archaea, where it has to date only been retrieved from methanogens within the orders of the Methanococcales, Methanobacteriales and Methanosarcinales (Euryarchaeota) [104, 105].

1.3.3. Nitrogenase

In general, nitrogenase is a dinitrogen reducing, ATP-hydrolyzing, oxygen-sensitive two protein component system, composed of the dinitrogenase heterotetramer ($\alpha_2\beta_2$, metal[6]-Fe-protein), with the α-subunit containing the active site for N_2 reduction, and the dinitrogenase-reductase homodimer (γ_2, Fe-protein) metalloproteins [108, 109]. Remarkably, even after decades of research, the overall reaction stoichiometry is still not unambiguously determined:

$$N_2 + (6+2n)H^+ + (6+2n)e^- + p(6+2n)ATP \rightarrow 2NH_3 + nH_2 + p(6+2n)ADP + p(6+2n)P_i$$

Most frequently, however, n = 1 and p = 2, meaning that per reduced molecule of N_2 a total of 16 ATP are hydrolyzed [110].

Based on the metal cofactor present in the active site (part of the α-subunit), three distinct nitrogenases have been identified until now. The majority of the present-day dinitrogen reduction is catalyzed by the Nif nitrogenase, which is Mo-dependent and considered the most universally distributed form in nature, making it the most extensively studied nitrogenase. Here, the dinitrogenase subunits are encoded by *nifD* and *nifK* genes, whereas the dinitrogenase-reductase homodimer is encoded by *nifH*. Sequences of the latter gene are remarkably homologous between different diazotrophs, indicating they encode a highly conserved protein family. As a result, *nifH* is an excellent marker to study the phylogeny of nitrogen fixing organisms and characterize aspects about their diversity and ecology [108, 109, 111]. In addition to reducing N_2, Mo-nitrogenase is also capable of reducing alternative substrates, including cyanide (CN^-), protons (H^+) and acetylene (C_2H_2), the latter being reduced to ethylene (C_2H_4) [112].

Although it was believed, for many years, that molybdenum was essential for biological nitrogen fixation, several studies performed in the 1970s suggested the existence of a vanadium-dependent nitrogenase. This was, however, only confirmed in 1986, with the isolation of a V-containing nitrogenase from two *Azotobacter* species [109, 113, 114]. Similar to the Mo-dependent dinitrogenase-reductase, the vanadium-dependent Fe-protein is a homodimer encoded by the *vnfH* gene. The VFe-protein, on the other hand, differs from the MoFe counterpart in terms of its structure. They are mostly hexameric proteins ($\alpha_2\beta_2\delta_2$) due to the presence of an additional δ-subunit, encoded by the *vnfG* gene [109, 115]. Compared to its Mo-counterpart, this significantly less widespread enzyme consumes considerably more ATP,

[6] The metal is Mo, V or Fe.

lowering its efficiency [109]. It does, however, show an increased activity at low temperatures (*e.g.* 5 °C), compared to Mo-nitrogenase, which could be an important advantage in cold environments [116]. In addition, V-nitrogenase is capable of reducing the same substrates as its Mo-counterpart. However, unlike Mo-nitrogenase, which generates C_2H_4 as the sole product of C_2H_2 reduction, V-nitrogenase forms a small amount of C_2H_6 (~5 %) in addition to C_2H_4 [109, 112].

Finally, in 1980, Bishop et al. reported the extremely oxygen-sensitive Fe-only nitrogenase system [117]. Similar to the Mo- and V-dependent nitrogenases, the α-, β- and γ-subunits are encoded by *anfH*, *anfD* and *anfK* genes, respectively. The mature dinitrogenase, however, is a hexameric protein with the δ-subunits being the products of the *anfG* gene [109, 115]. Although this form has similar abilities as the aforementioned ones, it is considered the nitrogenase system with the lowest efficiency [109].

The Mo-dependent nitrogenase has been found to more widespread, and to more efficiently bind and reduce nitrogen than the other two forms (Nif > Vnf > Anf). Furthermore, molybdenum represses synthesis of alternative nitrogenases. However, interestingly, no organism has been found harboring an alternative (V- or Fe-dependent) nitrogenase without the Mo-dependent form [105, 108]. The metal-dependencies of these three nitrogenases, however, explain their co-existence. Since molybdenum is one of the least abundant essential trace metals in soils, its rareness would seriously constrain nitrogen fixation [118]. However, vanadium is usually 50-100 times more abundant than molybdenum in soils, suggesting this metal could be available upon Mo depletion [119]. In combination with the retrieval of genes encoding alternative nitrogenases in a variety of terrestrial environments, this suggests that alternative nitrogenases may contribute significantly to nitrogen fixation worldwide [120].

Apart from the three aforementioned oxygen-sensitive nitrogenases, a fourth nitrogenase system has been reported in the Actinobacterium *Streptomyces thermoautotrophicus* by Ribbe et al. in 1997. Unlike the three conventional enzymes, this novel nitrogenase was said to be oxygen-insensitive and to differ significantly in structure [121]. Recent analyses, however, stated there is no evidence for nor against the existence of an oxygen-tolerant nitrogenase in this bacterium. Genome sequencing failed to detect the proposed nitrogenase enzyme, while multi-site fixation experiments failed to demonstrate the incorporation of labelled $^{15}N_2$ gas into biomass. Finally, growth experiments suggested *Streptomyces thermoautotrophicus* to be highly effective nitrogen scavengers, rather than nitrogen fixers [122].

1.3.4. Nitrogenase classification

Phylogenetic analyses of each of the nitrogenase protein components indicates they segregate into five distinct clades (Figure 6). The majority of all known sequences groups in cluster I, which contains all cyanobacterial, as well as most proteobacterial *nif*, and *nif* from certain Actinobacteria (*Frankia*) and Firmicutes (*Paenibacillus*). The second cluster contains a relatively small number of alternative (V- and Fe-dependent) nitrogenases as well as sequences belonging to certain methanogenic Archaea. Cluster III is dominated by *nif* from predominantly anaerobic members of the Bacteria and Archaea including several methanogens, acetogens, spirochetes, clostridia, sulfate-reducing bacteria and green sulfur bacteria. Clusters IV and V contain *nif* paralogues detected only in some anoxygenic photosynthetic bacteria and methanogens (Cluster IV); and (bacterio)chlorophyll synthesis genes common to all phototrophs (Cluster V) [108, 123]. More detail about the genes/enzymes enclosed in Cluster V is provided in section 1.4.2.2.2.

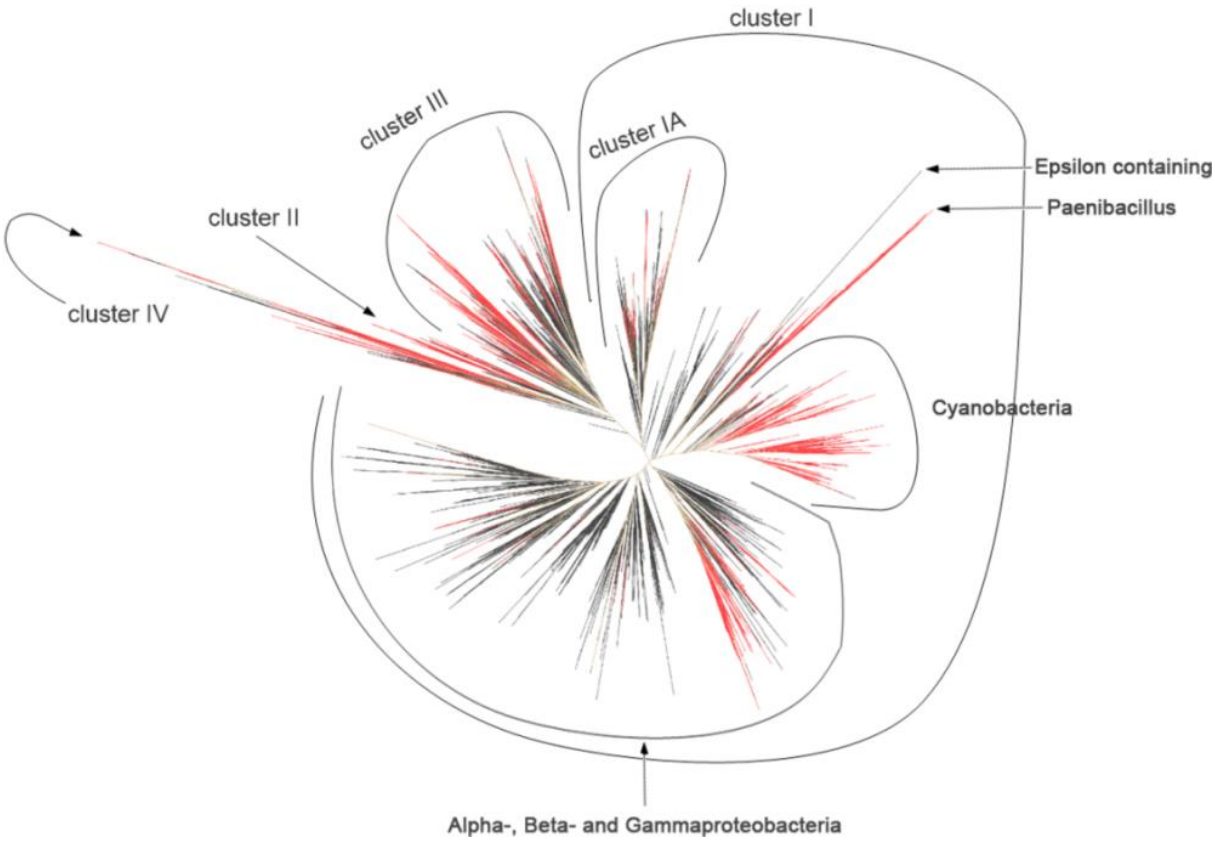

Figure 6 *nifH*/*vnfH*/*anfH* phylogenetic tree. Major phylogenetic groups are labeled in the Figure. Cluster IA contains primarily obligate anaerobes. Cluster V, which contains paralogues involved in (bacterio)chlorophyll synthesis is not included. Red branches are low G+C sequences (modified from Gaby and Buckley, 2014 [124])

1.3.5. Diazotrophy in Antarctica

Even though *nifH* is one of the most, and maybe even the most studied bacterial protein encoding gene in the world, few research groups have focused on its presence and diversity in Antarctica. Even today, sampling efforts clearly underrepresent Antarctica relative to other regions. Furthermore, the majority of all Antarctic nitrogenase research is restricted to only a very small region, the McMurdo Dry Valleys. Here, the investigation of diverse terrestrial samples, predominantly taken in the proximity of glaciers and lakes, revealed a broad diazotrophic community. Nitrogenase activity assays at different study sites indicated the presence of an active phototrophic (cyanobacterial) and heterotrophic N_2 fixing community, although substantial variability between nitrogen fixation rates could be observed at the different sites [125-129]. However, for some studied sites, the rates of nitrogen fixation activity were of the same magnitude as seen in temperate, subtropical and tropical shallow marine systems, indicating that temperature may not be a driving factor in nitrogen fixation rates [127]. PCR amplification of *nifH*, on the other hand, showed similar observations, although clear differences existed between the different sites, with *Nostoc* being the most dominant cyanobacterial component [125, 126, 128]. Studies on nitrogen fixing bacteria in other Antarctic regions are, however, much more scarce and focus mainly on the Antarctic Peninsula and the sub-Antarctic region [130-133]. However, in agreement with results obtained from the McMurdo Dry Valleys, the main nitrogen fixing organism in other Antarctic regions appears to be *Nostoc*, indicating that diazotrophic Cyanobacteria are the major contributors to N_2 fixation in Antarctica.

1.4. Phototrophy

1.4.1. Phototrophy on Earth

Ultimately, life revolves around entropy, and thus energy. All living organisms depend on a constant flux of energy in order to grow and reproduce. Of the two energy sources used by bacteria – light and chemical compounds – the first has been present since long before the dawn of our planet ~4.5 billion years ago [134]. Since the formation of our planet, light, and in particular sunlight, has always influenced Earth and, more importantly, the life on it. On an annual basis, the sun provides solar energy to our planet at a rate of ~3850 zettajoules, which is the equivalent to ~7000 times the energy mankind currently uses per year [135]. Although humankind developed technologies to capture sunlight and produce electricity [135], the world as we know it today would not exist without several natural light-harvesting (*i.e.* phototrophic) processes. Our atmosphere, environment, food, culture, etc. all depend, either directly or indirectly, on these processes, making phototrophy one of the most fundamental biological processes on Earth [136].

Two major types of phototrophy can be distinguished: (bacterio)chlorophyll-based phototrophy, which can be further subdivided in to oxygenic and anoxygenic photosynthesis, and rhodopsin-based phototrophy [137].

1.4.2. (Bacterio)chlorophyll-dependent phototrophy

1.4.2.1. Evolution of photosynthesis

Photosynthesis is the most important type of light-harvesting. In addition to fueling the production of more than 100 billion tons of dry biomass per year [135], it is responsible for maintaining the current normal levels of oxygen in our atmosphere [136]. Photosynthesis arose early in Earth's history, when our planet was characterized by extremely harsh environmental conditions [138, 139], and less light was present than now, due to the lower intensity of the young sun [140, 141]. The earliest forms of photosynthetic life appeared approximately 3.5 Ga ago and were almost certainly non-oxygen evolving, hence their name anoxygenic phototrophs (Figure 7). Based on carbon isotope composition measurements from ancient rocks, they are assumed to have used reductants such as H_2, H_2S and Fe^{2+}, but not

H_2O, using either a type 1 or 2 reaction center [142-144]. At that time, the Earth's atmosphere was composed mainly of nitrogen, carbon dioxide and methane, whereas oxygen was not present to an appreciable extent. Therefore, the organisms that lived then are thought to have been strict anaerobic species [143]. The genesis of oxygenic photosynthesis, only in Cyanobacteria, did not happen for another 700 million years, although there is still some uncertainty about the exact time of its first appearance (Figure 7). The strongest evidence, however, was found in lacustrine stromatolites in iron- and sulphur-depleted basins, suggesting that by ~2.72 Ga ago, oxygen-producing phototrophs had evolved [145]. In contrast to anoxygenic photosynthetic bacteria (APB), these oxygenic photosynthetic bacteria, using H_2O as a reductant and evolving O_2, had a type 2 reaction center connected in series with a type 1 (Figure 7) [144]. The appearance of oxygenic photosynthesis and hence the evolution of molecular O_2 by Cyanobacteria, however, did not immediately lead to significant changes in the Earth's atmosphere. Only 300 million years later, at the beginning of the Proterozoic era, the Great Oxidation Event led to a rise of atmospheric oxygen, altering the redox balance on Earth and forcing organisms to either adapt to oxygen, become extinct or retreat to anaerobic niches [142, 143, 145]. Under these new conditions, many of the older anaerobic anoxygenic photosynthetic bacteria probably disappeared or retreated to the remaining anoxic habitats. Some groups, however, adapted to the new aerobic conditions and embarked on an aerobic life style (aerobic anoxygenic phototrophic bacteria (AAP)) [146]. This new atmosphere revolutionized complex life, leading to the emergence of photosynthetic eukaryotes and eventually vascular plants that, in turn, increased oxygen levels to those observed in today's atmosphere [143].

1.4.2.2. Bacterial (bacterio)chlorophyll-dependent phototrophy

Phototrophic organisms show a broad distribution in the bacterial and eukaryotic domains of Life. In Bacteria, species using (bacterio)chlorophyll-dependent reaction centers have currently been found in seven phyla: Acidobacteria, Chlorobi, Chloroflexi, Cyanobacteria, Firmicutes, Gemmatimonadetes and Proteobacteria [147]. Within the phyla of the Acidobacteria and Gemmatimonadetes, however, only one phototrophic representative has been found so far [147, 148]. It should be noted, furthermore, that a potential phototrophic candidate has been found in the Actinobacteria (acc. no. KC465424). This, however, remains to be confirmed, as the data result from an unpublished study.

Phototrophic bacteria of different phyla are grouped together based on their phototrophic lifestyle (*i.e.* oxygenic photosynthetic bacteria, anaerobic photosynthetic bacteria and aerobic anoxygenic phototrophic bacteria). The latter organisms, discovered in 1978, are defined as oxygen-requiring bacteria which synthesize Bchl and use light energy as an auxiliary energy source for their mostly heterotrophic metabolism [146, 149, 150]. To date, no AAP bacterium is known to contain carbon fixation enzymes [151]. Although they rely on heterotrophy for 80 % or more of their cellular energetics, sunlight can double their organic carbon assimilatory efficiency compared to that of strict heterotrophs [152]. AAP predominantly belong to the Proteobacteria, although a single representative is known in the Gemmatimonadetes and in the Acidobacteria [146-148]. However, although aerobic anoxygenic phototrophy is a frequently studied topic, the inclusion of several bacterial groups in the AAP remains a matter of dispute even after nearly four decades of research. In many aquatic and terrestrial habitats, phototrophic bradyrhizobia and methylobacteria, of which some are facultative autotrophs, are common inhabitants [146, 153-155]. Currently, however, due to the potential presence of carbon fixation enzymes, the phototrophic species belonging to the genera *Bradyrhizobium* and *Methylobacterium* are not considered AAP although some authors originally chose to include them [156]. Phototrophic species of these two excluded groups share the same basic physiology with standard AAP. Most of them are aerobic heterotrophic species that synthesize Bchl and perform photosynthesis under aerobic conditions [155]. Furthermore, Kramer et al. showed that photosynthetic electron transport in *Bradyrhizobium* sp. BTAi is oxygen-dependent [157]. Therefore, in the future, these groups could be included in the AAP. This would, however, require a modification of the current definition of the term AAP, or the introduction of a novel term, grouping all these bacteriochlorophyll-containing bacteria.

1.4.2.2.1. (Bacterio)chlorophyll biosynthesis

Although (bacterio)chlorophyll-dependent phototrophy is a rather complex process taking place in the photochemical reaction center (RC), ultimately, absorption of light energy by chlorophyll (Chl) and bacteriochlorophyll (Bchl) pigments results in the preservation of chemical energy (*i.e.* as ATP) [158]. In turn, ATP synthesis often fuels one or more assimilatory reactions, for example assimilation of carbon dioxide or molecular nitrogen [144].

Before the actual energy conversion can take place, a series of complex reactions is required to generate (bacterio)chlorophyll. Throughout all (B)Chl-dependent phototrophic organisms, the metabolic precursor required to initiate the process is protoporphyrin IX, that, through a series of conversion reactions called the core pathway, is converted to protochlorophyllide (Pchlide) [159]. This highly unsaturated molecule, however, is unable to drive the necessary charge separation in the photochemical reaction centers due to its absorption in the low-energy region of the spectrum [160]. Therefore, during the final step of the core pathway, the saturation of protochlorophyllide is increased by converting it to chlorophyllide/chlorin (Chlide), a pigment that absorbs light in higher-energy regions of the spectrum, through one of two enzymes (Figure 7). The first enzyme, the oxygen-insensitive light-dependent protochlorophyllide oxidoreductase (LPOR), is found in angiosperms, gymnosperms and Cyanobacteria, whereas the second enzyme, the oxygen-sensitive light-independent or dark-operative protochlorophyllide oxidoreductase (DPOR), is found in Cyanobacteria, gymnosperms and anoxygenic phototrophic bacteria (Figure 7). Given that APB do not contain the light-sensitive variant, the evolution of LPOR seems to be in close connection with the development of oxygenic photosynthesis and hence the evolution of chlorophyll [159, 161, 162]. Although LPOR and DPOR catalyze the same conversion reaction, they are, however, very different enzymes.

After the conversion of Pchlide to Chlide, a major difference can be seen between oxygenic and anoxygenic phototrophs. In the APB, the biosynthesis pathway of Bchl is extended with an additional step to convert chlorin to bacteriochlorin (Figure 7). Interestingly, the enzyme responsible, chlorin oxidoreductase (COR), resembles DPOR, and probably evolved from the latter during a gene duplication event. Subsequently, bacteriochlorin is transformed into Bchl, whereas in oxygenic phototrophs, Chl is formed out of chlorin. The loss of COR was a main event during evolution, since the change of Bchl to Chl was a prerequisite for oxygenic photosynthesis to arise, because light absorbed by Bchl is not energetic enough to oxidize water [160, 162].

These (bacterio)chlorophyll pigments are eventually used in the photochemical reaction centers and the antenna complexes (Figure 7), where they absorb photons, hence starting a series of reactions that ultimately generate ATP.

1.4.2.2.2. Enzymes and genes involved in (bacterio)chlorophyll biosynthesis

DPOR. The reduction of the protochlorophyllide ring structure is a major regulatory step in de biosynthesis of (bacterio)chlorophyll. Two enzymes are responsible for this conversion: LPOR and DPOR. With part of the focus of this thesis on light-harvesting mechanisms in bacteria, the focus will be on the dark-operative variant (DPOR) in bacteria.

Dark-operative or light-independent protochlorophyllide oxidoreductase contains three subunits, making up two components: an oxygen-sensitive homodimeric L-protein (BchL) as the reductase component and a heterotetrameric NB-protein (BchN-BchB) as the catalytic component. The three subunits are encoded by *bchL*, *bchN* and *bchB*, which are possessed by all APB. Since Cyanobacteria use Chl, the cyanobacterial DPOR genes, that are homologs of the *bchLNB* genes, are called *chlLNB*, respectively [162-164].

Interestingly, the deduced amino acid sequences of *bchLNB*/*chlLNB* show significant similarity to those of the nitrogenase subunits. Similar to DPOR, nitrogenase is a three-subunit complex – composed out of a NifH dimer and a NifD-NifK heterotetramer – providing the molecular basis of biological nitrogen fixation by converting N_2 to NH_3. The highest similarity exists between BchL/ChlL and NifH (~30 %), whereas the degree of similarity between the N and B subunits with NifD and NifK, respectively, is only ~15-20 %.

These nitrogenase-like features of DPOR led to the hypothesis that both evolved from a common ancestor (Figure 7) [162, 163].

COR. As mentioned above, the Bchl biosynthesis pathway in APB is extended with an additional step to convert the chlorin ring to a bacteriochlorin ring. Interestingly, chlorin is further reduced by a second nitrogenase-like enzyme, COR. Similar to DPOR and nitrogenase, COR also is a three-subunit enzyme, with BchX, BchY and BchZ making up two components. The X-protein (BchX) plays a role as the reductase, whereas the YZ-protein (BchY-BchZ) provides the catalytic site for the reduction. Thus, ultimately, protochlorophyllide is converted to bacteriochlorophyllide by the sequential actions of two nitrogenase-like enzymes [160, 162].

Importantly, the three proteins of COR exhibit sequence similarity not only with those of nitrogenase, but also with those of DPOR. Thus, three structurally related enzyme complexes exist: nitrogenase, DPOR and COR, which all evolved from a common ancestor (Figure 7) [162, 164].

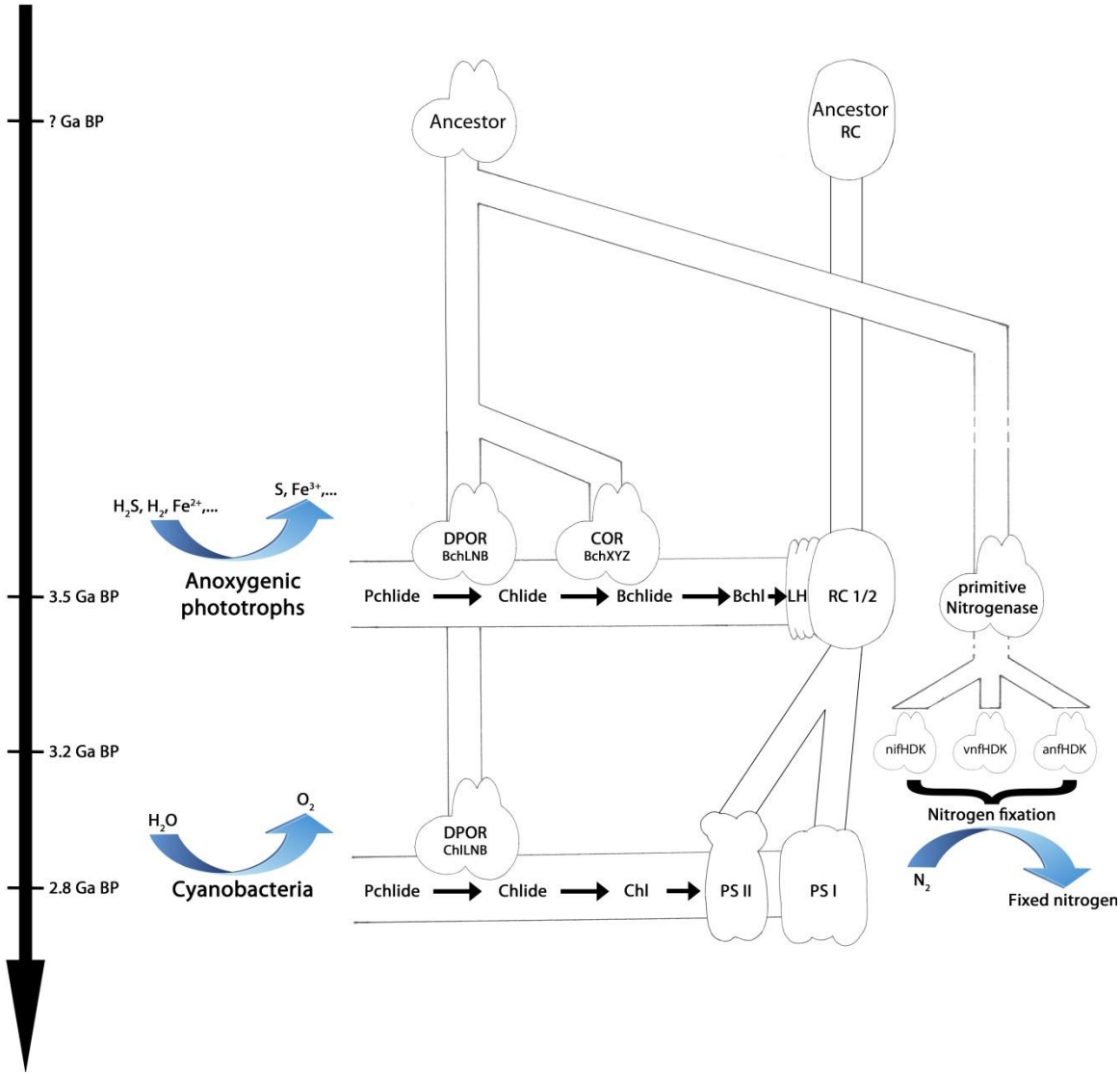

Figure 7 Evolutionary scheme of bacterial (bacterio)chlorophyll biosynthesis systems. Dark-operative protochlorophyllide oxidoreductase (DPOR), chlorin oxidoreductase (COR) and nitrogenase evolved from a common ancestor. In anoxygenic phototrophs, DPOR and COR catalyze the reduction of protochlorophyllide (Pchlide) to chlorophyllide (Chlide), and chlorophyllide to bacteriochlorophyllide (Bchlide), the immediate precursor to bacteriochlorophyll (Bchl), respectively. During evolution of Cyanobacteria (oxygenic photosynthetic bacteria), COR was lost due to a truncation of the bacteriochlorophyll biosynthesis pathway, with Chlide as the precursor of chlorophyll (Chl). Photosynthetic reaction center (RC) 1 and 2 both evolved from a common ancestral RC. During evolution to Cyanobacteria, RC 1 and 2 were incorporated in photosystem (PS) I and II, respectively. Ga: giga-annum; BP: before present; LH: light-harvesting. Modified from Fujita and Bauer (2003) [162]

1.4.2.2.3. (Bacterio)chlorophyll diversity

Evolution produced (bacterio)chlorophylls that allow the absorption of light. Because light quality and quantity can differ substantially depending on the time of day or the habitat, a broad range of (bacterio)chlorophylls have evolved, capable of harvesting light in different regions of the solar spectrum. In many organisms, multiple types of (bacterio)chlorophyll are present, hence allowing photon-harvesting under varying light conditions [165]. Nearly 100 different chlorophyll structures have been reported to date, that, due to their many double bonds, assure absorption form the near UV (~350 nm) to the near-infrared (~1050 nm) regions (Table 2) [159, 160, 166-168]. All these structures, however, are generally grouped into 16 classes of (bacterio)chlorophyll, although several other hypothetical classes – not yet detected in nature – have also been proposed (*e.g.* Bchl *f*) (Table 2) [160, 165, 169]. Depending on the organism, one or more types of (bacterio)chlorophyll are present (Table 2), with Chl *a* and Bchl *a* being the most common and abundant.

Table 2 Distribution and basic spectroscopic properties of natural (bacterio)chlorophylls in Bacteria

Pigment / Phylum	Acidobacteria	Chlorobi	Chloroflexi	Firmicutes	Gemmatimonadetes	Proteobacteria	Cyanobacteria	Absorption *in situ* (λ_{max})
Chl *a*		+					+	~440, 670 - 720
8-vinyl Chl *a*							+	~450, 670 - 720
8-hydroxy Chl *a*				+				~450, 670 - 720
Chl *b*							+	~460, 630 - 680
8-vinyl Chl *b*							+	450 - 480, 630 - 680
Chl *c*[1]							+	~400, 500 - 620
Chl *d*							+	~440, ~690
Chl *f*							+	~440, 720 - 740
Bchl *a*	+	+	+		+	+		<400, ~600, 800 - 900
Bchl *b*						+		<400, ~600, 800 - 1020
Bchl *c*	+	+	+					~460, 730 - 760
Bchl *d*		+	+					~440, 720 - 750
Bchl *e*		+	+					~460, ~715
Bchl *g*				+				<400, 780 - 850

[1] Chl *c* consists of three related types (Chl c_1, Chl c_2 and Chl c_3)

1.4.2.2.4. Photochemical reaction centers

(Bacterio)chlorophylls are the defining feature for the charge-separating reaction centers (RCs), the heart of (B)Chl-dependent phototrophy. RCs, however, have a low pigment density leading to little light absorption. In addition, only Chl *a*, Chl *b*, Bchl *a*, Bchl *b* and Bchl *g* are found in the photochemical reaction centers, greatly reducing the range of light absorbance. Therefore, in most RC-containing organisms, the reaction centers are surrounded by light-harvesting antenna complexes that typically contain hundreds of (bacterio)chlorophylls of the same or different classes as those present in the RC, thus increasing the absorption cross-section of the RCs and expanding the range of wavelengths utilizable for phototrophy. At the low light intensities found in nature, these energy-concentrating arrangements allow RCs to collect light energy that would otherwise not be available [160, 165, 166].

The photochemical reaction centers, however, are where the magic happens, *i.e.* the conversion of light energy to chemical energy. They are pigment-protein complexes that are the site of the early electron transfer reactions. In all currently known (B)Chl-dependent phototrophic organisms, the overall organization of the pigments in the RCs is similar.
Over the years, analyses segregated all RCs into two types, called type 1 and type 2 (Figure 7), according to the chemical identity of their immediate secondary electron donors and acceptors. Although multiple differences exist, structural studies strongly suggest that both descended from the same ancestral reaction center (Figure 7) [143, 170, 171]. An overview of the distribution of reaction centers in photosynthetic bacteria is shown in Table 3.

Type 1 RCs. In anoxygenic phototrophic bacteria, type 1 reaction centers form a homodimeric structure containing a core complex which is composed of a single protein, PshA, or PscA, in Firmicutes, or Chlorobi and Acidobacteria, respectively [148, 172, 173]. Type 1 RCs use the energy of an absorbed photon to take an electron from a donor and pass it to an iron-sulfur (FeS) acceptor, hence reducing the latter. Generally, the donor and acceptor are a cytochrome and a ferrodoxin, respectively. Subsequently, the RC drives a linear electron transfer from a second, inorganic donor to the reduced ferrodoxin. Finally, this ferrodoxin will supply its additional electron to drive one or more coupled assimilatory reactions. In oxygenic Cyanobacteria, however, type 1 RCs are part of a bigger complex called photosystem I (Figure 7), which has a heterodimeric structure composed of two proteins: PsaA and PsaB [171]. Here, in contrast to the APB, electrons are, also via ferrodoxin, supplied to $NADP^+$ and H^+, giving NADPH [144].

Type 2 RCs. In contrast to the type 1 RC, type 2 reaction centers found in APB all show a heterodimeric structure, with *pufL* and *pufM* encoding the conserved proteins, hence being convenient markers to study the diversity of these bacteria [146, 174, 175]. All type 2 RCs have pheophytin-quinone complexes as secondary electron acceptors. The secondary donor is a cytochrome, which is re-reduced with electrons from the proton-translocating cytochrome b-c_1 complex, itself reduced by the reduced pheophytin-quinone. As a result, there is no formation of NADPH. Although the overall pathway is cyclic, in oxygenic photosynthetic bacteria, it is predominantly non-cyclic. Similar to the type 1 RCs, type 2 RCs are part of a larger, heterodimeric complex called photosystem II (Figure 7), where D1 and D2 are the core proteins containing all the electron transport cofactors. Typically, the cyanobacterial photosystem II is associated with phycobilisomes which function as light-harvesting antannae. These phycobilisomes are aggregates of different types of phycobiliproteins (*i.e.* phycoerythrin, phycocyanin, phycoerythrocyanin and allophycocyanin) capable of absorbing light at different wavelengths, hence allowing Cyanobacteria to collect otherwise unavailable light [137]. Furthermore, where APB use either a type 1 or 2 reaction center, Cyanobacteria link the two types (*i.e.* the eventual electron acceptor of type 2 RC is the type 1 RC) to perform photosynthesis (Figure 7) [144, 171, 176].

Table 3 Distribution of photochemical reaction centers in phototrophic bacteria

Reaction center / Phylum	Acidobacteria	Chlorobi	Chloroflexi	Firmicutes	Gemmatimonadetes	Proteobacteria	Cyanobacteria
Type 1	+	+		+			+
Type 2			+	+[1]	+	+	+

[1] Putative phototroph *Alkalibacterium* sp. NP13 [177]

1.4.2.2.5. Distribution of photochemical reaction centers

Based on our current knowledge, bacteria harboring only type 1 reaction centers are significantly less abundant compared to the type 2. Only a handful of bacteria are known to possess a type 1 RC. A possible explanation is that organisms dependent only on a type 1 RC are in most cases also obligate anaerobes and thus, are hiding, as it were, from oxygen [144]. In addition, these RCs are found in bacterial phyla of which very little is known (*e.g.*

Acidobacteria), and the genes coding for the core proteins are different in the different type 1-dependent taxa. Thus, perhaps they are a lot more abundant and widespread, but have not yet been found due to the difficulties to sample their habitats and the lack of data to design primers targeting the genes encoding subunits of their reaction centers.

Type 2 reaction centers, however, are very widespread, especially among Proteobacteria. Originally it was believed all APB were anaerobic species, since Bchl had only been reported in species grown under anaerobic or semiaerobic conditions. In 1978, however, the first AAP was discovered in a seawater sample taken near Tokyo, Japan [150]. Subsequent to the isolation of this bacterium, described as *Erythrobacter longus*, many new obligate aerobic strains containing Bchl as the main light-harvesting pigment have been described from various aquatic habitats [146]. During the last couple of decades, AAP research grew and both culture-dependent, and culture-independent studies using the *pufM* gene as a marker, revealed their importance in the environment. The majority of studies, however, focused on aquatic habitats, ranging from marine to freshwater ecosystems all over our planet, where it was shown that AAP bacteria contribute to the recycling of organic carbon and other biogenic elements [146, 178]. Over the years, however, terrestrial ecosystems received little to no attention.

1.4.3. Rhodopsins

Organisms using photoreceptor proteins to sense and respond to light are found in all domains of life. One of the protein families involved, rhodopsins, also known as retinal proteins, consist of an opsin apoprotein and a covalently linked retinal chromophore. Little however is known about its origin in time.

All rhodopsins have a similar structure where the opsin consist of a seven-transmembrane α-helical architecture with the N- and C- terminals facing out- and inside the cell, respectively, whereas the retinal is connected to a cavity inside the protein with a Schiff base[7] linkage. Two types of rhodopsins, sharing practically no sequence similarity, can be distinguished. Type II rhodopsin is widely distributed in non-microbial Eukaryotes and has the sole purpose to initiate intra- or intercellular signaling. Type I rhodopsin, on the other hand, is widely distributed in Bacteria, Archaea and viruses, hence its name microbial rhodopsin. In addition

[7] A Schiff base is a aldehyde- or ketone-like compound (structure RR'C=NR") in which the carbonyl group is replaced by an imine or azomethine group.

to initiating intra- or intercellular signaling, it is also employed to absorb photons for energy conversion [179-181].

1.4.3.1. Microbial rhodopsins

Since their discovery in 1971, microbial rhodopsins were epitomized by haloarchaeal proteins (bacteriorhodopsin, halorhodopsin and sensory rhodopsin I and II) [179, 182]. With the start of the new millennium also came the discovery of related photoactive proteins with similar or different functions. Since then, thousands of rhodopsin gene sequences, mostly from environmental surveys, have been found in Archaea, viruses and Eubacteria, hence changing our concept of a small group of proteins – having a limited number of functions and present only in haloarchaea – towards that of an overwhelming variety of protein species with versatile functions, ecology, taxonomy and sequences that may even represent the most abundant phototrophic system on Earth [179, 183, 184].

1.4.3.1.1. Ion-pumping rhodopsins

Anno 2017, characterized rhodopsin proteins have been classified into three distinct groups, based on their functional role: light sensors, cation channels and light-driven ion pumps [185]. The latter type, however, has the widest ecological niche distribution and is found throughout Archaea and Bacteria. Based on the pumping direction, it can additionally be subdivided into two clades (*i.e.* inward- and outward-directed pumps), each containing multiple rhodopsin types that function optimally when absorbing light of one or more wavelengths. The outward-directed pumps, however, show most diversification and are represented by the proteorhodopsins, named after the initial identification in Gammaproteobacteria [181, 183, 184].

1.4.3.1.1.1. Proteorhodopsin

Since the discovery of proteorhodopsin in a marine metagenome from Monterey Bay in California in 2000 [186], it soon became clear that this rhodopsin family is by far the most abundant and widespread of all microbial rhodopsins [183]. Fueled by the absorption of photons, proteorhodopsin pumps protons to the outside of the bacterial membrane. Afterwards, the membrane potential resulting from this proton translocation is used by the cell to generate

ATP [186]. This large protein family is comprised of blue- and green-absorbing proteorhodopsins [183]. Whereas the blue type is spectrally tuned to absorb the most dominant wavelength of light in open ocean and deep sea waters (λ_{max} ~490 nm), the green type (λ_{max} ~525 nm) is characteristic for bacteria living near the surface [179]. None of the cultured proteorhodopsin-containing bacteria have been found to contain carbon fixation pathways [187, 188]. Thus, the benefit of this mechanisms seem to be that light may enhance ATP production by heterotrophic bacteria that otherwise rely on organic matter oxidation for energy [189].

Because bacteria carrying proteorhodopsin genes have been found to account for a substantial fraction of all oceanic planktonic bacteria, it probably contributes significantly to the input, in the biosphere, of solar energy. Although most of the currently available sequences originate from marine environments (~93 %) [190], they are, however, not confined to these systems. Approximately 6 % of proteorhodopsin sequences have been obtained from freshwater habitats [191, 192] and Antarctic sea ice [193]. The remaining fraction of proteorhodopsins originates from Siberian permafrost [194], Andean hot springs [195] and leaf surfaces of terrestrial plants [196]. Over the years, many experiments have been conducted to link the presence of proteorhodopsin to light-stimulated growth. Although in some cases it was concluded that proteorhodopsin expression can give significant advantages to the host under stress conditions, this is, however, not the general case [179, 183].

1.4.3.1.1.2. Actinorhodopsin

In 2008, additional analyses on metagenome data from the Global Ocean Survey [197, 198] revealed another novel microbial rhodopsin type, originally identified as proteorhodopsin and found in a variety of non-marine sites. The sequences exhibited a nearly exclusive linkage to open reading frames displaying a close relation to homologues in actinobacterial genomes [199]. Combined with the high abundance of actinobacterial SSU rRNA sequences in non-marine samples [197] and the knowledge that Actinobacteria are common inhabitants of freshwater ecosystems [200], the novel rhodopsins were designated actinorhodopsins [199].

Because they encode the conserved amino acid residues required for proton pumping, these actinorhodopsins hinted at a possible ion transport function [199]. Over the years, actinorhodopsins have been retrieved from the genomes of cultivated Actinobacteria and a variety of freshwater habitats as well as marine ecosystems [201, 202]. Similar to proteorhodopsin, however, their presence and distribution in terrestrial environments has

received little attention. The broad taxonomic and ecological distribution of proteo- and actinorhodopsins (and possibly also other currently unknown types) in diverse aquatic environments suggests they are potentially very important in the adaptation of their hosts to life on Earth's surface [191, 197, 199, 201].

1.4.4. Rhodopsin vs. (bacterio)chlorophyll: why less is not always more

In the end, however, the question remains why phototrophic niches are dominated by only a few groups of phototrophs containing (bacterio)chlorophyll-dependent reaction centers, while rhodopsin-based phototrophy seemingly has a global distribution and, more importantly, can easily be transferred among organisms. Part of the answer lies in the inefficient use of light by rhodopsin-containing microorganisms, as twice as much photons have to be absorbed to produce the same amount of ATP (bacterio)chlorophyll-dependent phototrophs do. Furthermore, the latter group can have hundreds to thousands of chromophores, while rhodopsin-dependent phototrophs mostly only contain one. As a result, microorganisms containing rhodopsins would have to synthesize a lot more rhodopsin molecules to absorb similar amounts of light energy. The high efficiency of (bacterio)chlorophyll-based phototrophy compared to that of rhodopsin-based phototrophy, however, raises a second question: why is it not more widespread and spread laterally? The answer here probably lies in the high number of genes required for (bacterio)chlorophyll-dependent photosystems that would have to be transferred. Since rhodopsin-based phototrophy only requires a fraction of the genes (bacterio)chlorophyll-based phototrophy does, it might explain its broader distribution. In conclusion, the globally widespread and abundant rhodopsin-based phototrophy is the perfect example why simplicity is not efficiency and why less is not always more [137, 203, 204].

1.4.5. Light-harvesting in terrestrial Antarctica

Despite an increasing number of descriptions of light-harvesting bacteria in sampling sites throughout the world, polar regions remain a virtually unexplored terrain. Furthermore, even though terrestrial areas account for ~30 % of our planet, only a handful of studies has been conducted towards unraveling the light-harvesting potential of bacterial communities inhabiting these regions. The main contribution on the Arctic region was made by Feng et al.

who provided baseline information about the distribution and abundance of anoxygenic phototrophic bacteria in European, Alaskan and Canadian Arctic soils [205]. Information about these biogeochemically important bacteria in Antarctica, however, is far rarer. Anoxygenic phototrophic bacteria have been reported from lakes and ponds throughout the continent [206], but never in Antarctic soils. Proteo- and actinorhodopsin-containing microbes, on the other hand, have, to our knowledge, never been detected in polar terrestrial regions (based on metagenome data available on MG-RAST [207] and IMG [208]). With global climate change affecting the ecology of animals and plants throughout the polar regions [26, 28, 209-211], it goes without saying that warming and other changes may also severely affect microbial communities, which are often the major contributors to total ecosystem biomass, nutrient cycling, biodiversity and energy flow. Therefore, understanding the patterns of (functional) bacterial diversity will provide essential information to characterize the impact of a changing climate.

1.5. Objectives and thesis outline

For many years it has been thought that Cyanobacteria are the main primary producers in Antarctic soils, providing not only themselves, but also other members of the community with carbon and nitrogen. During the last decade, the growing number of studies investigating Antarctica's microbial communities revealed the near absence of this phylum in several vegetated and fell-field sites on the Peninsula and in soils of the McMurdo Dry Valleys and the Sør Rondane Mountains. Little, however, is known about other bacterial life that may potentially perform carbon and/or nitrogen fixation in the absence of Cyanobacteria.

In case of nutrient limitations, the mainly microscopic life in these areas may be expected to use alternative energy sources. Certain microorganisms, phototrophs, can rely on light-harvesting mechanisms. These rhodopsin- and (bacterio)chlorophyll-dependent pathways may enhance ATP production in their host. In turn, this ATP may fuel assimilatory reactions. In this respect, the abundant sunlight during the austral summer may be a useful resource for the phototrophic microorganisms inhabiting these extreme oligotrophic environments.

Given the sometimes low abundance of Cyanobacteria in ice-free Antarctic areas, our focus was on non-cyanobacterial taxa.The main objectives of this thesis were (i) to investigate the diversity and distribution of autotrophic, diazotrophic and phototrophic bacteria in terrestrial Antarctic samples by sequencing key protein encoding genes involved in these processes, (ii) to provide new cultured representatives of carbon fixing, nitrogen fixing and light-harvesting prokaryotes, (iii) compare the diversity retrieved using culture-dependent and -independent approaches and (iv) to test the potential contribution of auto-, diazo- and phototrophic mechanisms for growth in laboratory experiments.

In a first study (**Chapter 2**), the autotrophic, diazotrophic and phototrophic bacterial community composition of four terrestrial samples, taken in the proximity of the Belgian Princess Elisabeth Station, was explored. We used a clone library approach, targeting *cbbL* and *cbbM* genes, to assess the autotrophic community dependent on the Calvin-Benson-Bassham cycle. To study the diazotrophic guild community, *nifH* genes, encoding a subunit of the most widespread Mo-dependent nitrogenase, were targeted. Finally, to obtain a better insight in non-cyanobacterial phototrophs, we examined the diversity of key protein encoding genes involved in the two major bacterial light-harvesting processes: rhodopsin-based

phototrophy (proteorhodopsin genes) and (bacterio)chlorophyll-based phototrophy (*pufLM* genes).

While the clone libraries provided a first view on the diversity of these important genes, they were, however, limited to ~100 clones per sample per gene. Furthermore, rarefaction analyses indicated that the diversity was much larger. Therefore, in a second and third study (**Chapters 3 and 4**), the total bacterial community was examined, and the diversity of the aforementioned mechanisms was further unraveled, using an Illumina MiSeq sequencing approach. We also compared the results with those obtained in the clone library study.

Although diversity of functions can be probed using metagenomics approaches, linking to organisms remains problematic. Thus, isolation and characterization of microorganisms is of great scientific relevance for investigating ecophysiology and adaptive strategies, and linking function to identity. Cultured microorganisms also permit testing of certain functions. In addition, isolates may help extend identifications obtained from metagenome data. Therefore, in **Chapter 5**, a large-scale isolation campaign was performed to cultivate and characterize the aerobic anoxygenic phototrophic bacterial communities from the same four terrestrial samples. Potential phototrophic strains were tested under laboratory conditions to investigate the influence and contribution of light on growth. In addition, near complete 16S rRNA gene data from newly obtained strains was compared with partial sequences obtained from the same samples using Illumina MiSeq sequencing to allow more insight in the overlap between culture-dependent and –independent approaches.

In addition to potential aerobic anoxygenic phototrophic bacteria, the isolation campaign gave access to a broad range of bacteria, including first cultured representatives of the candidate phylum FBP. In **Chapter 6**, the genomic landscape, morphology and physiology of one of these strains (R-68213^{T}) is described. The draft genome was obtained using Illumina HiSeq sequencing. On the basis of various analyses, strain R-68213^{T} represents a novel species and genus affiliated with a phylum-level lineage previously recognized as candidate phylum FBP. Formal descriptions of the proposed novel species (*Abditibacterium utsteinense*), phylum (Abditibacteria) and other taxonomic ranks are included in **Chapter 6**.

Chapter 2

Analysis of *cbbL*, *nifH* and *pufLM* in soils from the Sør Rondane Mountains, Antarctica, reveals a large diversity of auto- and phototrophic bacteria

Redrafted from:

Tahon G., Tytgat B., Stragier P. and Willems A. (2016). Analysis of *cbbL*, *nifH*, and *pufLM* in Soils from the Sør Rondane Mountains, Antarctica, Reveals a Large Diversity of Autotrophic and Phototrophic Bacteria. Microbial Ecology 71(1): 131-149 doi: 10.1007/s00248-015-0704-6.

Author's contributions:

GT and AW designed the experiments. GT performed the experiments. GT, BT, PS and AW analyzed the data. BT and PS contributed analysis tools. GT, BT and AW wrote the manuscript.

Summary

Cyanobacteria are generally thought to be responsible for primary production and nitrogen fixation in the microbial communities that dominate Antarctic ecosystems. Recent studies of bacterial communities in terrestrial Antarctica, however, have shown that Cyanobacteria are sometimes only scarcely present. Therefore we studied other bacteria that presumably take over their role as primary producers and diazotrophs. We also investigated the occurrence of proteorhodopsin and anoxygenic phototrophy as mechanisms for non-Cyanobacteria to exploit solar energy. The diversity of key genes in these processes was studied in surface samples from the Sør Rondane Mountains, Dronning Maud Land, using clone libraries of the large subunit of ribulose-1,5-biphosphate carboxylase/oxygenase (RuBisCO) genes (*cbbL*, *cbbM*), dinitrogenase-reductase (*nifH*) genes, proteorhodopsin genes and genes coding for the light and medium subunits of the type 2 phototrophic reaction center (*pufLM*). We recovered a large diversity of non-cyanobacterial *cbbL* type IC in addition to cyanobacterial type IB, suggesting that non-cyanobacterial autotrophs may contribute to primary production. The *nifH* diversity recovered, was predominantly related to Cyanobacteria, particularly members of the *Nostocales*. While proteorhodopsin genes were not detected, a large diversity of *pufLM* was observed, suggesting for the first time, that the aerobic photoheterotrophic lifestyle may be important in oligotrophic high altitude ice-free terrestrial Antarctic habitats.

2.1. Introduction

The Antarctic continent is dominated by extreme environmental conditions, including limited organic nutrients, low humidity, frequent freeze-thaw and wet-dry cycles, rapid drainage, low and transient precipitation and low thermal capacity of the substratum [16]. Ice-free regions are limited to a minute proportion (~0.32%) of the continent's surface area [4]. The Sør Rondane Mountains, located in Dronning Maud Land, East-Antarctica and home of the Belgian Princess Elisabeth Station (PES) contain an important number of ice-free mountain tops and nunataks [212].

Antarctica has a very limited fauna and flora, with just two native species of flowering plants, occurring only on the Peninsula [24, 213]. Surface substrates near the PES consist mostly of weathered rocks with relatively little organic material, given the absence of vascular plants. It is assumed that Cyanobacteria, with their ability to fix carbon dioxide and nitrogen, are the main primary producers in this barren region [128, 214]. Recent studies however, have shown that Cyanobacteria are not always highly abundant in vegetated and fell-field sites on the Peninsula and in mineral soils of the McMurdo Dry Valleys [215-217] and in such cases, other bacteria can be assumed to take over their role in carbon and/or nitrogen fixation.

Of several mechanisms for carbon dioxide fixation, the Calvin-Benson-Bassham (CBB) cycle is considered to be the most important autotrophic pathway [218]. The enzyme responsible for the actual CO_2 fixation in the CBB cycle is ribulose-1,5-biphosphate carboxylase/oxygenase [75]. Four different natural types of RuBisCO are known: Type I (*cbbL* gene) is subdivided in to a green and a red subgroup, each containing two subclasses [218]. In the green subgroup, *cbbL* IA is predominantly found in Alpha-, Beta- and Gammaproteobacteria, while type IB is found in Cyanobacteria, green algae and green plants [76]. The red group contains type IC, found in Alpha-, Beta- and Gammaproteobacteria [72, 76], but recently also in Verrucomicrobia [82], Firmicutes [83] and Actinobacteria [79, 80], and type ID, found in non-green algae [72, 76]. Some authors have suggested that the type I found in Verrucomicrobia and Actinobacteria may form a separate subtype in the red subgroup, type IE *cbbL* [79, 80, 82]. Type II (*cbbM* gene) is found in various Proteobacteria (purple non-sulfur bacteria), aerobic and facultative anaerobic chemoautotrophic bacteria and dinoflagellates [70, 84]. RuBisCO type III has only been found in Archaea [87]. Finally, type IV RuBisCO is designated as RuBisCO-like [72] and is considered not to be involved in the Calvin cycle [88].

Besides fixation of carbon dioxide, biological nitrogen fixation, a multistep process catalyzed by the nitrogenase enzyme complex [219], is also very important in Antarctic soils [133, 220]. The nitrogenase complex is composed of two metalloproteins: dinitrogenase and dinitrogenase-reductase. The latter, encoded by *nifH*, is very useful in phylogenetic studies because of its conserved nature [221] and high level of congruence with 16S rRNA gene phylogeny [222].

Both nitrogen and carbon dioxide fixation require a substantial amount of ATP. In the poor soils of the ice-free zones in continental Antarctica, sunlight – an abundant energy source during the Antarctic summer – may therefore represent an important resource, also for non-Cyanobacteria. Over the past decades, several alternative light-harvesting mechanisms, including microbial rhodopsins and aerobic anoxygenic phototrophy, have been discovered. By far the most abundant family of microbial rhodopsins, proteorhodpsins (PR) is found in Bacteria where these membrane embedded light-driven proton pumps contribute to proton motive force [181, 223]. They have been reported from diverse marine and non-marine habitats [183]. To our knowledge, their presence in Antarctica, however, has not been reported.

The aerobic anoxygenic phototrophic bacteria (AAP) are photoheterotrophs that can use bacteriochlorophyll *a* to harvest light energy. Light-mediated CO_2 incorporation also occurs, although this activity is rather low [151] and remains controversial [224]. Diversity of AAP has been examined by screening the *pufL* and *pufM* genes, which encode the light and medium subunit of the type 2 photoreaction center [225]. Until now, AAP have been predominantly documented in the Alpha-, Beta- and Gammaproteobacteria (purple bacteria) [226, 227], although the light-harvesting complex is also present in anaerobic green non-sulfur bacteria, heliobacteria and other Firmicutes [177, 228]. Previous studies have revealed presence of AAP in various aquatic habitats, ranging from the world's oceans [174, 229-232] to inland lakes [226, 227, 233, 234] and even Antarctic sea ice [175]. Until today, little is known about the presence and diversity of AAP in terrestrial environments, including continental Antarctica.

We hypothesized that in ice-free areas of terrestrial Antarctica, in addition to Cyanobacteria, other prokaryotic primary producers, and bacteria that exploit solar energy, may contribute to carbon and/or nitrogen fixation. Our area of study was the Sør Rondane Mountains, in the vicinity of the PES, where a previous pyrosequencing study indicated scarcity of

Cyanobacteria in some of the samples of terrestrial microbial mats [54]. The aim of this study was to investigate terrestrial samples for the presence and diversity of primary producers and other bacteria exploiting sunlight, by targeting *cbbL* and *cbbM* genes (RuBisCO), *nifH* (nitrogen fixation), *pufLM* (phototrophy with type 2 reaction centers, including AAP) and proteorhodopsin genes. This work represents the first study of these mechanisms in oligotrophic high altitude soils of the Sør Rondane Mountains and is the first report of the possible presence of AAP in terrestrial Antarctic samples.

2.2. Materials and Methods

Sample collection and environmental data

Surface samples consisting of weathered granite parent material (depth of 0 to 5 cm) were taken aseptically during the Antarctic summer of 2009 near the Princess Elisabeth Station (71°57'S, 23°20'E) at Utsteinen, Sør Rondane Mountains, East Antarctica. Upon collection, samples were frozen at -20°C until processing. Four samples were used in this study: KP15, KP43 and KP53 were collected on the Utsteinen ridge in the immediate vicinity of the PES, KP2 was collected on the north face of the nunatak 1.3 km south of the research station (Figure 1). Sample description, coordinates, altitude, geochemical parameters (pH, water content, total organic carbon (TOC) and conductivity) and relative abundance of Cyanobacteria according to Tahon et al. (unpublished data, see Chapter 3) are listed in Table 1.

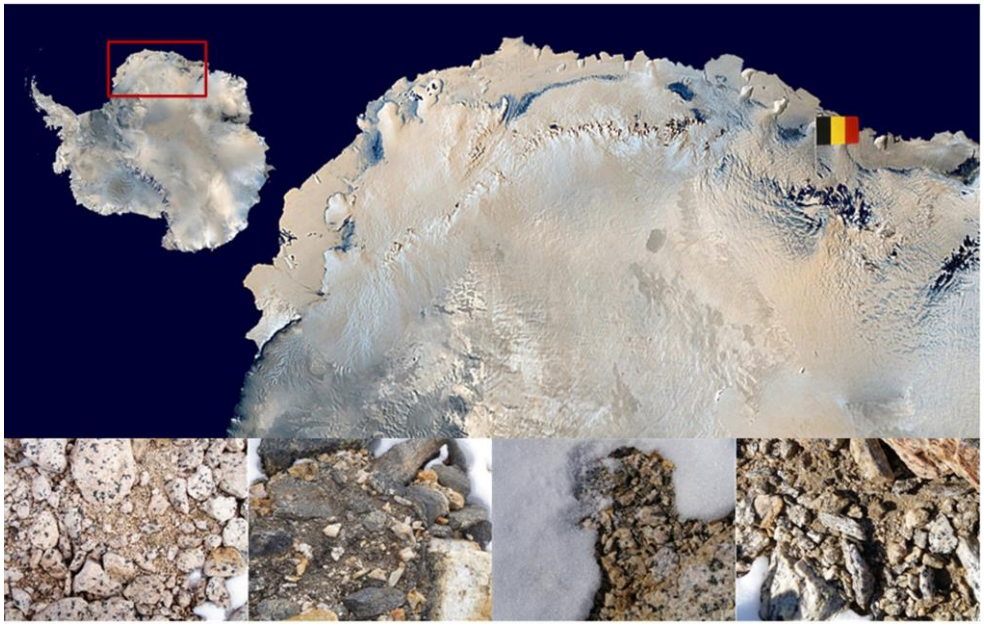

Figure 1 Location of the Princess Elisabeth Station (*base of flag post*) in the Sør Rondane Mountains, East-Antarctica and, from *left to right*, photographs of the terrestrial samples used in this study (KP2, KP15, KP43 and KP53)

Isolation of DNA

Seven DNA extraction methods were tested and their performance to yield maximum diversity of bacterial DNA from Antarctic terrestrial surface samples was verified by denaturing gradient gel electrophoresis of partial 16S rRNA and partial *pufM* genes. The

Table 1 Parameters associated with analyzed samples

Sample	Sample Coordinates	Altitude (m)	Description of sample area	Conductivity (µS/cm)	pH	Water Content	TOC	Relative abundance Cyanobacteria[a]
KP2	71° 57' 28.6" S, 23° 19' 45.8" E	1320	Small gravel particles in between the rocks, moraine slope west	19	6.54	6.28 %	0.08 %	19.62 %
KP15	71° 56' 45.8" S, 23° 20' 43.6" E	1366	Brown soil under lichen	33	5.57	3.38 %	0.33 %	51.85 %
KP43	71° 56’ 47.3” S, 23° 20’ 44.6” E	1362	Brown soil with dark green fragments	520	6.22	0.91 %	2.57 %	0.04 %
KP53	71° 56’ 45.3” S, 23° 20’ 42.4” E	1362	Grey soil on second part of the ridge	312	6.34	0.23 %	0.21 %	0.12 %

[a] Data based on Illumina MiSeq 2 x 300 bp paired-end sequencing of partial 16S rRNA genes (unpublished data, see Chapter 3)

PowerLyzer® PowerSoil® DNA isolation kit (MoBio Laboratories) performed best (Figure S1) and was therefore used for analyzing the samples. Total genomic DNA was extracted in triplicate from 400 mg of homogenized sample, following the manufacturer's protocol. The lysis step was performed using an alternative protocol separately provided by the manufacturer. Briefly, 500 µl phenol:chloroform:isoamylalcohol (25:24:1), 500 µl Bead Solution and 60 µl of C2 solution were added to a PowerLyzer™ Glass Bead Tube together with 400 mg of sample. Tubes were bead beaten for 10 minutes (30.0 Hz) and the supernatant was transferred to a clean 2 ml collection tube. A volume of 150 µl of component C3 was added and tubes were cooled for 5 minutes at 4°C, after which the original protocol was continued. Extracted DNA was quantified with a Qubit® 2.0 fluorometer (Life Technologies) and stored at -20°C until processing.

PCR

For every DNA extract, a PCR was performed in triplicate in a total volume of 25 µl containing 0.2 mM of each deoxynucleotide triphosphate (dNTP), 1X Qiagen PCR buffer (Qiagen), 0.625 units of Qiagen *Taq* DNA polymerase (Qiagen), 100 µM bovine serum albumin (BSA), 3 µl of template solution and a forward and reverse primer (Table 2). Amplification was performed using a Veriti thermal cycler (Life Technologies). The temperature profiles of all PCRs are shown in Table 2. For each sample and primer set, all nine PCR products (3 replicate DNA extracts x 3 replicate PCRs), showing bands of the expected size, were pooled and purified using a Nucleofast 96 PCR clean up membrane system (Macherey-Nagel) and Tecan Genesis Workstation 200 (Tecan).

Clone library construction and sequencing

Purified PCR products were cloned with a pGEM®-T Vector System II (Promega) following the manufacturer's instructions, in duplicate or triplicate. Competent *Escherichia coli* JM109 cells were transformed with the ligation product and screened using blue/white coloration. For each sample, 150 white transformants of each PCR type were purified by streaking. To release plasmid DNA, cells were suspended in 15 µl of MilliQ water and lysed by heating to 100°C (10 min). Inserts were amplified using the T7/SP6 primer set (Promega) in a 25 µl reaction mixture containing 1 µl of DNA solution, 1X Qiagen PCR buffer (Qiagen), 0.2 mM

Table 2 PCR Primers and conditions used for screening different genes

Gene	Target	Primer	Sequence 5'-3'	Final Concentration	Region	Amplicon size	Program
cbbL	RuBisCO IA & IB	RubIgF[a]	GAY TTC ACC AAR GAY GAY GA	0.4 µM	571-1382[i]	± 800 bp	95°C (3min); 3x 95°C (1min), 49°C (2min 15sec), 72°C (2min 15sec); 30x 95°C (35sec), 49°C (1min 15sec), 72°C (1min 15sec); 72°C (7min)
		RubIgR[a]	TCR AAC TTG ATY TCY TTC CA	0.4 µM			
cbbL	RuBisCO IA & IC	K2f[b]	ACC AYC AAG CCS AAG CTS GG	0.2 µM	496-990[i]	492-495 bp	95°C (3min); 35x 95°C (1min), 62°C (1min), 72°C (1min 30sec); 72°C (10min) [235]
		V2r[b]	GCC TTC SAG CTT GCC SAC CRC	0.2 µM			
cbbM	RuBisCO II	cbbM343F[c]	GGY AAY AAC CAR GGY ATG GG	0.1 µM	343-1126[c]	700-800 bp	95°C (3min); 30x 95°C (1min), 50°C (2min), 72°C (3min); 72°C (7min) [59]
		cbbM1126R[c]	CGY ARB GCR TTC ATR CCR CC	0.1 µM			
pufLM	AAP	pufLF[d]	CTK TTC GAC TTC TGG GTS GG	0.2 µM	64-1612[j]	± 1500 bp	94°C (3min); 30x 94°C (1min), 60°C (1min), 72°C (2min); 72°C (10min) [236]
		pufMR[e]	CCA TSG TCC AGC GCC AGA A	0.2 µM			
PR	*Flavo-bacteria*	PR-Flavo-F[f]	GAY TAY GTW GSW TTY ACD TTY TTT GTR GG	0.2 µM	58-572[k]	460 bp	94°C (3min); 15x 94°C (30sec), 60°C (45sec) (-0.5°C/cycle), 72°C (30sec) ; 20x 94°C (30sec), 52°C (45sec), 72°C (30sec) ; 72°C (10min) [237]
		PR-Flavo-R[f]	GCC CAW CCH ACW ARW ACR AAC CAR CAT A	0.2 µM			
PR	*Flavo-bacterium*	PR-Flavo-2F[f]	GGC TAT GAT GGC HGC WK	0.2 µM	93-586[k]	460 bp	
		PR-Flavo-2R[f]	CWA DWG GRT ARA TNG CCC A	0.2 µM			
PR	Universal	PR-1aF[g]	GAT CGA GCG NTA YRT HGA RTG G	1.87 µM	340-665[l]	± 335 bp	94°C (2min), 30x 94°C (30sec), 52°C (30sec), 72°C (30sec) ; 72°C (7min) [189]
		PR-1aR[g]	GAT CGA GCR TAD ATN GCC CAN CC	1.87 µM			
nifH	Universal	PolF[h]	TGC GAY CCS AAR GCB GAC TC	0.3 µM	115-476[h]	360 bp	95°C (5min); 30x 94°C (30sec), 53°C (1min), 72°C (40sec); 72°C (5min) [238]
		PolR[h]	ATS GCC ATC ATY TCR CCG GA	0.3 µM			

[a] data from [239], [b] data from [240], [c] data from [59], [d] data from [241], [e] data from [242], [f] data from [237], [g] data from [189], [h] data from [243], [i] Based on the *cbbL* IA sequence of *Bradyrhizobium* sp. ORS278 (CU234118), [j] Based on the *pufLM* sequence of *Sphingomonas sanxanigenens* DSM 19645 (CP006644), [k] Based on the PR sequence of *Dokdonia* sp. PRO 95 (FJ627053), [l] Based on the PR sequence of *Vibrio campbellii* BAA-1116 (FJ985782)

of each dNTP, 0.1 µM of each primer, 0.3125 units of Qiagen *Taq* polymerase (Qiagen) and 0.5 µl $MgCl_2$ (25 mM). Amplification was carried out as follows: 95°C (5 min), 3 cycles of 95°C (1 min), 49°C (2 min 15 sec), 72°C (2 min 15 sec) and 30 cycles of 95°C (35 sec), 49°C (1 min 15 sec), 72°C (1 min 15 sec). A final extension at 72°C (7 min) and subsequent cooling at 4°C completed the reaction. PCR products of the expected size were purified as described above and gene fragments were sequenced with primers T7 and SP6 and the original amplification primers (Table 2) using a BigDye Xterminator™ purification kit (Applied Biosystems) and an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems).

Analysis of clones

For each functional gene, a custom-made database was prepared in BioNumerics 7.5 (Applied Maths) by downloading all related sequence records from the NCBI and IMG (https://img.jgi.doe.gov/) [244] databases as available per December 1st 2014. Redundant sequences were not removed not to lose metadata information on the habitats reference sequences originated from. Records were manually checked to eliminate low-quality sequences (*i.e.* presence of stop codons, ambiguous bases, indels). Newly obtained sequences were added to the corresponding BioNumerics database using the Assembler module. For all clones, between 4 and 8 overlapping sequences were assembled. Vector sequences and amplification primer sequences were trimmed off and sequences were manually curated. For additional quality curation, nucleotide sequences were translated into amino acids using MEGA 6 [245]. Putative chimeric sequences (detected using the Uchime module in Mothur [246, 247]), sequences with no similarity to our genes and sequences that were too short (≥ 1 AA) or contained stop codons were removed from the dataset before further analysis. For rarefaction analysis, binning at both 95% and 100% nucleotide sequence identity and 100% protein sequence identity was performed using the "Fill field with cluster number" option in BioNumerics 7.5. For phylogenetic analyses of *cbbL*, *nifH* and *pufLM*, all unique protein sequences were included. They are referred to as operational RuBisCO units (ORU), operational nitrogenase units (ONU) and operational puf units (OPU). For each gene, a first amino acid alignment was made with all sequences present in our database (*nifH*: 46.371; *pufLM*: 3.706; *cbbL* type IA, IB, IC: 8.004), using Clustal Omega [248, 249]. Alignments were trimmed to the size of our cloned fragments and visually inspected, excluding all non-overlapping sequences or sequences that were too short (≥ 1 AA) from further analysis. These alignments were used to construct a maximum likelihood (ML) tree (1.000 bootstrap

replicates) with FastTree software [250] using the Whelan and Goldman evolutionary model and the discrete gamma model with 20 rate categories. Analysis of the *cbbL* IA/IB sequences, together with reference data, resulted in a final tree containing 2.103 protein sequences. For *cbbL* IA/IC, *nifH* and *pufLM*, these trees contained 3.610, 26.370 and 3.529 protein sequences respectively (data not shown). From the resulting phylogenetic trees, closest relatives of our newly obtained clone sequences as well as representative sequences from the entire tree were selected in order to prepare a smaller tree representing the initial complete tree, following the same protocol. Trees were visualized using the iTOL software [251, 252]. Clone sequences sharing the same nearest neighbor were grouped into clusters.

Statistical analyses

Statistical analyses of clone sequences were performed using the Vegan package [253] in R (https://cran.r-project.org). The total expected number of OTUs was determined by rarefaction analysis. For all clone libraries, different indices, based on derived protein sequences, were calculated, including evenness (Pielou), species richness (Chao1), and Bray-Curtis dissimilarity.

Accession numbers

The sequences determined in this study have been deposited in the National Center for Biotechnology GenBank database under the accession numbers KT154121 to KT154253 (IA/IB *cbbL* sequences), KT154254 to KT154456 (IA/IC *cbbL* sequences), KT154018 to KT154120 (*nifH* sequences) and KT154457 to KT154682 (*pufLM* sequences).

2.3. Results

Four terrestrial samples (Table 1), collected in the proximity of the Belgian Princess Elisabeth Station, were examined. To target different subtypes of the *cbbL* gene, coding for RuBisCO type I, two primer sets were used, RubIgF/R and K2f/V2r which robustly capture subtypes IA/IB and IA/IC, respectively (Figure S2, Table 2) [86, 87, 235, 239, 240, 254, 255]. A total of 353 clones of subtypes IA/IB and 444 of types IA/IC were obtained (Table 3). The *cbbM* gene (type II RuBisCO) failed to amplify from all four samples even though a recently designed primer set reported to target a broad diversity of the gene was used (Table 2) [59]. For *nifH* and *pufLM*, a total of 339 and 317 clones were obtained in this study (Table 3). Proteorhodopsin genes failed to amplify from all four samples.

When applying OTU binning at 95% DNA similarity, for all clone libraries rarefaction curves reached saturation (Figure 2), reflecting that at a 5% DNA divergence level the number of OTUs is relatively limited. Indeed, for *cbbL* subtypes IA/IB, *cbbL* subtypes IA/IC, *nifH* and *pufM* at 95% DNA sequence similarity 14, 43, 3 and 21 OTUs were recovered (Table 3). However, analysis of unique protein sequences or unique DNA sequences indicated that not all diversity was sampled as graphs were still rising (Figure 2), in line with overall coverage values of about 35 to 50% (Table 3).

For the *cbbL* gene the two primer sets used to amplify subtypes IA/IB and IA/IC target slightly different but overlapping regions of the gene that were therefore investigated separately (Figure S2, Table 2). For both primer sets, at least 98 high quality sequences were retained per sample, except for samples KP15 and KP53, where fewer sequences were obtained for one of the primer sets (Table 3). The coverage based on estimated species richness (Chao1) for the separate samples was only between 23.1% and 79.2% for IA/IB and between 27.1% and 74.7% for IA/IC, whereas the total coverage of the IA/IB and IA/IC libraries was only 35.4% and 40.7%. The overall evenness of the clone libraries was quite high in all the cases. Only for *cbbL* clone libraries of sample KP43, the value was lower. Finally, Bray-Curtis dissimilarity analysis revealed that, with the available data, very little overlap exists between the data from all terrestrial samples (Table 3, Figure 3).

Table 3 Overview of clone library characteristics. Diversity indices were calculated on the basis of derived unique protein sequences

			Clone Library				
			KP2	**KP15**	**KP43**	**KP53**	**All**
cbbL	No. of clones		115	39	123	76	353
(Primers for IA/IB)	No. of OTUs (100 % DNA)		44	17	43	30	133
	No. of OTUs (95 % DNA)		6	5	5	4	14
	No. of ORUs (100 % AA)		31	16	30	25	98
	evenness (H/H_{max})		0.652	0.896	0.478	0.708	0.728
	Chao1		52.86	20.20	130.00	82.00	276.75
	Coverage %		58.6 %	79.2 %	23.1 %	30.5 %	35.4 %
	Bray-Curtis	KP15	0.97				
		KP43	0.99	1.00			
		KP53	0.99	1.00	0.98		
cbbL	No. of clones		98	108	122	116	444
(Primers for IA/IC)	No. of OTUs (100 % DNA)		41	55	39	70	203
	No. of OTUs (95 % DNA)		14	14	4	16	43
	No. of ORUs (100 % AA)		31	44	31	56	158
	evenness (H/H_{max})		0.813	0.811	0.587	0.842	0.805
	Chao1		41.50	102.13	77.00	206.50	388.22
	Coverage %		74.7 %	43.1 %	40.3 %	27.1 %	40.7 %
	Bray-Curtis	KP15	0.96				
		KP43	1.00	1.00			
		KP53	0.99	0.96	0.87		
nifH	No. of clones		115	119	105	-	339
	No. of OTUs (100 % DNA)		33	31	41	-	103
	No. of OTUs (95 % DNA)		2	3	2	-	3
	No. of ONUs (100 % AA)		24	21	23	-	62
	evenness (H/H_{max})		0.415	0.392	0.459	-	0.466
	Chao1		87.33	47.25	50.20	-	50.48
	Coverage %		27.5 %	44.4 %	45.8 %	-	37.7 %
	Bray-Curtis	KP15	0.23				
		KP43	0.93	0.93			
pufM	No. of clones		-	98	98	121	317
	No. of OTUs (100 % DNA)		-	37	16	19	68
	No. of OTUs (95 % DNA)		-	16	4	6	21
	No. of OPUs (100 % AA)		-	29	8	17	49
	evenness (H/H_{max})		-	0.814	0.496	0.525	0.612
	Chao1		-	46.14	8.75	56.00	98.60
	Coverage %		-	62.8 %	91.4 %	30.4 %	49.7 %
	Bray-Curtis	KP43		0.92			
		KP53		0.91	0.57		

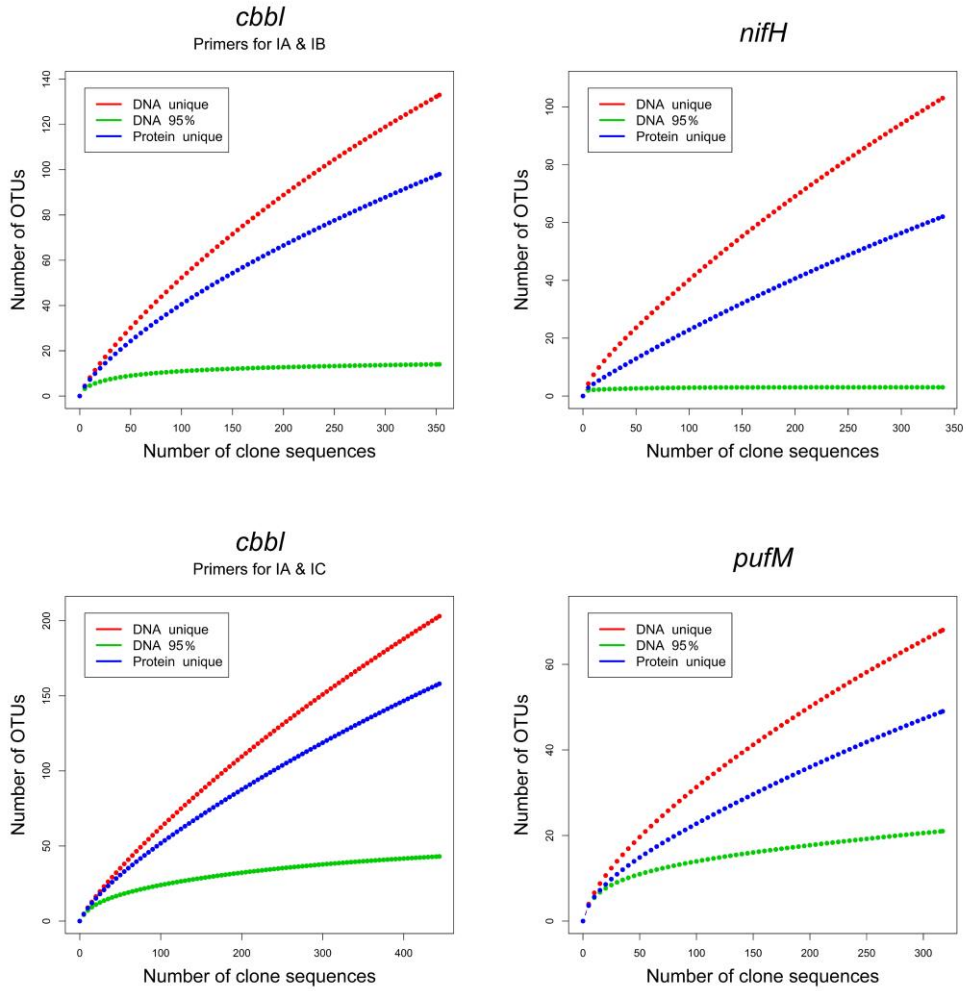

Figure 2 Rarefaction curves based on OTUs grouping clones that have 95 and 100 % similarity for nucleotide sequences and 100 % for protein sequences. Analysis was performed using the Vegan package in R

For *cbbL* IA/IB, the 353 clones were binned into 98 ORUs that had a unique amino acid sequence. Similarly, for types IA/IC, a total of 158 ORUs were defined from 444 clones (Table S1). For both subtypes, most ORUs comprised only one or two derived protein sequences and little overlap existed between ORUs originating from different terrestrial samples (Figure 3). After maximum likelihood analysis of the *cbbL* IA/IB sequences, together with reference data, all 98 ORUs grouped into 9 visual clusters and 3 separate ORUs dispersed between other available sequence data (Figure 4). Clusters 4 to 9 and two separate ORUs grouped with cyanobacterial type IB. The number of type IB sequences recovered from samples KP2, KP15, KP43 and KP53, was 115, 35, 15 and 4, respectively (Table 4). Blast searches revealed none of them to be similar to eukaryotic *cbbL*. They made up 50 ORUs that showed very little overlap between samples. In fact, only two visual clusters (clusters 6 and 7) (Table 4) in the ML tree (Figure 4) contained ORUs recovered from multiple samples. Most

type IB ORUs contained sequences originating from a single sample. The type IB sequences showed a clear taxonomic grouping of Cyanobacteria with clusters of *Oscillatoriales*, *Nostocales* and *Chroococcales* sequences (Figure 4) and two clone clusters (4 and 8) showed a close affiliation to *Chroococcales cbbL* IB sequences. Separate ORU KP2.RuBisCO.AB.Clone89 and clusters 5 – 7, all from samples KP43 and KP53, grouped with sequences of *Nostocales*, while cluster 9 and singleton KP15.RuBisCO.AB.Clone61 grouped with the *Oscillatoriales*. Our type IB sequences mostly appear cosmopolitan as their closest relatives originate from samples taken in a variety of ecosystems all over the world. A few (*e.g.* KP15.RuBisCO.AB.Clone61, Cluster 8, Cluster 9), however, have a sequence from Antarctica or other cold environments as closest neighbor (Table 4) [87, 256].

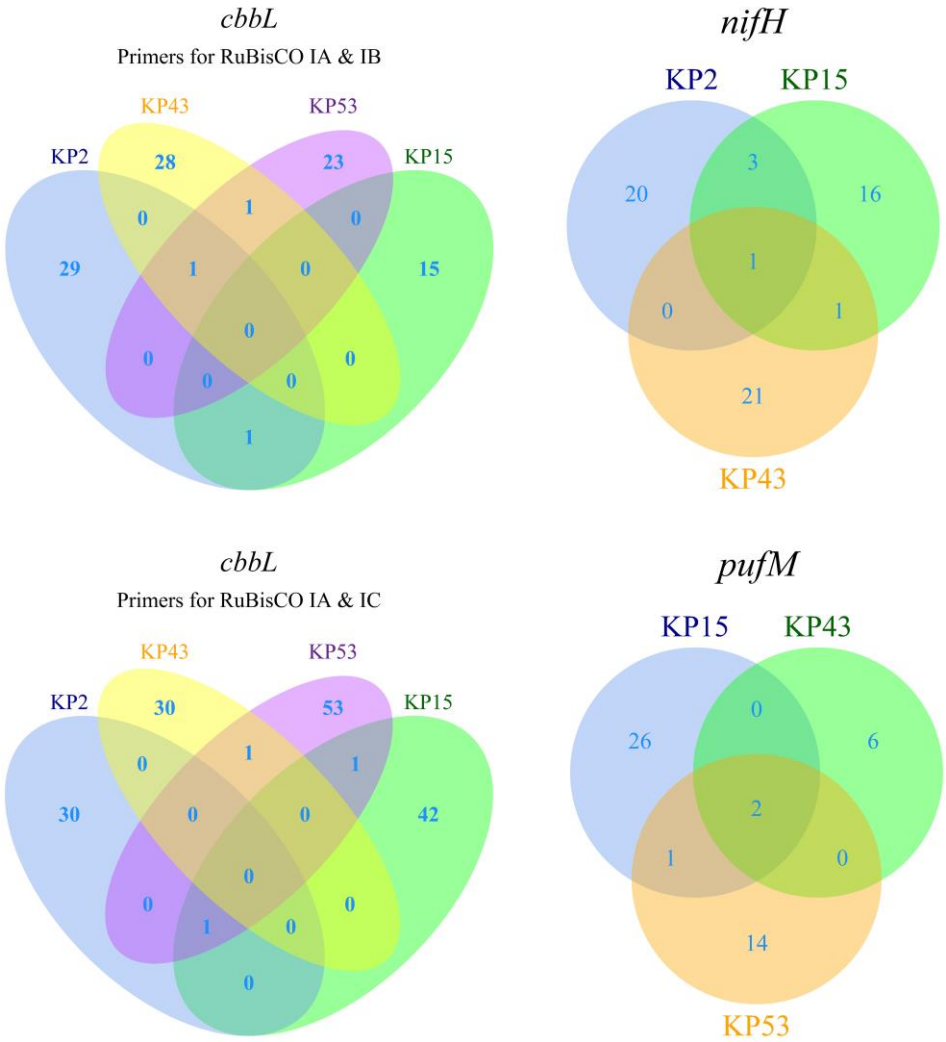

Figure 3 Unweighted Venn diagram of gene protein ORUs/ONUs/OPUs, calculated using the VennDiagram package (http://cran.r-project.org/web/packages/VennDiagram/index.html) in R. *Colored areas* show the number of ORUs/ONUs/OPUs present in the terrestrial samples or shared between multiple samples

Cluster 3 plus a separate ORU grouped with type IA and these sequences showed a close affiliation with the *cbbL* type IA sequence from *Bradyrhizobium* sp. BTAi1 (Figure 4). Remarkably, despite the amplification with primers targeting only type IA/IB, two clusters (1 and 2) grouped with type IC sequences (Figure 4), in particular with unnamed clone sequences obtained from Chinese arid (Accession no. FR849177, FR849160, FR849127) and German agricultural (Accession no. AY572135) soil.

A similar analysis of the clone sequences for *cbbL* IA/IC grouped all 148 ORUs into 16 clusters and 26 separate ORUs (Figure 5, Table 4). Only 1 separate ORU, containing one sequence (KP15.RuBisCO.AC.Clone113), grouped with sequences of type IA. It grouped with the *cbbL* IA sequence of *Bradyrhizobium* sp. BTAi1, as was the case for all type IA sequences obtained with the other primer set. The remaining 147 ORUs, containing 443 sequences, all grouped with *cbbL* IC (Table 4). Apart from cluster 12 which grouped with *Nocardia* species and cluster 16, which grouped with *Nitrosospira* species, most new cloned sequences grouped with sequences of other uncultured bacteria (Figure 5, Table 4).

For amplification of a ~360 bp fragment of the *nifH* gene, a degenerate primer set, previously described by Poly and colleagues [243], was used. The primers target two of the three main conserved *nifH* regions, that have been used to design nearly all existing *nifH* primers [111]. For sample KP53, no successful amplification could be obtained. For the other samples, amplification was successful: 115, 119 and 105 clones were retained for samples KP2, KP15 and KP43 respectively (Table 3). Overall, 62 ONUs were obtained. Bray-Curtis dissimilarity again revealed hardly any overlap between the data from different samples, except for samples KP2 and KP15 (Table 3), which shared 4 protein sequences (Figure 3), of which one (ONU 1) was recovered from all three samples (KP2: 85 clones, KP15: 92 clones, KP43: 5 clones) (Table S1). The overall coverage of the *nifH* clone libraries, based on the Chao1 parameter, was highest for sample KP43 (45.8%), but dropped to 44.4% and 27.5% for clone libraries of KP15 and KP2 respectively. Evenness was low for all samples (Table 3). After maximum likelihood analysis of the protein sequences, together with our *nifH* database sequences, revealed all our ONUs to group with Cyanobacteria (Figure 6). They formed one cluster together with several *Nostocales* reference strains, especially *Nostoc* (Table 4).

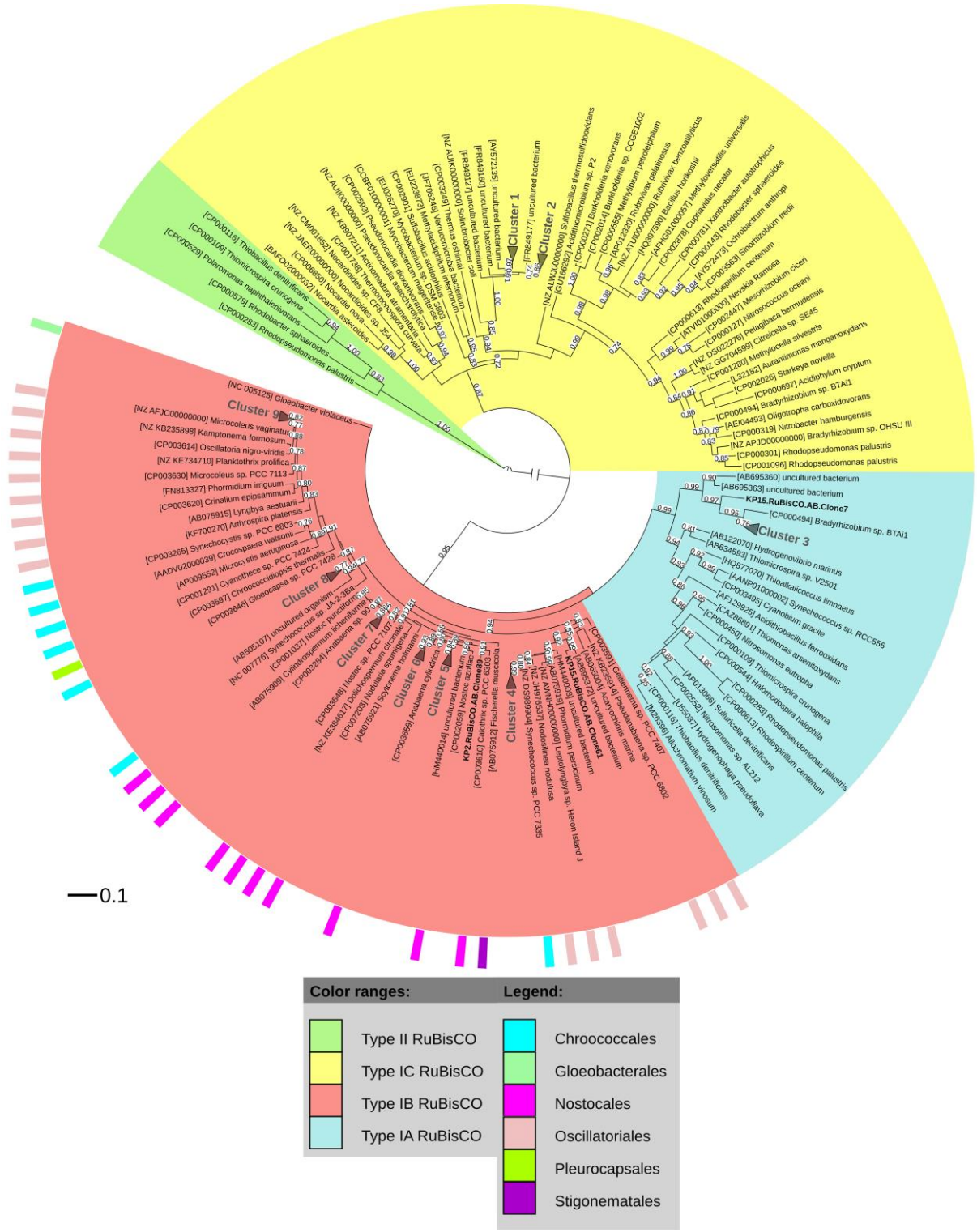

Figure 4 Maximum likelihood phylogenetic tree (1000 bootstrap replicates) of *cbbL* IA and IB amino acid sequences (267 AA). Scale bar indicates 0.1 substitutions per amino acid position. Clones are named by terrestrial sample_gene name_index number. Separate ORUs are labelled in bold. Accession numbers are listed next to each sequence name. Bootstrap values <70 % are not shown. *cbbM* sequences of *Rhodopseudomonas palustris* BisB5, *Rhodobacter sphaeroides* ATCC 17029, *Polaromonas naphthalenivorans* CJ2, *Thiomicrospira crunogena* XCL-2 and *Thiobacillus denitrificans* ATCC 25259 (accession numbers CP000283, CP000578, CP000529, CP000109 and CP000116) were used as an outgroup. Different cyanobacterial orders are labelled with a *colored bar*. Visual clades are displayed as simplified *gray triangles*

Table 4 Composition of visual clusters and singletons and the origin of, and similarity with their closest neighbor, per clone library. Colors indicate *cbbL* subtype (IA (*blue*), IB (*red*), IC (*yellow*)), *ENV* refers to environmental clones

Gene	Cluster	No. of unique protein sequences	No. of Sequences				Closest neighbor		
			KP2	KP15	KP43	KP53	Accession no.	Origin	Similarity (%)
cbbL	1	38			108	43	AY572135	ENV; Germany; Agricultural soil	94.1-96.1
Primers for IA/IB	2	7				29	FR849177	ENV; China; Arid soil	96.5-97.3
	3	2		2			CP000494	*Bradyrhizobium* sp. BTAi1; USA; Stem nodules of *Aeschynomene indica*	91.4-93.2
	KP15.RuBisCO.AB.Clone7	1		2			CP000494	*Bradyrhizobium* sp. BTAi1; USA; Stem nodules of *Aeschynomene indica*	88.0
	4	4	9				NZ_DS989904	*Synechococcus* sp. PCC 7335; Mexico; Snail shell	94.1-95.7
	5	3			9		NZ_KK073768	*Scytonema hofmanni*; Limestone	99.6-100
	6	11	16		5	4	CP003659	*Anabaena cylindrica* PCC 7122; United Kingdom; Water	94.1-97.7
	7	2	2	2	1		CP003284	*Anabaena* sp. 90; Finland; Lake water	96.9-98.0
	8	17	83				AB505107	ENV; Japan; Iron-rich snow	91.0-91.8
	9	11		32			AB695365[a]	ENV; Antarctica; Aquatic moss pillar	98.4-100
	KP15.RuBisCO.AB.Clone61	1		1			AB695372	ENV; Antarctica; Aquatic moss pillar	96.9
	KP2.RuBisCO.AB.Clone89	1	5				CP003610	*Calothrix* sp. PCC 6303; USA; Freshwater	95.7
cbbL	1	2	3				JQ964878	ENV; China; Agricultural soil	78.5-84.6
Primers for IA/IC	2	32			55	46	AY422925	ENV; Hawaii; 210-yr old Kilauea Caldera Rim volcanic deposit	89.3-91.3
	3	17		45			KF523929	ENV; China; Soil	92.6-94.0
	4	7		15		2	NZ_AUIK00000000	*Solirubrobacter soli* DSM 22325; Korea; Ginseng field soil	80.5-83.2
	5	3	9				NZ_AUIK00000000	*Solirubrobacter soli* DSM 22325; Korea; Ginseng field soil	86.6-89.2
	6	5		15			JQ964837	ENV; China; Agricultural soil	88.6-90.6
	7	2			3		JAOB01000069	*Mycobacterium xenopi* 4042; Human lung	90.6-91.9
	8	4				6	JQ836503	ENV; China; Soil	85.2-86.6
	9	14	21			34	JQ836503	ENV; China; Soil	85.2-87.2
	10	2				2	NZ_JAER00000000	*Nocardioides* sp. J54; Chile; Atacama desert	86.6-87.2
	11	5		9		1	FO082843	*Nocardia cyriacigeorgica* GUH-2; Human	85.2-86.6
	12	2		3			NZ_BAFO00000000	*Nocardia asteroides* NBRC 15531	90.6-92.6
	13	11	36				GU166292	*Acidithiomicrobium* sp. P2; Pyrite enrichment culture	88.6-90.6
	14	2	3				HQ174601	ENV; China; Cropland soil	87.2-87.9
	15	6				12	HQ388736	ENV; China; Paddy soil	84.6-88.6
	16	17	4		63		CAUA01000016	*Nitrosospira* sp. APG3; USA; Freshwater lake sediment	94.6-96.6
	KP15.RuBisCO.AC.Clone47	1		1			JQ836457	ENV; China. Soil	84.6
	KP53.RuBisCO.AC.Clone141	1				1	JQ836514	ENV; China. Soil	88.6
	KP2.RuBisCO.AC.Clone100	1	2				NZ_AUIK00000000	*Solirubrobacter soli* DSM 22325; Korea; Ginseng field soil	91.3
	KP53.RuBisCO.AC.Clone4	1				3	AY422925	ENV; Hawaii; 210-yr old Kilauea Caldera Rim volcanic deposit	97.3
	KP15.RuBisCO.AC.Clone109	1		2			AY422925	ENV; Hawaii; 210-yr old Kilauea Caldera Rim volcanic deposit	98.0
	KP2.RuBisCO.AC.Clone102	1	1				JQ836516	ENV; China; Soil	87.2
	KP43.RuBisCO.AC.Clone116	1			1		JQ836516	ENV; China; Soil	85.9
	KP2.RuBisCO.AC.Clone84	1	1				NZ_CM001852	*Nocardioides* sp. CF8; USA; Soil	87.2
	KP53.RuBisCO.AC.Clone67	1				1	JQ964872	ENV; China; Soil	94.6
	KP15.RuBisCO.AC.Clone5	1		1			DQ149782	ENV; USA; Pine forest and agricultural soils	97.3
	KP2.RuBisCO.AC.Clone103	1	4				JQ964901	ENV; China; Soil	93.3
	KP15.RuBisCO.AC.Clone8	1		2			NZ_ATUI00000000	*Rubrivivax benzoatilyticus* JA2; India; Paddy soil	92.6
	KP15.RuBisCO.AC.Clone148	1		2			KJ720517	ENV; China; Soil	96.0
	KP15.RuBisCO.AC.Clone104	1		5		3	HQ174649	ENV; China; Cropland soil	100
	KP15.RuBisCO.AC.Clone152	1		1			HQ174649	ENV; China; Cropland soil	99.3
	KP15.RuBisCO.AC.Clone52	1		1			HQ174649	ENV; China; Cropland soil	98.0
	KP53.RuBisCO.AC.Clone92	1				1	KJ720371	ENV; China; Soil	98.7

Gene	Cluster	No. of unique protein sequences	No. of Sequences				Closest neighbor		
			KP2	KP15	KP43	KP53	Accession no.	Origin	Similarity (%)
	KP53.RuBisCO.AC.Clone47	1				1	KJ720371	ENV; China; Soil	98.7
	KP53.RuBisCO.AC.Clone25	1				1	AY422900	ENV; Hawaii; 300-yr old forest soil volcanic deposit	86.5
	KP2.RuBisCO.AC.Clone77	1	1				HQ997091	ENV; China; Cropland soil	98.0
	KP2.RuBisCO.AC.Clone117	1	1				KJ720363	ENV; China; Soil	98.7
	KP2.RuBisCO.AC.Clone1	1	2				KJ720363	ENV; China; Soil	98.7
	KP53.RuBisCO.AC.Clone83	1				1	HQ997138	ENV; China; Cropland soil	98.7
	KP15.RuBisCO.AC.Clone103	1	7	4		1	KJ720441	ENV; China; Soil	100
	KP15.RuBisCO.AC.Clone18	1		1			JQ836500	ENV; China; Soil	98.7
	KP2.RuBisCO.AB.Clone110	1	3				DQ149787	ENV; USA; Pine forest and agricultural soils	97.3
	KP15.RuBisCO.AC.Clone113	1		1			CP000494	*Bradyrhizobium* sp. BTAi1; USA; Stem nodules of *Aeschynomene indica*	95.3
nifH	1[b]	60	110	118	105	-		*Nostocaceae* Cyanobacteria	95.1-100
pufM	1	10	-	5	68	22	AB486028	ENV; China; Paddy soil	91.4-94.8
	2	13	-	24		24	AVFL01000011	*Skermanella stibiiresistens* SB22; China; Iron mine soil	86.2-94.8
	3	4	-	24			JF906277	ENV; Germany; Freshwater lake	96.6-98.3
	4	2	-	17			CP001280	*Methylocella silvestris* BL2; Germany; Acidic forest cambisol	93.1-94.8
	5	4	-	14			KC900142	ENV; Antarctica; Great Wall Cove	94.8-96.6
	6	11	-	8	30	75	JF906279	ENV; Germany; Freshwater lake	93.1-91.4
	KP15.pufLM.Clone55	1	-	2			KC900120	ENV; Antarctica; Great Wall Cove	94.8
	KP15.pufLM.Clone78	1	-	1			KC465437	*Erythrobacter* sp. HU12-14; Mongolia; Lake Hulunhu	94.8
	KP15.pufLM.Clone51	1	-	1			HF947099	ENV; China; Lake Taihu	98.3
	KP15.pufLM.Clone14	1	-	1			JQ340684	ENV; Pacific Ocean	94.8
	KP15.pufLM.Clone13	1	-	1			AB486028	ENV; China; Paddy soil	96.6

[a] The environmental sequence with accession number AB695365 is enclosed in the cluster

[b] The following reference sequences are enclosed in the cluster: L23514 (*Nostoc commune*), CP001037 (*Nostoc punctiforme* PCC 73102), KC243672 (*Nostoc* sp. LEGE 06106), U04054 (*Nostoc muscorum*), CP007203 (*Nodularia spumigena* CCY9414), CP003642 (*Cylindrospermum stagnale* PCC 7417), CP003943 (*Calothrix* sp. PCC 7507), and environmental sequences DQ995890, EU915059, AY819602, HM140755, DQ140640, GU196863, KC667460 and KC140445

For amplification of *pufLM* genes a degenerate primer set was used (Table 2) that amplifies a ~1500 bp long fragment, providing additional sequence information compared to other primer sets that only amplify small parts of the *pufL* or *pufM* gene. All samples, except for KP2, showed successful amplification. Cloning resulted in 98, 98 and 121 positive sequences for samples KP15, KP43 and KP53, respectively (Table 3).

Analysis of the complete ~1500 bp fragment revealed 193 OPUs with a unique amino acid sequence (data not shown). To date, most available sequences comprise a smaller part of the *pufM* gene. Therefore, to broaden phylogenetic analysis by comparing our new sequences with as much reference data as possible, only this small part of the *pufM* gene, covering 58 AA, was used for further analysis. This increased the reference data in the phylogenetic analysis by 17 times, while reducing our sequences to 49 unique OPUs (Table S1), 46 of which consisted of sequences originating from just one of the terrestrial samples. Of the other 3 OPUs, two (OPU 22 and OPU 40) were shared between three terrestrial samples and one (OPU 2) between samples KP15 and KP53 (Figure 3).

Bray-Curtis analysis indicated samples KP43 and KP53 to show only 57% dissimilarity, while all other relationships show a very high dissimilarity (Table 3). Evenness was highest for the KP15 library (81.4%) and 49.6% and 52.5% for the KP43 and KP53 libraries respectively. Coverage of the libraries was 62.8%, 91.4% and only 30.4% for samples KP15, KP43 and KP53, respectively (Table 3).

For phylogenetic analysis, a maximum likelihood tree (1.000 bootstraps), was generated (Figure 7), in which the 49 OPUs could be grouped in 6 visual clusters and five separate OPUs (Table 4). Sequences obtained from sample KP15 proved to be the most diverse, with OPUs belonging to all six clusters and all five separate OPUs. Sequences from samples KP43 and KP53 were much less diverse, belonging to only two and three of the visual clusters, respectively (Table 4).

Only for cluster 5 and singleton KP15.pufLM.Clone55, the closest affiliated unknown *pufM* sequences (Accession no. KC900142 and KC900120) also originates from Antarctica (Figure 7), however, they were not found in a terrestrial environment, but in surface water from Great Wall Cove, King George Island, Antarctica. Furthermore, cluster 1 and singleton KP15.pufLM.Clone13 (Figure 7) group with unknown *pufM* sequences originating from Chinese paddy soil samples [149, 257, 258]. Clusters 2 and 3, and singletons KP15.pufLM.Clone14, -Clone55 and –Clone78 grouped into a larger cluster (from *Sphingomonas* sp. PB56 [AY853583] to *Sphingomonas echinoides* ATCC 14820 [NZ_JH584235]) (Figure 7) of *pufM* sequences belonging mainly to organisms of the

alphaproteobacterial orders of the *Rhizobiales*, originating from terrestrial and aquatic environments, and *Sphingomonadales*, predominantly originating from aquatic environments.

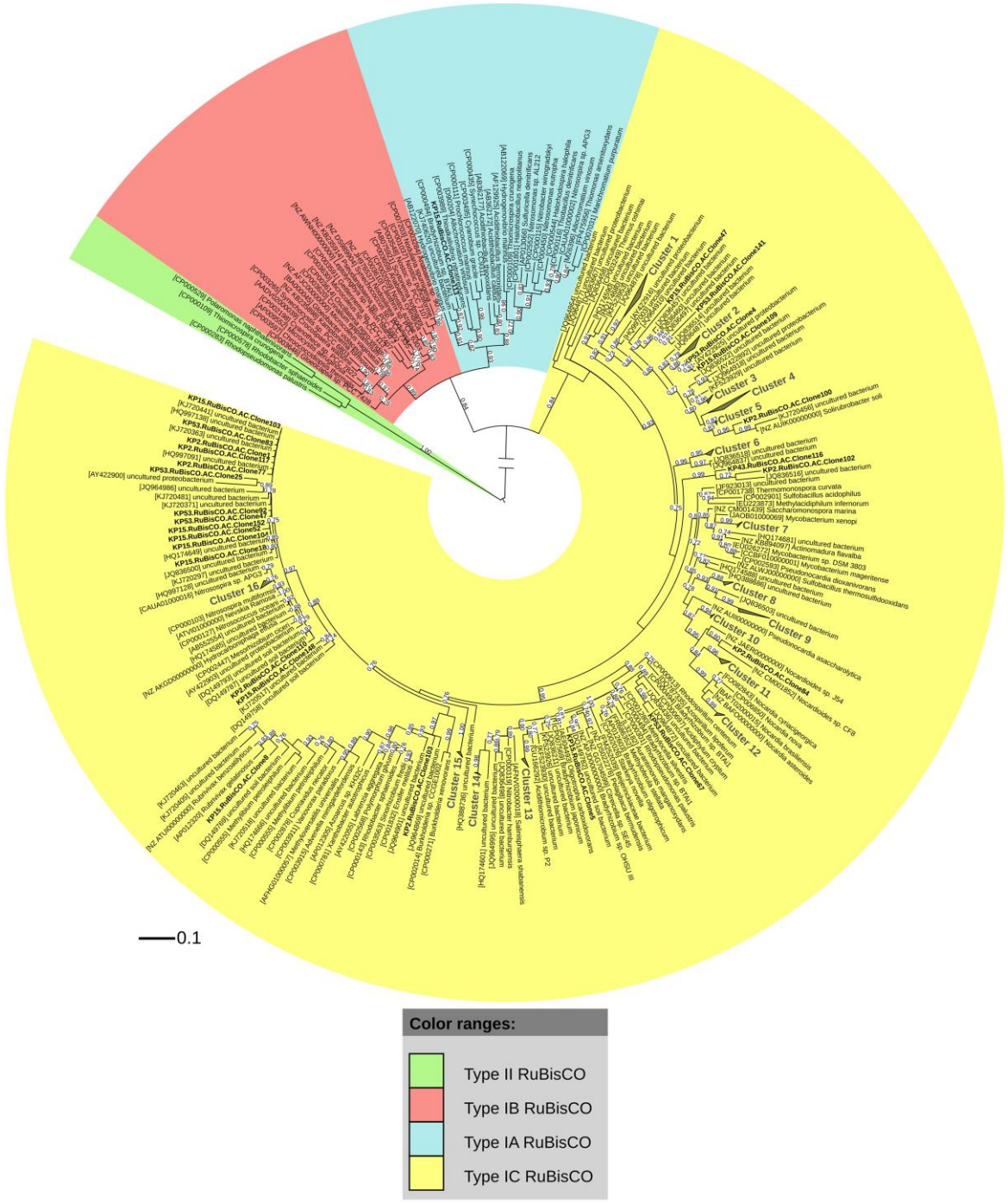

Figure 5 Maximum likelihood phylogenetic tree (1000 bootstrap replicates) of *cbbL* IA and IC amino acid sequences (154 AA). Scale bar indicates 0.1 substitutions per amino acid position. Clones are named by terrestrial sample_gene name_index number. Separate ORUs are labelled in bold. Accession numbers are listed next to each sequence name. Bootstrap values <70 % are not shown. *cbbM* sequences of *Rhodopseudomonas palustris* BisB5, *Rhodobacter sphaeroides* ATCC 17029, *Polaromonas naphthalenivorans* CJ2 and *Thiomicrospira crunogena* XCL-2 (accession numbers CP000283, CP000578, CP000529 and CP000109) were used as an outgroup. Visual clades are displayed as *gray triangles*. Total branch lengths to the closest and the farthest leaf are used as sides *of the triangle*

2.4. Discussion

Based on previous research, where Cyanobacteria were reported in small numbers in some exposed Antarctic communities [215-217], we hypothesized that other members of the bacterial community may take their role as primary producers in the oligotrophic soils of the Antarctic continent. We investigated this in four samples from soils near the Princess Elisabeth Station in the Sør Rondane Mountains. As part of a larger biogeographic study these samples were recently included in Illumina sequencing of 16S rRNA genes (Tahon et al, unpublished data, see Chapter 3) and preliminary analyses indicate that Cyanobacteria make up 0.04 to 51.85% of the reads recovered from our four samples (Table 1). We studied the diversity of genes encoding RuBisCO as a marker for phototrophic and chemolithotrophic primary producers and the nitrogenase gene *nifH* as a marker for nitrogen fixation. The scarcity of organic matter and abundant availability of sunlight would favor bacteria that are able to exploit the latter resource. Therefore we also investigated the diversity of type 2 phototrophic reaction centers (*pufLM* genes), which are also found in aerobic photohetertrophs, and PR light harvesting systems.

Few studies have reported the presence and diversity of *cbbL* and/or *cbbM* genes in Antarctica, most of them investigating aquatic ecosystems [87, 99, 100], leaving the ice-free terrestrial regions unexplored. Here, two different degenerate primer sets were used to target the diversity of RuBisCO type I (*cbbL* gene) in soils near the Princess Elisabeth Station. The first amplified both types IA and IB, the second types IA and IC. This division of RuBisCO type I into different subtypes can be seen clearly in the phylogenetic trees (Figure 4 and Figure 5). The *cbbL* t*ype* IC, previously found in Alpha-, Beta- and Gammaproteobacteria, Verrucomicrobia, Firmicutes and Actinobacteria, was the most frequently recovered type in all samples (202 ORUs) (Table 4) and showed very high diversity, making up 18 clusters and 26 separate ORUs (Figure 4 and Figure 5, Table 4), suggesting the presence, in the invesitgated terrestrial areas near the Belgian base, of multiple non-cyanobacterial autotrophs, presumably belonging to a wide range of taxa. Remarkably, the primer set targeting types IA/IB also picked up a large number of type IC sequences from samples KP43 and KP53 (108 and 72 sequences, respectively) (Table 4), indicating that the primer set, although successfully amplifying only *cbbL* IA/IB in previous studies [239, 254, 259-261], is not 100% specific for these two subtypes.

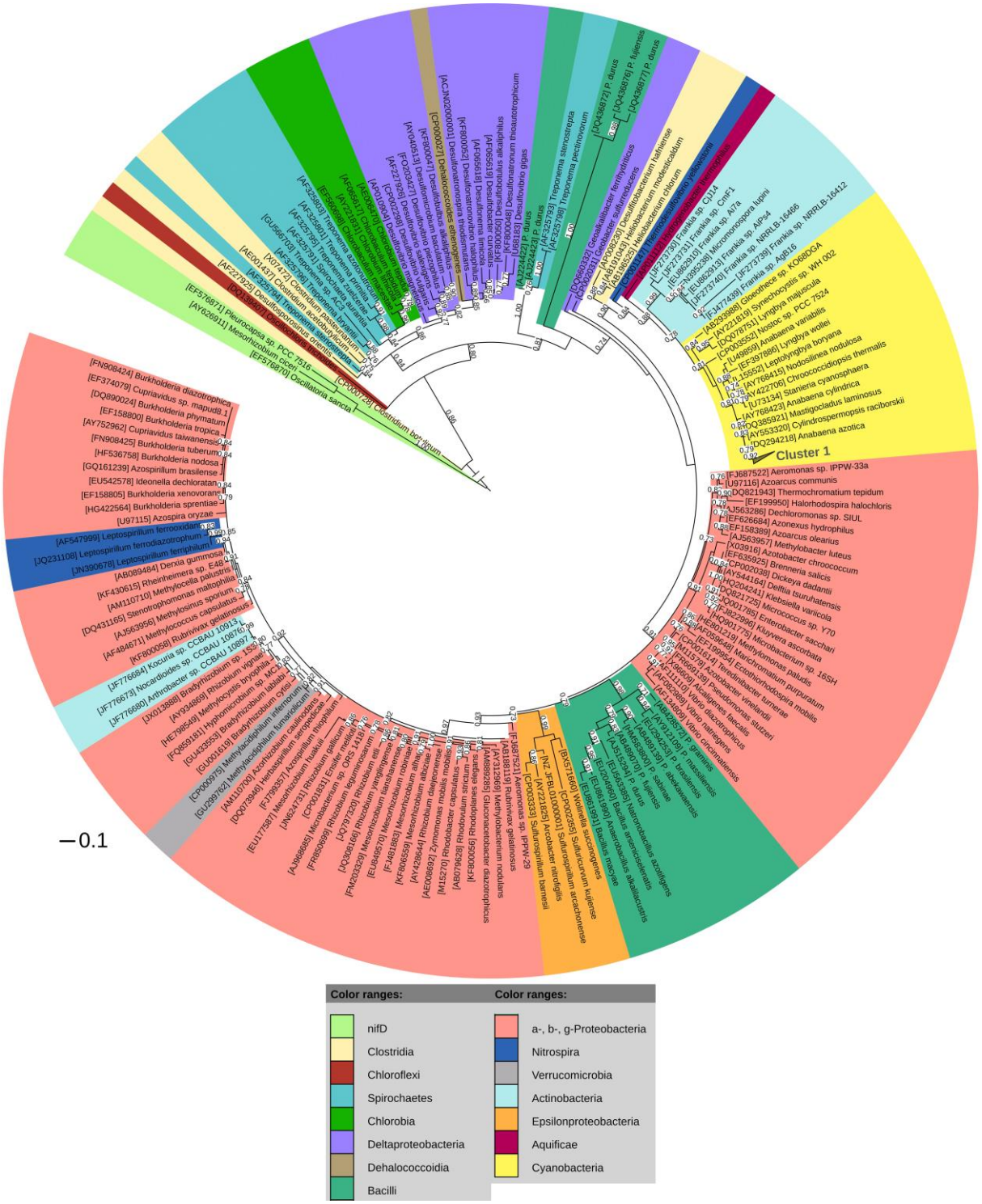

Figure 6 Maximum likelihood phylogenetic tree (1000 bootstrap replicates) of *nifH* amino acid sequences (106 AA). Scale bar indicates 0.1 substitutions per amino acid position. Clones are named by terrestrial sample_gene name_index number. Separate ONUs are labelled in bold. Accession numbers are listed (between brackets) next to each sequence name. Bootstrap values <70 % are not shown. *nifD* sequences of *Pleurocapsa* sp. PCC 7516 (EF576871), *Mesorhizobium ciceri* (AY626911) and *Oscillatoria sancta* PCC 7515 (EF576870) were used as an outgroup. Visual clades are displayed as *gray triangles*. Total branch lengths to the closest and the farthest leaf are used as sides *of the triangle*

Type IA *cbbL*, typical of Alpha-, Beta- and Gammaproteobacteria, was amplified only in sample KP15 with a very limited diversity of only 4 ORUs (5 DNA sequences), despite the use of two primer sets that target this type (Table 4). Possible reasons for this low diversity of type IA may be a primer bias, the absence or low abundance of the targeted *cbbL* type IA. Based on literature data, the latter may be the case, as both primer sets were shown to display only limited bias amongst the different *cbbL* types [240] and they successfully amplified type IA in other studies [86, 87, 235, 240, 254, 255]. A further possibility is the presence of slightly different type IA genes that were not picked up by the primers. Primer mismatch is a weakness for all PCR-based diversity surveys and may cause a portion of the actual diversity to be missed.

Finally, a large diversity of cyanobacterial *cbbL* type IB was recovered (Figure 4). While there was very little overlap between samples, KP2 and KP15 yielded the most diversity, in line with the higher relative abundance of Cyanobacteria in those samples (Tables 1 and 4).

Taken together, the diversity of RuBisCO type I sequences recovered with two primer sets from the terrestrial Antarctic samples revealed a larger relative diversity of non-cyanobacterial primary producers (202 ORU of type IC and 4 ORU of type IA) than Cyanobacteria (50 ORU), confirming our hypothesis that these groups may potentially contribute to primary production in these systems. Furthermore, the presence of *cbbL* type IA sequences similar to those retrieved from the alphaproteobacterium *Bradyrhizobium* sp. BTAi1, type IB sequences from *Oscillatoriales*, *Nostocales* and *Chroococcales*, and type IC from *Nitrosospira* and *Rubrivivax* corroborates previous reports from Antarctic or subglacial environments [87, 262].

It should be noted that while many of the clone sequences show relatively high similarity to sequences of cultivated bacteria, this does not imply taxonomic relatedness. The *cbbL* phylogeny shows a high degree of incongruency with the 16S rRNA phylogeny, due to horizontal gene transfer and gene duplication associated with differential gene loss [263]. Considering the type of habitat related reference sequences originate from, no clear grouping of cold habitat sequences was apparent for any of the *cbbL* types (Table 4).

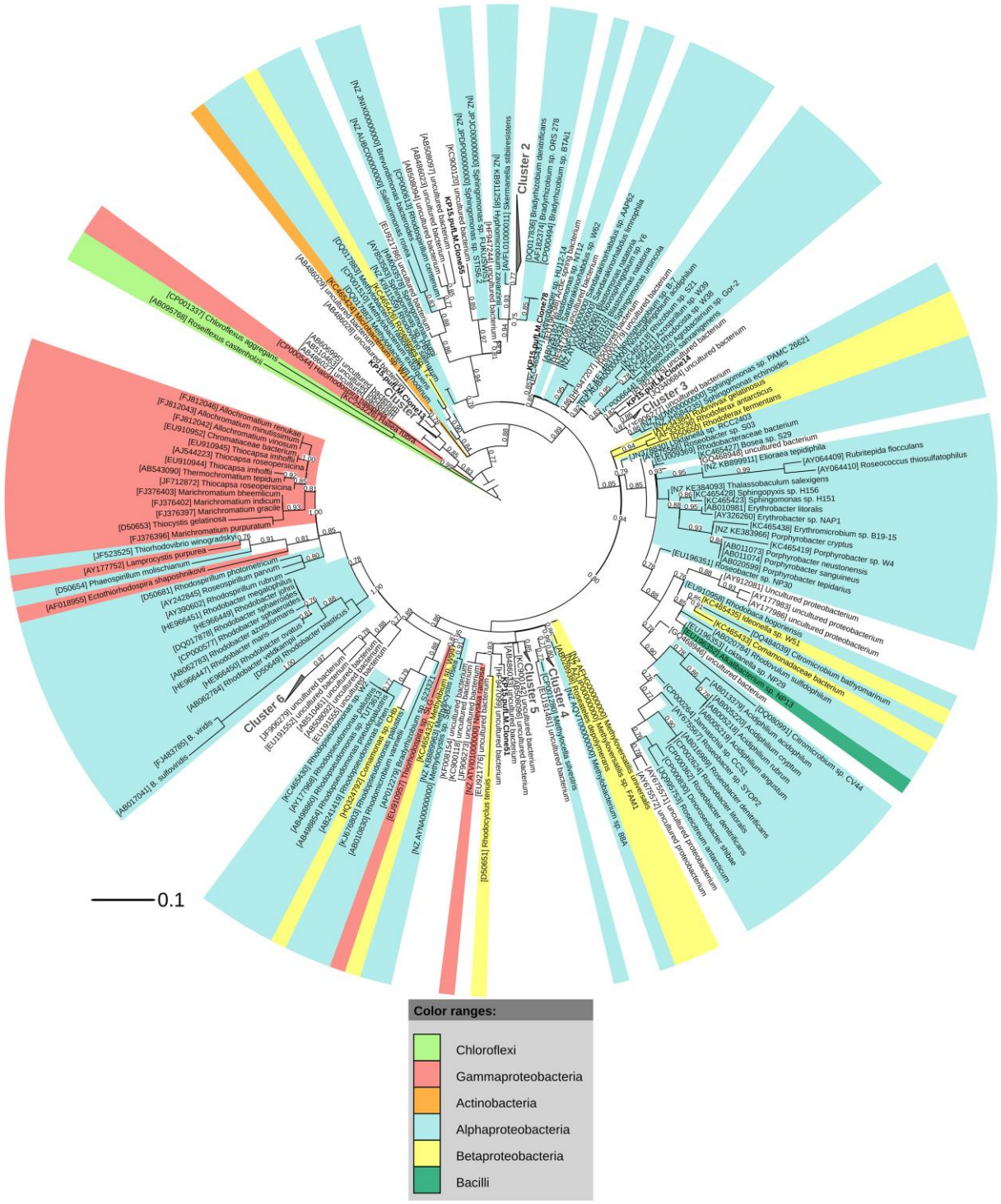

Figure 7 Maximum likelihood phylogenetic tree (1000 bootstrap replicates) of *pufM* amino acid sequences (58 AA). Scale bar indicates 10 substitutions per 100 amino acid positions. Clones are named by terrestrial sample_gene name_index number. Separate OPUs are labelled in bold. Accession numbers are listed next to each sequence name. Bootstrap values <70 % are not shown. *Chloroflexus aggregans* and *Roseiflexus castenholzii* (accession numbers CP001337 and AB095768) were used as an outgroup. Visual clades are displayed as *gray triangles*. Total branch lengths to the closest and the farthest leaf are used as sides *of the triangle*

Previous studies reported the presence of type II RuBisCO in Antarctica [99, 100], however, despite the use of a primer set recently shown to amplify a 700-800 bp fragment of the *cbbM* gene in a broad variety of bacteria [59], no positive amplification or no amplicon of the correct size could be obtained from our samples (data not shown). These results imply that type II RuBisCO sequences are absent or below the detection limit in our samples or that they do not match the primers used.

We also studied the diversity of dinitrogenase-reductase (*nifH*) genes to assess to what extent nitrogen fixation might be performed by non-cyanobacterial taxa. The primer set used targets a broad variety of Alpha-, Beta- and Gammaproteobacteria, Gram positive bacteria and only a small percentage of known cyanobacterial *nifH* sequences [111, 243]. All obtained sequences, however, showed a close relationship to known *nifH* sequences belonging to *Nostocales* cyanobacteria and especially the heterocystous cyanobacterium *Nostoc*. No non-cyanobacterial sequences were recovered, suggesting that non-Cyanobacteria contribute little to nitrogen fixation in these terrestrial Antarctic samples. Although we recovered mostly *Nostocales*-affiliated *nifH* sequences, this does not imply these are the main nitrogen fixers because the primers used did not target the majority of Cyanobacteria [111]. Since rarefaction curves based on unique sequences did not reach saturation (Figure 2) and coverage estimates are below 50% (Table 3), it is possible that other Cyanobacteria as well as non-Cyanobacteria may contribute to nitrogen fixation in the terrestrial Antarctic. The presence of *Nostocales* Cyanobacteria, and in particular *Nostoc*, in our terrestrial Antarctic samples as suggested by *nifH* findings, is corroborated by the *cbbL* type IB data (Figure 4). *Nostoc* is indeed frequently found in Antarctic environments [106, 125, 128, 216] as well as other terrestrial ecosystems characterized by aridity and nutrient limitations [264]. Its success in these extreme environments is due to its ability to remain desiccated for extended periods of time and, after rehydration, recover its metabolic activity completely within hours to days [264]. Moreover, *Nostoc* can withstand repeated cycles of freezing and thawing, a typical Antarctic environmental condition [23, 265]. No particular psychrophilic nitrogenase associations could be detected: the most closely related *nifH* sequences originated from a variety of ecosystems from all over our planet (Table 4).

In the present study, two mechanisms to harvest sunlight were examined. The first, proteorhodopsin, has been almost exclusively studied in aquatic ecosystems, including from Arctic and Antarctic marine environments [191, 193, 266]. Nevertheless, PR-like variants

have been found in Siberian permafrost samples [183], suggesting that they could also play a role in continental Antarctica. However, proteorhodopsin genes were not detected in any of the samples tested here, despite the use of multiple primer sets (Table 2), amplifying proteorhodopsin in a wide range of bacteria [191, 203, 237, 266]. Possibly proteorhodopsins are not abundant or maybe not present at all in the investigated soils. Another likely explanation may be that the terrestrial Antarctic variants of proteorhodopsin are not targeted by these primer sets: as most reference data are from marine systems, the available primers may not fully capture terrestrial proteorhodopsin diversity.

A second mechanism to harvest sunlight that was investigated here is anoxygenic phototrophy using type 2 reaction centers, encoded by *pufLM* genes and found in Chlorofexaceae, Proteobacteria (purple bacteria), Gemmatimonadetes and Firmicutes [267]. Diverse *pufLM* genes were recovered from all but one sample, indicating that anoxygenic phototrophs are present. We did not recover sequences representative of strictly anaerobic photoautotrophic Purple sulfur bacteria or Chloroflexi and only cluster 6 grouped among sequences of Purple non-sulfur bacteria (Figure 7). Most of our clusters grouped among diverse lineages originating from aerobic photoheterotrophic taxa of Alpha- and Betaproteobacteria referred to as AAP (Figure 7). Studies on AAPs in soils are extremely rare to date. Only Feng and colleagues [149, 205, 257, 258, 268] investigated their presence in paddy soils from China and in Arctic soils. Other studies to date focused on aquatic ecosystems [174, 175, 269], where it has been shown that AAP represent an important part of the total bacterial community [270]. This work, with the Sør Rondane Mountains as area of study, represents the first report of AAP bacteria in Antarctic soils.

Since to date most available sequences only comprise a small part of the 3' side of the *pufM* gene, this 58 AA long region was used for analysis. When expanding sequence data to the larger, originally obtained *pufLM* protein sequences, these affiliations with known AAP bacteria can still be seen, indicating that this 58 AA long region of *pufM* is indeed a useful indicator for *pufLM* phylogeny. In the terrestrial samples, a high diversity of *pufM* genes was present and clear differences were observed between the samples.

Most of the *pufM* diversity recovered here affiliated with alphaproteobacterial *pufM* (Table 4, Figure 7), an observation made previously in the Arctic [205] and Antarctic [175, 233], but also in Chinese paddy soils [258]. Furthermore, some of the alphaproteobacterial-like OPUs were found grouping with *pufM* sequences originating from *Sphingomonadales* bacteria, an order that has previously been shown to be present in soils of the Sør Rondane Mountains [46]. Gammaproteobacterial-like *pufM* sequences were also found in our samples (Table 4,

Figure 7), but contributed less to the general *pufM* diversity, as was the case for Arctic soils [205], but not Antarctic sea ice and seawater, where no *pufM* of the gammaproteobacterial group was detected [175]. Again, no clear psychrophilic *pufM* groupings could be deduced from the geographic origin of these closest relatives, as they originate from marine and terrestrial environments from all over the world, a trend that could also be seen for the other functional genes (Table 4).

2.5. Conclusion

Overall, the data presented in this study suggest that, in soils in the vicinity of the Princess Elisabeth Station, a broad diversity of microorganisms harboring RuBisCO genes is present. Type IA was only scarcely recovered and all sequences showed close affiliation to the *cbbL* type IA sequence of *Bradyrhizobium* sp. BTAi1. Non-cyanobacterial type IC was dominantly retrieved and appeared to be more diverse than cyanobacterial type IB in the bacterial autotrophic communities near the research station, and may potentially contribute to the input of organic matter into the oligotrophic Antarctic systems. The *nifH* diversity recovered was low, as only sequences affiliated with *Nostocales* were recovered from our samples, suggesting non-Cyanobacteria are not important contributors to nitrogen fixation in the poor Antarctic soils. Our study of mechanisms to harvest solar energy could not detect rhodopsin genes. However, it did, for the first time, show the presence of aerobic anoxygenic phototrophic bacteria in Antarctic soils, suggesting that photoheterotrophy may be a useful life strategy during the austral summer.

The functional genes studied here did not appear to show Antarctic types: the majority of sequences grouped among neighbors from various ecosystems worldwide. For all genes coverage estimates and rarefaction analysis of unique sequences, however, showed that saturation was not reached, indicating that a large proportion of diversity remains unknown. Consequently, more large-scale analyses, including deep sequencing, is required to assess the diversity of these systems.

2.6. Acknowledgements

This work was funded by the Fund for Scientific Research – Flanders (project G.0146.12). Additional support was obtained from the Belgian Science Policy Office (project CCAMBIO). The authors declare that they have no conflict of interest.

2.7. Supplementary information

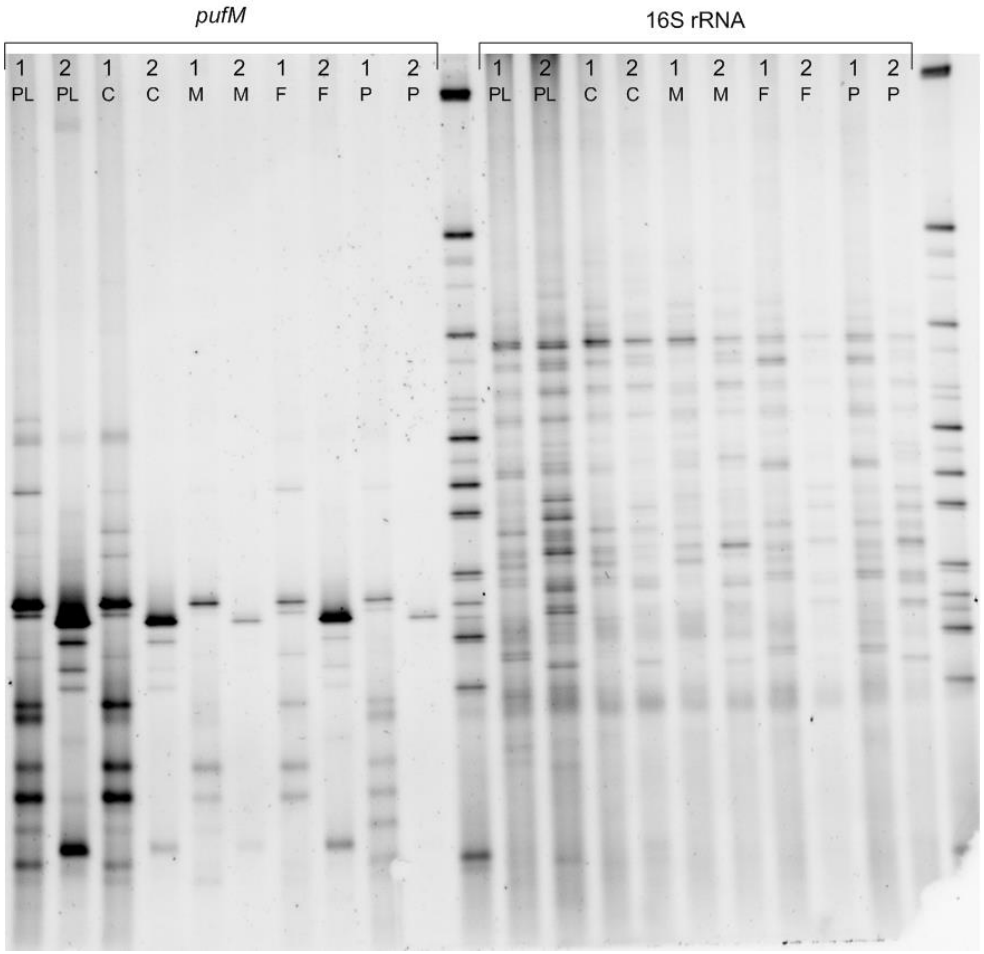

Figure S1 Comparison of genetic fingerprints (DGGE) of two terrestrial samples collected near the Princess Elisabeth Station, by use of five different DNA extraction methods. Partial 16S (V1-V3 region) was amplified using primers V3-357F_GC and V3-518R [271]. Partial *pufM* gene fragments were amplified using primers pufM557F [228] and GC_WAWR [272]. Gene fragments retrieved were visualized using DGGE. Lanes designated as 1 and 2 correspond with the two terrestrial samples tested. Different DNA extraction methods used are labeled as PL (PowerLyzer® PowerSoil® DNA Isolation Kit), C (CTAB), M (Muyzer), F (FastDNA™ SPIN Kit for Soil) and P (Pitcher). In total, seven different DNA extraction methods were tested: CTAB, Pitcher, Muyzer, PowerLyzer® PowerSoil® DNA Isolation Kit, FastDNA™ SPIN Kit for Soil, PowerSoil® DNA Isolation Kit and UltraClean Soil DNA Isolation Kit. For all methods, protocols were tested with and without modifications (*e.g.* the influence of bead bating or an extra lysis step), to optimize DNA yield and diversity. Partial *pufM* and 16S rRNA genes were amplified as described above and visualized using DGGE. The resulting best five DNA extraction methods were compared in the figure

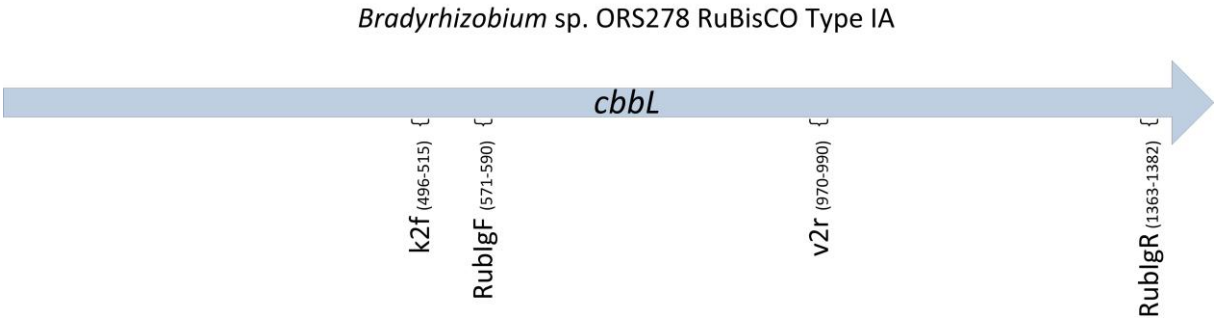

Figure S2 Schematic illustration of the cbbL gene of *Bradyrhizobium* sp. ORS278 (accession number CU234118). Curly brackets indicate the positions of forward and reverse primers used in this study (Table 2), to target the different subtypes of RuBisCO type I. Positions of the primers are assigned according to the DNA sequence of the cbbL gene of *Bradyrhizobium* sp. ORS278 and given next to the primer name, between brackets

Table S1 Composition of all ORUs, ONUs and OPUs per clone library

Table S1 is available at:
https://static-content.springer.com/esm/art%3A10.1007%2Fs00248-015-0704-6/MediaObjects/248_2015_704_MOESM1_ESM.pdf

Chapter 3

Diversity of key genes for primary production and diazotrophy in soils from the Sør Rondane Mountains, East-Antarctica, revealed by amplicon sequencing

Redrafted from:

Tahon G., Tytgat B. and Willems A. (2017). Diversity of key genes for primary production and diazotrophy in soils from the Sør Rondane Mountains, East-Antarctica, revealed by amplicon sequencing. BMC Microbiome, *under review*.

Author's contributions:

GT performed experiments, analyzed the data and drafted the manuscript. BT performed geochemical analyses and contributed to data analyses. AW conceived the project and helped draft the manuscript. All authors have read and approved the final manuscript.

Summary

Background

Although Cyanobacteria are generally considered the most important primary producers and diazotrophs in Antarctic ecosystems, they appear to be scarce in several high-altitude Antarctic regions. In view of these observations, we investigated the presence of non-cyanobacterial carbon and nitrogen fixing microorganisms. An Illumina sequencing approach was used to analyze the bacterial community composition and the diversity and abundance of key genes involved in the carbon and nitrogen fixation processes in oligotrophic exposed soils from the Sør Rondane Mountains, East-Antarctica.

Results

Analysis of the large subunit of type I ribulose-1,5-biphosphate carboxylase/oxygenase genes (*cbbL*), revealed a large actinobacterial, and alpha-, beta- and gammaproteobacterial diversity of cbbL type IC, whereas type IA diversity was restricted to *Bradyrhizobium* sp. BTAi1-like sequences. Although a large portion of the cbbL sequences grouped with those of cultivated bacteria, some belonged to currently unknown phylotypes. Dinitrogenase-reductase genes (*nifH*) most similar to those of Nostocales cyanobacteria, were dominantly retrieved from these oligotrophic soils.

Conclusions

These findings suggest that diverse bacteria capable of assimilating carbon dioxide through the Calvin-Benson-Bassham cycle inhabit these extreme terrestrial systems and may contribute to primary production. Cyanobacteria, present in greatly varying numbers as assessed by Illumina sequencing of a 16S rRNA gene fragment, appear to be the most important nitrogen fixers in these habitats.

3.1. Introduction

Because of their capability to perform carbon and in many cases also nitrogen fixation, as well as withstand severe environmental conditions, Cyanobacteria are generally thought to be the main primary producers and diazotrophs in the microbial communities that dominate Antarctic ecosystems [128, 214, 273]. Over the course of the last decade, however, multiple studies have reported this group to sometimes occur in very low numbers in several samples from different Antarctic environments, ranging from mineral soils in the McMurdo Dry Valleys [215] to fell-field sites on the Peninsula [216, 217] and terrestrial high-altitude samples from the Darwin [274] and Dronning Maud Land mountains [54, 275]. Despite these findings, little is known about which microorganisms might take over the cyanobacterial role of primary production and nitrogen fixation in these Antarctic systems.

While in the McMurdo Dry Valleys organic matter in soils has multiple origins due to the proximity to the sea, aeolian redistribution, and presence of lichens, mosses, algae and Cyanobacteria [17], in the more elevated inland regions it appears to be mainly of microbial origin [53, 274, 276, 277]. As a result, organic matter is among the most limiting factors for microbial growth in soils of the Sør Rondane Mountains [38, 275]. The bacterial communities inhabiting these environments therefore mainly depend on the sequestration of atmospheric CO_2 by primary producers [67]. We hypothesize that in communities with few Cyanobacteria, other bacteria may take over as primary producers of fixed carbon and nitrogen and this should be reflected in the diversity of key genes for these functions in these communities.

Multiple pathways for carbon dioxide reduction have been discovered [67]. The most widespread and important autotrophic pathway is the Calvin-Benson-Bassham cycle [218], in which ribulose-1,5-biphosphate carboxylase/oxygenase (RuBisCO) performs a key function [74]. The highly conserved *cbb* genes – encoding the large subunit of RuBisCO – serve as a useful phylogenetic marker. Horizontal gene transfer and gene duplication have, however, resulted in some incongruence with the phylogeny of 16S ribosomal RNA (rRNA) [73]. Of the four known types of RuBisCO (I-IV), only the first two are involved in the actual fixation of carbon dioxide in Bacteria [72]. In RuBisCO type I, the dominantly found type in bacteria, the large subunit is encoded by the *cbbL* gene [278]. Type I enzymes show considerable diversification and are therefore further subdivided into four subtypes (IA-ID), of which three (IA-IC) occur in Bacteria [72]. Subtypes IA and IB are found in Cyanobacteria. Subtype IA is,

however, predominantly found in Alpha- Beta and Gammaproteobacteria, whereas, subtype IB has been found exclusively in Cyanobacteria [76]. Subtype IC, is found in Alpha-, Beta- and Gammaproteobacteria [72, 76] and was more recently reported in Thermi [81], Actinobacteria [79, 80], Verrucomicrobia [82] and Firmicutes [83]. Since the discovery of RuBisCO in the latter three bacterial groups, however, it has been suggested to group their *cbbL* sequences in a novel subtype, subtype IE [79, 80, 82]. Type II RuBisCO, the large subunit of which is encoded by the *cbbM* gene, is found in Alpha-, Beta-, Gamma- and Zetaproteobacteria [70, 84].

Biological nitrogen fixation is mediated by the nitrogenase enzyme complex [109], which is composed of two oxygen-labile metalloproteins: dinitrogenase and dinitrogenase-reductase [108]. Although various forms of the nitrogenase complex exist, the molybdenum-dependent one (encoded by *nif* genes) is the most widespread and abundant. The alternative vanadium- (v*nf* genes) and iron-dependent (*anf* genes) forms are only present in a limited number of diazotrophs and have so far always been found secondarily to the molybdenum-dependent form [108]. As a result, only key genes encoding the latter form, and especially the *nifH* gene, encoding the dinitrogenase-reductase metalloprotein, have been widely used to study the diversity of diazotrophs [108, 221].

During a limited exploratory cloning survey we investigated the presence and diversity of the two bacterial RuBisCO types in exposed terrestrial samples gathered in the vicinity of the Belgian Princess Elisabeth Station (PES), located in the Sør Rondane mountains, East-Antarctica. We selected four samples showing little visible sign of cyanobacterial biofilm, to test the hypothesis that other primary producers might be present in such oligotrophic high altitude sites. While we did not detect RuBisCO type II (*cbbM* gene), RuBisCO type I (*cbbL* gene) was found to be quite diverse in this system, with type IC, absent from Cyanobacteria, the most dominantly retrieved [277]. The diversity of *nifH* genes detected was restricted to sequences grouping with those of Nostocales Cyanobacteria. While our initial clone library set up was limited to ~100 clones/sample, in the present study, we aimed to more comprehensively assess the diversity of bacteria capable of carbon or nitrogen fixation in the same samples. We therefore used Illumina MiSeq paired-end (PE) 300 bp sequencing of *cbbL* and *nifH* genes to assess the diversity of genes involved in carbon and nitrogen fixation. In addition we sequenced 16S rRNA genes in the same samples to study the total bacterial communities present in these exposed oligotrophic soils of the Sør Rondane Mountains.

3.2. Materials and Methods

Sample collection and environmental data

To allow comparison, the same samples as used in the clone library study [277] were used. These surface samples consisting of weathered granite material were collected in the proximity of the Belgian Princess Elisabeth Station (71° 57’ S, 23° 20’ E) at Utsteinen, Dronning Maud Land, East-Antarctica, during the Antarctic summer of 2009, and were frozen at -20 °C upon collection. Four samples, showing no visual signs of macroscopic growth, were used in this study, as previously described by Tahon et al. [277] (Table 1). Samples KP15 (71° 56' 45.8" S, 23° 20' 43.6" E), KP43 (71° 56’ 47.3” S, 23° 20’ 44.6” E) and KP53 (71° 56’ 45.3” S, 23° 20’ 42.4” E) were collected on the Utsteinen ridge approximately 500 m north of the research station. Sample KP2 (71° 57' 28.6" S, 23° 19' 45.8" E) was collected 1.3 km south of the PES.

DNA extraction

Total genomic DNA was extracted and purified using the PowerLyzer® PowerSoil® DNA isolation kit (MoBio Laboratories). In a comparative analysis of several methods, this was previously identified as the DNA extraction method yielding most bacterial diversity [277]. For each sample, DNA was extracted from three homogenized 400 mg subsamples, following the manufacturer’s protocol and modifications according to Tahon et al. [277]. DNA was quantified using a Qubit® 2.0 fluorometer (Life Technologies) and stored at -20 °C until further processing.

PCR and preparation for Illumina sequencing

Primers were selected to amplify a broad diversity of taxa and to produce amplicon sizes suitable for Illumina MiSeq PE 300 bp sequencing. They were extended with an adapter to complement the Nextera XT index kit (Illumina). PCR primers used to amplify fragments of the *cbbL* (IA+IC), *nifH* and 16S rRNA genes are listed in Table 2. Primers previously used to target *cbbL* IA + IB [277] were not included because they did not produce an amplicon size suitable for Illumina MiSeq 300 bp paired-end sequencing.

For each sample, PCR for the protein encoding *cbbL* and *nifH* genes was performed in triplicate on all three DNA extracts, for each primer set, resulting in a total of nine PCR products per sample per functional gene. PCRs were carried out in 25 µl reaction mixtures containing 1x Qiagen PCR buffer (Qiagen), 0.2 mM of each deoxynucleotide triphosphate, 0.625 units of Qiagen *Taq* polymerase (Qiagen), 100 mM bovine serum albumin and forward and reverse primers (Table 2). Each reaction mixture received 3 µl of template DNA.

For amplification of the V1-V3 hypervariable regions of the 16S rRNA gene, for each sample, PCR was performed in duplicate on all three DNA extracts, resulting in six PCR products per sample. Each reaction mixture contained 2.5 µl 10x High Fidelity PCR buffer (Roche), 2.5 µl deoxynucleotide triphosphates (10 mM) (Life Technologies), 0.1 units of High Fidelity Hot Start polymerase (Roche), 0.5 µl of each primer (Table 2) and 0.5 µl of DNA template. Finally, the mixture was adjusted to a final volume of 25 µl with sterile HPLC water.

Amplification was performed using a Veriti thermal cycler (Life Technologies) following the temperature profiles shown in Table 2. Subsequent to PCR, all nine or six PCR products were pooled and purified using the Ampure beads XT (Agencourt) protocol with slight modifications: only 0.8 reaction volume of beads was used and DNA was resuspended in MilliQ water. Barcoding of pooled PCR products was performed using the Nextera XT indices (Illumina) during an eight cycle version of the amplicon PCR with the indices replacing the primers. Resulting PCR products were purified as described above, with resuspension in Tris buffer (0.1 M, pH 8.5). Integrity and amplicon sizes of the PCR products were checked using a BioAnalyzer (Agilent). DNA was quantified using a Qubit, as described above. Afterwards, samples were pooled equimolarly and sequenced on an Illumina MiSeq PE 300 bp platform (GATC).

Table 1 Parameters associated with analyzed samples

Sample	Sample Coordinates	Altitude (m)	Description of sample area	Conductivity (µS/cm)	pH	Water Content	TOC
KP2	71° 57' 28.6" S, 23° 19' 45.8" E	1320	Small gravel particles in between the rocks, Utsteinen nunatak	19	6.54	6.28 %	0.08 %
KP15	71° 56' 45.8" S, 23° 20' 43.6" E	1366	Brown soil under lichen, East part of Utsteinen ridge	33	5.57	3.38 %	0.33 %
KP43	71° 56’ 47.3” S, 23° 20’ 44.6” E	1362	Brown soil with dark green fragments, East part of Utsteinen ridge	520	6.22	0.91 %	2.57 %
KP53	71° 56’ 45.3” S, 23° 20’ 42.4” E	1362	Grey soil on East part of Utsteinen ridge	312	6.34	0.23 %	0.21 %

Table 2 PCR primers (without adapters) and conditions used for screening different genes

Gene	Target	Primer	Sequence 5’-3’	Final concentration	Region	Amplicon size	Program
cbbL	RuBisCO IA and IC	K2f[a]	ACC AYC AAG CCS AAG CTS GG	0.2 µM	496-990[e]	492-495 bp	95 °C (3 min); 35x 95 °C (1 min), 62 °C (1 min), 72 °C (1 min 30 s); 72 °C (10 min)
		V2r[a]	GCC TTC SAG CTT GCC SAC CRC	0.2 µM			
nifH	Universal	IGK3[b]	GCI WTH TAY GGI AAR GGI GGI ATH GGI AA	1.0 µM	19-413[f]	395 bp	95 °C (10 min); 40x 95 °C (45 s), 52 °C (30 s), 72 °C (40 s); 72 °C (10 min)
		DVV[b]	ATI GCR AAI CCI CCR CAI ACI ACR TC	1.0 µM			
16S rRNA	Universal	pA[c]	AGA GTT TGA TCC TGG CTC AG	0.5 µM	8-536[g]	529 bp	95 °C (3 min); 27x 95 °C (30 s), 55 °C (45 s), 72 °C (3 min); 72 °C (10 min)
		BKL1[d]	GTA TTA CCG CGG CTG CTG GCA	0.5 µM			

[a] data from [240], [b] data from [279], [c] data from [280], [d] data from [281], [e] Based on the *cbbL* IA sequence of *Bradyrhizobium* sp. ORS278 (CU234118), [f] Based on the *nifH* sequence of *Azotobacter vinelandii* (M20568), [g] Based on the 16S rRNA gene sequence of *Escherichia coli* (A14565)

Processing of Illumina MiSeq sequence data

For each gene, the forward and reverse reads were merged using the *fastq_mergepairs* command in USEARCH [282] with a maximum of six mismatches allowed in the overlapping region. After merging, forward and reverse primer sequences were removed from the sequences using cutadapt v1.8 [283]. Simultaneously, sequences that were too short or too long (*cbbL* IA/IC: < 441 bp and > 467 bp; *nifH*: < 323 bp and > 369 bp; 16S rRNA gene < 450 bp and >570 bp) were removed. During the subsequent quality filtering, by using USEARCH, all sequences with one or more nucleotides below the Phred Q20 threshold score and a maximum expected error of 0.5 were removed from further analysis. Afterwards, the protein encoding *cbbL* and *nifH* sequences were placed in reading frame +1 and those containing stop codons and/or indels resulting in a frameshift were removed. For all genes, putative chimeric sequences were detected using the UCHIME module in Mothur and removed [246, 247]. Finally, all remaining protein encoding *cbbL* and *nifH* sequences were translated to proteins using MEGA 6 [245]. For nifH, additionally the presence of the following conserved amino acid residues was checked: Ala43, the dipeptide Glu93-Pro94, Arg101, the dipeptide Ile104-Thr105 and Glu111 (positions based on the nifH protein sequence of *Azotobacter vinelandii*, accession number M20568) [162]. Deduced protein sequences lacking two or more of these conserved sites were discarded from further analysis.

Sequence analyses

Partial 16S rRNA gene sequences were grouped in operational taxonomic units (OTU) at a cut-off of 97 % nucleotide similarity, using UPARSE [284]. Taxonomy for these OTUs was inferred using the May 2013 GreenGenes training set [285-287] with the Bayesian classifier [288] implemented in Mothur [246]. Non-bacterial, chloroplast and mitochondrial OTUs were removed prior to downstream analysis.

For each protein encoding gene, updated versions of the custom-made databases described previously [277] were used. Databases contained all related sequence records from NCBI and IMG (https://img.jgi.doe.gov/) [244] as available per October 15th 2015. Newly obtained *nifH* and *cbbL* nucleotide sequences and their translated protein sequences were added to the databases using the import module of BioNumerics 7.5 (Applied Maths). For phylogenetic analyses, all Illumina cbbL and nifH protein sequences were clustered at a 95 % cut-off using CD-HIT [289, 290], grouping them into operational RuBisCO units (ORU) and operational

nitrogenase units (ONU) respectively. One representative of each ORU/ONU was used to construct a phylogenetic tree. For each gene, a first alignment was made with all sequences present in our database (nifH: 47 184 sequences; cbbL types IA/IB/IC: 8489), using Clustal Omega [248, 249] on the Stevin supercomputer at UGent. Alignments were trimmed to the size of our sequenced amplicons and visually inspected, excluding from further analysis all non-overlapping sequences. Remaining sequences were realigned, after which the alignment was used to construct a maximum likelihood phylogenetic tree (1000 bootstrap replicates), using the FastTree tree building software [250] with the Whelan and Goldman evolutionary model and the discrete gamma model with 20 rate categories. From the resulting phylogenetic trees, closest relatives of our newly obtained sequences as well as representative sequences from the entire tree were selected to prepare a smaller tree representing the initial complete tree, following the same protocol. Sequences originating from uncultured bacteria were not included in the final tree. Trees were visualized using the iTOL software [251, 252]. Related ORUs or ONUs were grouped into visual clusters that were named after cultivated bacteria that grouped in the cluster or, in the absence of cultivated members, were given a Utsteinen (UT) cluster number designation.

Statistical analyses

The Vegan package [253] in R (http://cran.r-project.org) was used for statistical analyses of the Illumina 16S rRNA nucleotide sequences, and the cbbL and nifH. For each gene, an OTU/ORU/ONU table was constructed, showing the total number of nucleotide sequences per OTU/ORU/ONU. Using this table, the total number of expected OTUs, ORUs or ONUs per sample and for the four samples combined was determined by rarefaction analysis. Different parameters were calculated for each dataset, including species richness (Chao1), Bray-Curtis dissimilarity and evenness (Pielou).

Availability of data and materials

Raw *cbbL*, *nifH* and 16S rRNA gene sequences are available from the NCBI sequence read archive (http://www.ncbi.nlm.nih.gov/sra) under accession number SRP067116.

3.3. Results

After Illumina sequencing and thorough quality control of all sequence data, 3204 high-quality sequences of 441-453 bp length and 12 698 high-quality sequences (length 323-369 bp) were left in the *cbbL* and *nifH* datasets, respectively (Table 3). For the 16S rRNA gene, 21 663 sequences were obtained (Table 3).

Table 3 Overview of clone library characteristics. Diversity indices were calculated on the basis of derived unique protein sequences

			Illumina data				
			KP2	**KP15**	**KP43**	**KP53**	**All**
cbbL	No. of sequences		1127	444	1298	335	3204
	No. of ORUs		135	74	50	55	277
	evenness (H/H_{max})		0.650	0.707	0.459	0.614	0.620
	Chao1		325.59	132.57	125.60	121.00	587.08
	Bray-Curtis	KP15	0.968				
		KP43	0.876	0.998			
		KP53	0.773	0.949	0.793		
nifH	No. of sequences		1247	11 324	85	42	12 698
	No. of ONUs		19	45	6	6	47
	evenness (H/H_{max})		0.251	0.368	0.253	0.531	0.356
	Chao1		23.67	50.63	7.50	7.50	54.86
	Bray-Curtis	KP15	0.802				
		KP43	0.988	0.985			
		KP53	0.983	0.993	0.465		
16S rRNA	No. of sequences		6264	4857	2827	7715	21 663
	No. of OTUs		333	246	211	275	703
	evenness (H/H_{max})		0.714	0.625	0.729	0.688	0.708
	Chao1		408.68	359.17	325.11	360.80	885.42
	Bray-Curtis	KP15	0.793				
		KP43	0.756	0.863			
		KP53	0.748	0.667			

For *cbbL*, all deduced protein sequences constituted 277 ORUs (95 % similarity), with the number of ORUs per sample varying from 50 to 135 (Table 3). Approximately 57 % of all ORUs (158 ORUs) only contained one sequence (Table S1). No ORU was found to be present in all four samples (Fig. 1). Six ORUs and 25 ORUs were detected in three and two of the samples, respectively, and 246 ORUs were recovered from one of the samples (Fig. 1). Only four ORUs (ORUs C9, C11, C12, C15) contained more than 5 % each of the total number of sequences (Table S1) and together made up more than 50 % of the cbbL sequences recovered. Of these, three (ORUs C9, C11, C15) were found to occur in three of the samples. ORUs C9 (17.45 % of sequences) and C11 (10.30 % of sequences) both had cbbL IC sequences of

Mycobacterium species as a closest match (~88 % amino acid (AA) similarity). ORU C15 showed 97.1 % AA similarity to the cbbL IC sequence originating from *Nitrosospira multiformis* ATCC 25196.

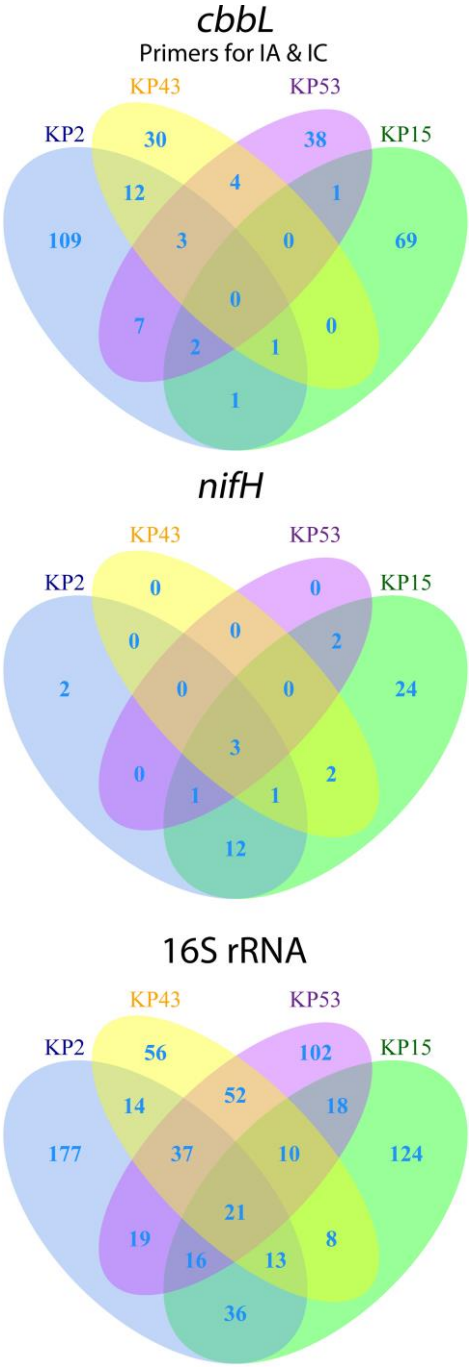

Fig. 1 Unweighted Venn diagrams of ORUs/ONUs/OTUs. Venn diagrams were calculated using the VennDiagram package (http://cran.r-project.org/web/packages/VennDiagram/index.html) in R. Colored areas show the number of ORUs/ONUs/OTUs present in the samples or shared between multiple samples

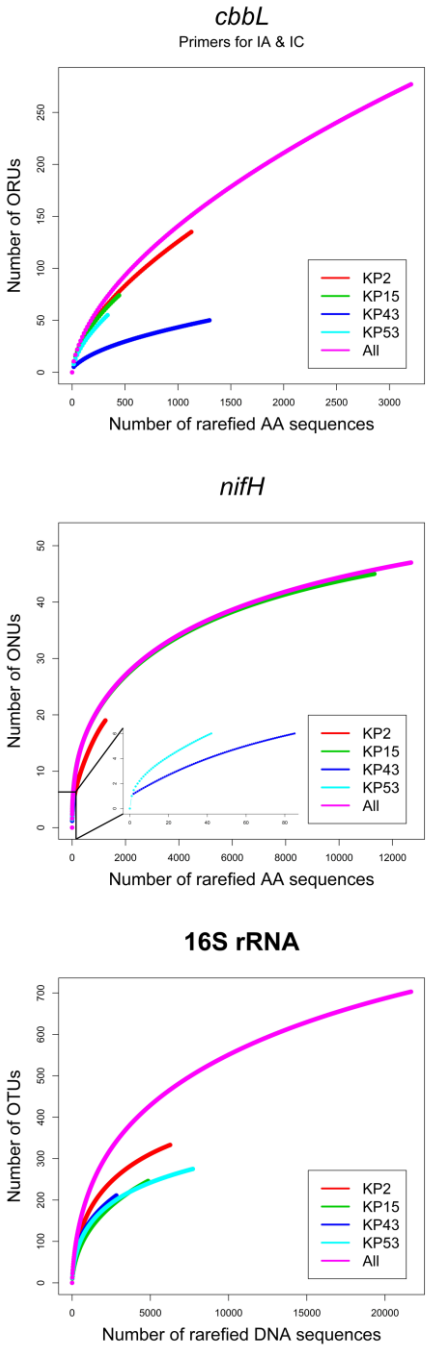

Fig. 2 Rarefaction curves of *cbbL*, *nifH* and 16S rRNA sequences. Rarefaction curves are based on grouping sequences that share 95 % similarity on protein level (*cbbL* and *nifH*) or 97 % nucleotide sequence similarity (16S rRNA). Analysis was performed using the Vegan package in R

Rarefaction analysis indicated that saturation was not yet reached (Fig. 2). For all samples, individually or taken together, the graph was still rising, indicating that the 3204 sequences, representing 277 ORUs, did not cover all the cbbL IA/IC diversity present. This was corroborated by the Chao1 values that were much higher than the number of observed ORUs (Table 3). Bray-Curtis dissimilarity analysis showed that most samples had little to nearly no overlap (Table 3). Overall, the distribution in all samples was quite even, except for sample KP43, which had a Pielou evenness index of only 0.459, even though it had most reads, grouping into the least number of ORUs (Table 3).

After maximum likelihood analysis, most ORUs could be grouped into 10 clusters, seven of which grouped with cbbL sequences from known bacteria. A total of 11 ORUs were separate. Of these, five grouped with a cbbL sequence from a cultured bacterium (Fig. 3, Table 4). Sample KP2 proved to be the most diverse sample, its sequences belonging to nearly all clusters and separate ORUs. All other samples were much less diverse, with sequences grouping mainly in the largest clusters (Tables 4 and S1). Type IC sequences were dominantly recovered from the samples, making up 96.79 % of all cbbL sequences. All newly obtained sequences of type IA grouped in a single cluster with that of *Bradyrhizobium* sp. BTAi1 (Fig. 3) and most of them originated from sample KP15 (99 sequences) and to a much lesser extent from KP2 (four sequences) (Table 4). ORUs grouping with cbbL type IC were scattered among the known type IC phylogeny, grouping with the cbbL sequences obtained from various cultured bacteria (Fig. 3, Table 4). Most of these ORUs, together representing most of the sequence data, grouped together in three of the clusters. These showed similarity to the cbbL IC sequences of the genus *Mycobacterium* (30.56 % of sequence data), *Solirubrobacter soli*/*Nocardia* species (22.28 % of sequence data) and *Nitrosospira multiformis* ATCC 25196 (18.32 % of sequence data) (Fig. 3, Table 4). Sequences grouping with the latter originated mostly from samples KP2 and KP43, whereas only one and two sequences – belonging to ORUs C15 and C35 – were obtained from samples KP53 and KP15, respectively (Table 4). All other clusters showing affiliation to cbbL IC sequences from cultivated bacteria contained only a small number of ORUs and sequences (Table 4). In addition, clusters UT 1 and UT 3 did not group closely with any known cbbL sequence, although they did contain 6.40 % and 16.51 % of the nucleotide sequences and 8.30 % and 15.16 % of the ORUs, respectively (Table 4, Fig. 3). Both mainly contained sequences from sample KP2 (Table 4).

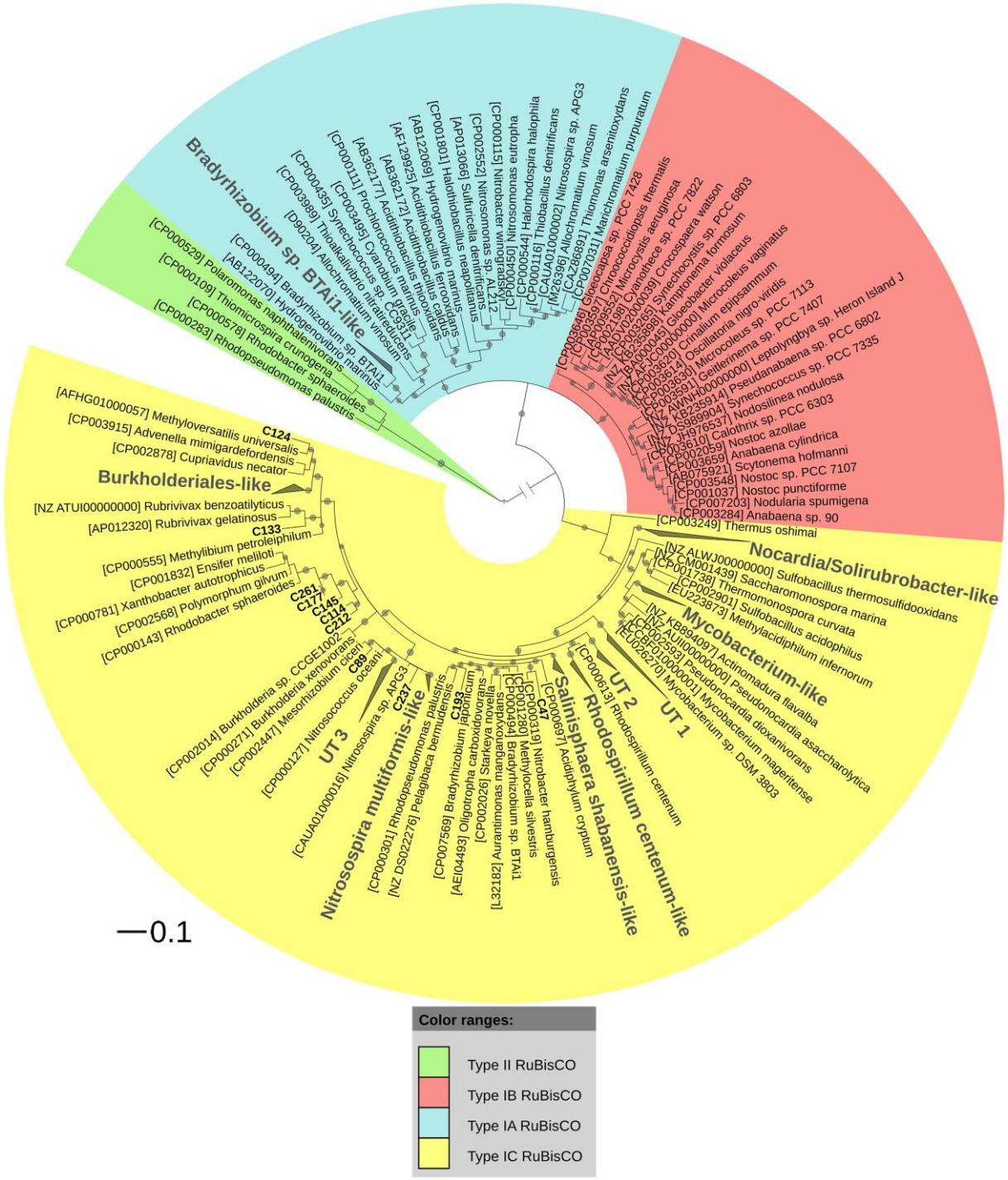

Fig. 3 Maximum likelihood phylogenetic tree of cbbL IA, IB and IC protein sequences. Maximum likelihood phylogenetic tree (1000 bootstrap replicates) of cbbL IA, IB and IC protein sequences (155 AA). Scale bar indicates 0.1 substitutions per amino acid position. Related IA and IC ORUs were grouped into clusters (displayed as gray triangles) that were named after cultivated bacteria that grouped in the cluster or, in the absence of cultivated members, were given a Utsteinen (UT) cluster number designation. Total branch lengths to the closest and the farthest leaf are used as sides of the triangle. Separate ORUs are labelled in bold. For reference data, accession number and taxon name is listed. Bootstrap values are displayed as circles. Smallest circles represent the lower cut-off of 70 %. *cbbM* sequences (RuBisCO type II) were used as an outgroup.

Finally, comparison of all 2316 unique *cbbL* nucleotide sequences with the previously obtained 203 unique clone sequences [277], revealed an overlap of only 37 sequences (data not shown). However, grouping at 95 % protein similarity (*i.e.* the threshold for our ORUs) revealed that all but six clone sequences were not recovered using the deep sequencing approach (data not shown).

At a 95 % similarity binning, the sequences in the nifH dataset could be grouped into 47 ONUs (Table 3, Table S1), all related to nifH sequences of Nostocales Cyanobacteria (88 to 95 % AA similarity). Most sequences originated from samples KP15 (11 324 sequences) and KP2 (1247 sequences), while samples KP43 and KP53 only yielded 85 and 42 sequences, respectively (Table 3).

Table 4 Characteristics of different clusters or separate ORUs

Cluster/Separate ORU	No. of ORUs	ORUs (%)	No. of reads	Reads (%)
Nocardia/*Solirubrobacter*-like	126 (37, 55, 14, 30)	45.49 %	714 (145, 315, 139, 115)	22.28 %
UT 3	42 (34, 5, 6, 6)	15.16 %	529 (466, 25, 10, 28)	16.51 %
Mycobacterium-like	34 (13, 0, 15, 16)	12.27 %	979 (150, 0, 640, 189)	30.56 %
UT 1	23 (23, 0, 1, 0)	8.30 %	205 (204, 0, 1, 0)	6.40 %
Nitrosospira multiformis-like	12 (5, 1, 8, 1)	4.33 %	587 (104, 2, 480, 1)	18.32 %
Bradyrhizobium sp. BTAi1-like*	11 (2, 10, 0, 0)	3.97 %	103 (4, 99, 0, 0)	3.21 %
Burkholderiales-like	7 (7, 0, 1, 0)	2.53 %	11 (10, 0, 1, 0)	0.34 %
Rhodospirillum centenum-like	5 (0, 1, 3, 2)	1.81 %	22 (0, 1, 19, 2)	0.69 %
UT 2	4 (4, 0, 0, 0)	1.44 %	8 (8, 0, 0, 0)	0.25 %
Salinisphaera shabanensis-like	2 (2, 0, 1, 0)	0.72 %	17 (10, 0, 7, 0)	0.53 %
Mesorhizobium ciceri-like	1 (1, 0, 0, 0)	0.36 %	16 (16, 0, 0, 0)	0.50 %
Rubrivivax-like	1 (1, 0, 0, 0)	0.36 %	2 (2, 0, 0, 0)	0.062 %
Methyloversatilis universalis-like	1 (0, 1, 0, 0)	0.36 %	1 (0, 1, 0, 0)	0.031 %
Nitrobacter hamburgensis-like	1 (1, 0, 0, 0)	0.36 %	1 (1, 0, 0, 0)	0.031 %
Nitrosospira sp. APG3-like	1 (0, 1, 0, 0)	0.36 %	1 (0, 1, 0, 0)	0.031 %

Type IA cbbL clusters are indicated with an asterisk. Separate ORUs without any close known cbbL sequence are not included. Numbers between brackets indicate the number of ORUs or reads found in the different samples (KP2, KP15, KP43, KP53)

The number of ONUs per sample varied from six, for samples KP43 and KP53, to a maximum of 45, for KP15 (Table 3). Of the 47 ONUs, three were found to be common to all samples (Fig. 1). The most abundant ONU (ONU C3), however, was not detected in sample KP53 (Table S1). A total of 10 ONUs was classified as singleton (Table S1) and 26 ONUs contained sequences obtained from a single sample (Fig. 1).

Rarefaction analysis showed that for the single as well as the pooled samples, saturation was not reached (Fig. 2). Indeed, the estimated number of ONUs (Chao1) was higher than the

number of observed ONUs for both the single and the pooled samples although the differences were relatively small, indicating that most diversity was retrieved from the samples (Table 3). Bray-Curtis dissimilarity analysis revealed that only between samples KP43 and KP53 more similarity existed while for all other relationships the dissimilarity values were very high (Table 3). Finally, the samples' *nifH* communities were very unevenly distributed, as the evenness indices (Pielou) for all samples were very low (Table 3).

All 21 663 sequences of the V1-V3 region of the 16S rRNA gene grouped into 703 OTUs at 97 % cut-off (Table 3). The number of OTUs per sample varied from 211 for sample KP43 to 246, 275 and 333 for samples KP15, KP53 and KP43, respectively (Table 3). A total of 21 OTUs was found to be present in all the samples, whereas 76 and 147 OTUs were detected in three and two of the samples, respectively (Fig. 1). Rarefaction analysis showed saturation was not reached (Fig. 2), indicating more diversity is present in the samples, which is confirmed by the high number of expected OTUs (Chao1, Table 3). Finally, Bray-Curtis analysis revealed only limited overlap between the four samples, whereas the Pielou evenness indices showed a quite uneven distribution of the sequences in the OTUs (Table 3).

Based on the taxonomy inferred using the GreenGenes training set, the 703 OTUs could be assigned to 16 phyla, 44 classes, 55 orders, 69 families and 55 genera (Tables 5 and S1). A total of six (13 sequences) and 539 (16 447 sequences) OTUs remained unclassified at phylum (Table 5) and genus (Table S1) level, respectively. In total, Acidobacteria (32.63 % of sequences) were most abundantly recovered, followed by Actinobacteria (22.98 % of sequences), Cyanobacteria (17.44 % of sequences) and Proteobacteria (12.17 % of sequences) (Table 5). Most of the sequences (95.74 %) belonging to the Acidobacteria were assigned to the family-level taxon Ellin6075 [291] in the class Chloracidobacteria (Table S1). Of the Actinobacteria, 25.65 % of all sequences could be assigned to the family Nocardioidaceae in the order Actinomycetales (Table S1). Most proteobacterial sequences grouped with the alphaproteobacterial orders of the Rhizobiales (17.15 %), Rhodospirillales (35.81 %) and Sphingomonadales (38.89 %) (Table S1). Sequences assigned to Cyanobacteria grouped mainly in the families Nostocaceae (20.30 %), Xenococcaceae (13.71 %) and Phormidiaceae (49.05 %).

In the individual samples, Acidobacteria and Actinobacteria remained the most frequently recovered phyla based on the number of sequences (Fig. 4). Only for sample KP15 this was not the case as 52.15 % of the recovered reads grouped with Cyanobacteria (Fig. 4). Other phyla with a relative abundance of more than 1 % were Chloroflexi, the candidate division

FBP, Bacteroidetes and Thermi (Table S1). Whereas Cyanobacteria comprised only a small number of all sequences in samples KP43 and KP53 (0.04 and 0.12 % of sequences), they were among the most abundantly recovered phyla in samples KP2 and KP15 (19.72 and 52.15 % of sequences) (Fig. 4, Table 5).

Looking at OTUs, however, Actinobacteria and Proteobacteria were most frequently recovered from all samples. Whereas in samples KP2 and KP15, a high percentage of sequences were assigned to Cyanobacteria, the phylum constitutes only between 3.60 and 6.50 % of the OTUs, respectively.

It should be noted that chloroplast 16S rRNA gene sequences were retrieved from all samples (Table S1). For samples KP2 and KP15, these were only 13 and 10 sequences, respectively. However, for samples KP43 and KP53, a total of 111 and 282 chloroplast sequences were retrieved. Most of these sequences grouped into one OTU at 97 % binning (OTU 137) and could be assigned to the Trebouxiophyceae. Although the GreenGenes training set could not assign the OTU to a lower taxon, an additional BLAST identification (http://blast.ncbi.nlm.nih.gov/Blast.cgi) grouped this OTU with the genus *Trebouxia*.

To compare the bacterial diversity retrieved with different DNA extraction protocols, 16S rRNA data of three of the samples investigated here (KP2, KP15 and KP43) was compared with data obtained by Tytgat et al. [275]. In their study a different DNA extraction protocol was used. Briefly, extracellular DNA was removed in a first step following the protocol of Corinaldesi et al. [292]. Afterwards, DNA was extracted according to Zwart et al. [293] using a lysis step (3x 1.25 min bead beating with zirconium beads) followed by phenol:chloroform:isoamylalcohol extraction. Finally the DNA was cleaned using the Wizard® DNA clean-up system (Promega). The two main differences between the two extraction protocols are 1) the removal of extracellular DNA in the protocol used by Tytgat et al. [275], a step that was not performed using the PowerLyzer® PowerSoil® DNA isolation kit (MoBio Laboratories) and 2) the use of a shorter bead beating step using different beads by Tytgat et al. [275].

To compare the results obtained using these different extraction methods, taxonomy analysis was performed together with the reads obtained by Tytgat et al. [275]. It revealed 422 shared OTUs, while 220 and 179 OTUs, were retrieved only with the extraction protocol of Tytgat et al. [275] or the PowerLyzer® PowerSoil® kit, respectively (data not shown). Although OTU distribution was mostly similar, using our extraction protocol a clear increase in OTUs assigned to Actinobacteria and Chloroflexi could be observed, whereas Bacteroidetes showed

a clear decrease (Fig. S1, Table S2). Furthermore, the distribution of reads showed remarkable differences (Fig. S1, Table S2). With the extraction protocol used by Tytgat et al. [275] 64.40 % of the reads were assigned to Acidobacteria compared to 27.34 % for the PowerLyzer® PowerSoil® DNA isolation kit. Actinobacterial, cyanobacterial and Chloroflexi reads rose from 3.50 to 16.99 %, 2.37 to 27.02 % and 0.96 to 7.24 % respectively, whereas reads assigned to Bacteroidetes dropped from 4.97 to 2.03 % (Fig. S1, Table S2). Similar observations could be seen for the separate samples.

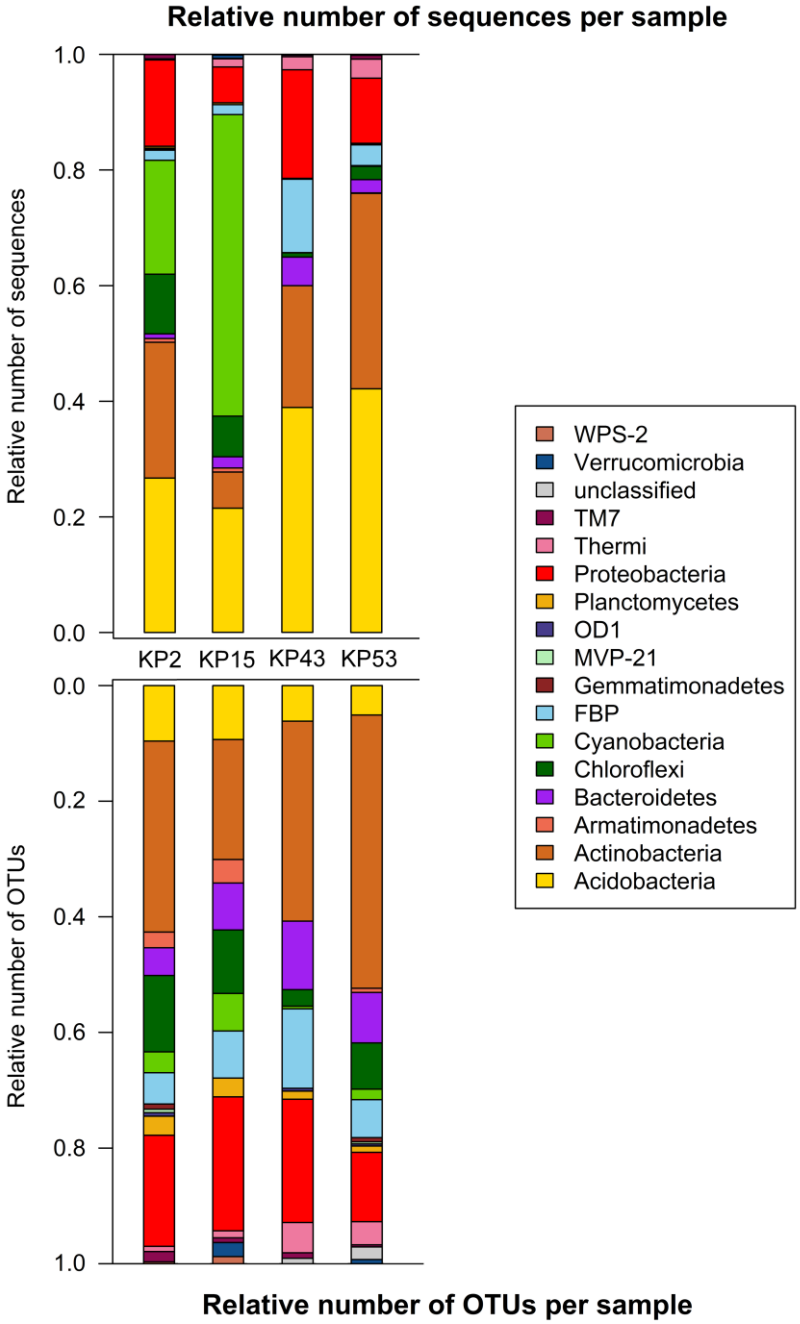

Fig. 4 Relative abundances of phyla based on number of 16S rRNA gene sequence reads (top) and OTUs (bottom)

Table 5 Distribution of recovered 16S rDNA OTUs per phylum and class in each of the samples. The number of reads per class is given between brackets

Phylum	Class	All	KP2	KP15	KP43	KP53
Acidobacteria	Acidobacteriia	2 (187)	0	2 (187)	0	0
	Chloracidobacteria	31 (6805)	27 (1657)	15 (797)	13 (1100)	13 (3251)
	Solibacteres	8 (77)	5 (16)	6 (59)	0	1 (2)
Actinobacteria	Acidimicrobiia	29 (405)	15 (170)	7 (60)	9 (30)	15 (145)
	Actinobacteria	154 (4036)	74 (1208)	35 (191)	58 (556)	96 (2081)
	Nitriliruptoria	1 (4)	0	0	1 (1)	1 (3)
	Rubrobacteria	4 (18)	0	0	2 (4)	3 (14)
	Thermoleophilia	27 (492)	16 (78)	8 (52)	2 (3)	13 (359)
	Unclassified	8 (23)	5 (15)	1 (1)	1 (1)	2 (6)
Armatimonadetes	0319-6E2	1 (2)	1 (2)	0	0	0
	Armatimonadia	14 (75)	7 (39)	7 (31)	0	2 (5)
	Chthonomonadetes	1 (1)	0	1 (1)	0	0
	Fimbriimonadia	1 (1)	0	1 (1)	0	0
	OPB50	1 (1)	1 (1)	0	0	0
	Unclassified	1 (1)	0	1 (1)	0	0
Bacteroidetes	Cytophagia	35 (291)	6 (22)	12 (77)	16 (98)	10 (94)
	Flavobacteriia	2 (5)	0	2 (5)	0	0
	Rhodothermi	3 (3)	2 (2)	0	1 (1)	0
	Saprospirae	25 (149)	8 (25)	6 (12)	7 (39)	12 (73)
	Sphingobacteriia	2 (13)	0	0	1 (2)	2 (11)
Chloroflexi	Anaerolineae	4 (24)	2 (17)	3 (7)	0	0
	C0119	6 (155)	5 (145)	2 (10)	0	0
	Chloroflexi	8 (75)	6 (31)	2 (9)	2 (14)	2 (21)
	Ellin6529	12 (445)	12 (394)	1 (1)	1 (1)	3 (49)
	Ktedonobacteria	7 (248)	0	6 (241)	0	1 (7)
	Thermomicrobia	32 (173)	11 (18)	9 (51)	2 (5)	14 (99)
	TK10	13 (65)	7 (40)	4 (23)	0	2 (2)
	Unclassified	2 (3)	1 (2)	0	1 (1)	0
Cyanobacteria	ML635J-21	2 (3)	1 (1)	1 (2)	0	0
	Nostocophycideae	4 (767)	1 (126)	4 (640)	1 (1)	0
	Oscillatoriophycideae	7 (2374)	3 (501)	6 (1866)	0	3 (7)
	Synechococcophycideae	6 (117)	2 (94)	4 (21)	0	2 (2)
	Unclassified	5 (517)	5 (513)	1 (4)	0	0
FBP	Unclassified	48 (828)	18 (109)	20 (84)	29 (359)	18 (276)
Gemmatimonadetes	Gemm-1	1 (6)	1 (6)	0	0	0
	Gemmatimonadetes	4 (9)	2 (3)	0	0	2 (6)
MVP-21	Unclassified	3 (13)	2 (4)	0	0	1 (9)
OD1	SM2F11	4 (13)	2 (9)	0	1 (1)	1 (3)
Planctomycetes	Phycisphaerae	10 (20)	6 (10)	2 (7)	3 (3)	0
	Planctomycetia	13 (22)	5 (10)	6 (8)	0	3 (4)
Proteobacteria	Alphaproteobacteria	98 (2569)	53 (908)	45 (285)	40 (516)	29 (860)
	Betaproteobacteria	14 (30)	6 (8)	9 (12)	2 (7)	2 (3)
	Gammaproteobacteria	3 (8)	2 (6)	0	0	1 (2)
	Deltaproteobacteria	3 (6)	1 (2)	1 (1)	2 (3)	0
	TA18	1 (1)	1 (1)	0	0	0
	Unclassified	2 (22)	1 (11)	2 (3)	1 (5)	1 (3)
Thermi	Deinococci	16 (398)	3 (11)	3 (68)	11 (64)	11 (255)
TM7	SC3	1 (2)	0	1 (2)	0	0
	TM7-1	4 (70)	3 (14)	1 (2)	1 (4)	1 (50)
	TM7-3	2 (30)	1 (26)	0	1 (4)	0
	Unclassified	2 (8)	2 (8)	0	0	0
Unclassified	Unclassified	6 (13)	0	0	2 (4)	6 (9)
Verrucomicrobia	Spartobacteria	6 (32)	0	6 (28)	0	2 (4)
WPS-2	Unclassified	4 (8)	1 (1)	3 (7)	0	0

3.4. Discussion

In this study we investigated the hypothesis that in the exposed soils around the Princess Elisabeth Station in Dronning Maud Land, in the visible absence of Cyanobacteria, other bacteria might take over as main primary producers of fixed carbon and nitrogen. We studied the diversity of RuBisCO and nitrogenase genes as proxies for the bacterial diversity capable of taking over these functions. While an initial clone library survey had hinted at considerable diversity [13], it was limited in depth (100 clones/sample). Here we used Illumina amplicon sequencing on the same samples to extend these findings and provide more comprehensive coverage of the diversity present. In addition we studied the taxonomic marker gene for 16S rRNA to document the taxonomic community composition.

We used the same broad-range primer set – amplifying *cbbL* types IA and IC – as selected before [277] for the Illumina sequencing. As in the pilot study [277], the cbbL type IC was most dominantly recovered from all samples (96.79 % of sequences) (Table 4) and sequences were dispersed over the type IC phylogeny (Fig. 3). This type, found in Alpha-, Beta- and Gammaproteobacteria, Actinobacteria, Verrucomicrobia, Thermi and Firmicutes, has not yet been retrieved from Cyanobacteria [72, 76, 79-83]. Sequences from all samples mainly grouped with Actinobacteria and to lesser extent with Alphaproteobacteria, Betaproteobacteria and Gammaproteobacteria, implying the presence of several non-cyanobacterial taxa that may be able to assimilate CO_2 via the Calvin-Benson-Bassham cycle. Moreover, the presence of cbbL sequences similar to those of *Mesorhizobium* in KP2 and *Nitrosospira* in KP2 and KP43 confirms our findings of the pilot study [277]. Similar type IC cbbL sequences were found in samples taken in the subglacial environment of the Robertson Glacier, Canada [262]. Although it is not possible to directly link cbbL sequences to 16S rRNA OTUs, it is notable that the 16S rRNA results tentatively corroborate the presence of several of the cbbL IC groups recovered from our samples. Taxa grouping with the dominantly recovered *Nocardia/Solirubrobacter*-like cbbL type (Table 4) are among the most represented Actinobacteria in the 16S rRNA gene dataset (Table S1).

On the other hand, some taxa, although frequently retrieved from the cbbL dataset, were not retrieved by sequencing of the 16S rRNA genes. For example, *Mycobacterium* was not detected among 16S rRNA OTUs, whereas 34 ORUs, comprising 30.56 % of the *cbbL* reads, grouped with this genus. This does not necessarily imply the absence of *Mycobacterium* from

our samples, as rarefaction analysis (Fig. 2) indicates more diversity is present in the investigated samples.

As in the first survey [277], cbbL type IA was again only recovered in very small numbers, mainly from KP15, and with a very low diversity (11 ORUs) (Table 4). The similarity to the cbbL IA sequence of *Bradyrhizobium* sp. BTAi1 also corroborates the 16S rRNA gene Illumina data where OTU 56 strongly hints at the presence of this alphaproteobacterial genus in our samples (Table S1). Furthermore, a sevenfold increase in 16S rRNA reads from this OTU could be observed between samples KP2 and KP15. This correlates with a high increase of *Bradyrhizobium*-related cbbL sequences observed here (Table 4). Similar cbbL IA sequences were also retrieved from an Antarctic moss pillar [87] and low latitude subarctic subglacial sediment [262]. This suggests that this organism, originally described from stem nodules of *Aeschynomene indica* in Africa and southeast North America [294], has a wide temperature range. Finally, the potential of *Bradyrhizobium* sp. BTAi1 and relatives to synthesize bacteriochlorophyll *a* [295] may provide advantages in the oligotrophic Antarctic environment.

In the pilot survey [277], a high abundance of cbbL type IB, grouping with sequences from Chroococcales and Oscillatoriales Cyanobacteria was recorded in samples KP2 and KP15, respectively. Here, in the Illumina data, indeed nearly all 16S rRNA reads assigned to those cyanobacterial orders originated from the same samples. Samples KP43 and KP53, on the other hand, showed a very limited cyanobacterial 16S rRNA gene diversity, in accordance with the very low retrieval of cbbL type IB [277]. In general, the latter two samples also displayed a low non-cyanobacterial cbbL diversity, suggesting a scarcity of bacterial primary producers. However, the retrieval of 16S rRNA gene sequences grouping with Trebouxiophyceae chloroplast 16S rRNA mainly from KP43 and KP53 suggests these microalgae may be important primary producers here. Indeed, this diverse class of mainly photosynthetic green algae displays a remarkable variation in ecology and has previously been found in Antarctica [296-298]. Furthermore, the unicellular green algae *Trebouxia*, present in most lichens, also occurs in a free-living state [298, 299].

Although our newly obtained cbbL sequences cannot be directly linked to those of cultivated bacteria, the broad variety of cbbL IA/IC sequences retrieved suggests that multiple non-cyanobacterial bacteria may indeed play a role in primary production. Furthermore, the presence of chloroplast 16S rRNA in samples with low numbers of Cyanobacteria also suggests that, instead of bacterial autotrophs, green algae may take over the cyanobacterial

role in these samples. The originally used type IB *cbbL* primers did, however, not target the plant and green algae *cbbL* IB form [239, 277]. Therefore, future research using broader *cbbL* primers or metagenome analyses should reveal whether green algae indeed play an important role in carbon dioxide fixation in this extreme environment.

To investigate the influence of bacteria on a key step in the nitrogen cycle, the diversity of *nifH* genes, involved in nitrogen fixation, was assessed. The PolF/R primer set [243], used for the clone libraries, was replaced as it was shown to only encompass ~25 % of the *nifH* diversity in the database composed by Gaby and Buckley [111]. To retrieve more diversity from our soil samples, the primer set IGK3/DVV, targeting more than 90 % of the *nifH* sequences *in silico* [111] was used. This primer set was reported to amplify PCR products of the expected size from all soil DNA samples and nitrogen-fixing strains tested and did not show amplification in the negative controls [111]. Where the previously used primer set designed by Poly and colleagues [243] failed to amplify *nifH* from one of the samples during the pilot study [277], the more universal IGK3/DVV primer set succeeded. However, the more universal primer set proved to be less specific. It mostly amplified DNA fragments of the same size which are not involved in nitrogen fixation, as they lacked several of the conserved amino acid residues typical of nifH [162], hence reducing *nifH* sequencing depth. Similar to the clone library data [277], all nifH sequences obtained by Illumina sequencing showed a close relationship to sequences belonging to Nostocales Cyanobacteria and in particular *Nostoc*, an organism that thrives under the extreme Antarctic conditions [106, 125, 128, 216, 265]. Presence of *Nostoc* and Nostocales deduced from nifH sequences is corroborated by the 16S rRNA Illumina data, where four of the OTUs (OTUs 4, 7, 11 and 1501) indicate presence of Nostocales, and in particular *Nostoc* (OTU 11), in our samples (Table S1). Moreover, the number of reads of these OTUs in each of the four samples correlates with the number of nifH reads from these samples. Sample KP15 shows a particular richness in nifH and cyanobacterial 16S rRNA variations, compared to the other samples. An explanation may be that this sample was originally taken underneath lichen growth (Table 1). Although 85 % of lichens contain green algal photobionts, another 10 % have Cyanobacteria as the main or only photosynthetic partner [300]. The extremely low retrieval of chloroplast 16S rRNA from KP15 suggests the lichen associated with our sample belongs to the latter form. Therefore, the presence of lichen might have contributed to a local enrichment of Cyanobacteria, resulting in the high 16S rRNA and nifH richness recovered.

Given the documented broad range of the nifH primers used [111] here, the limited *nifH* sequencing depth from samples KP43 and KP53 may be a result of the absence of more *nifH* diversity in our samples. Indeed, results of 16S rRNA gene sequencing based on two different DNA extraction protocols displayed an extremely limited number of cyanobacterial reads, indicating their near absence in these samples (Tables 5 and S2).

The higher number of sequences combined with the use of a more universal primer set did not lead to the recovery of non-cyanobacterial nifH sequences, supporting the previously made suggestion that non-cyanobacterial taxa may contribute little to nitrogen fixation in these oligotrophic high-altitude terrestrial Antarctic environments [277]. However, although Cyanobacteria appear to be the main nitrogen fixing group, rarefaction analysis indicated not all nifH diversity has yet been recovered (Fig. 2). Indeed, the 16S rRNA gene data (Table S1) also support the presence of several other bacterial groups with nitrogen-fixing potential, with OTUs assigned to the genera *Arthrobacter* and *Nocardioides*, and the families Bradyrhizobiaceae and Frankiaceae. Although *in silico* analysis, using our custom-made database, revealed that the primer pair IGK3/DVV indeed targets these taxa (data not shown), we did not detect related nifH sequences. Previous studies have reported the presence of multiple proteobacterial *nifH* genes in the McMurdo Dry Valleys [127], and diazotrophic bacteria have been recovered from fuel-contaminated Antarctic soils [301]. Possible reasons for the absence of non-cyanobacterial *nifH* sequences may be a very low abundance of these *nifH* types in our samples or primer limitations, as the selected primer pair does not target 100 % of the currently known *nifH* diversity [111]. Furthermore, the less abundant vanadium- and iron-dependent nitrogenases, where the dinitrogenase-reductase metalloprotein is encoded by *vnfH* and *anfH* genes [108], respectively, also remain un-retrieved using the current approach, although to date, these alternative nitrogenases have always been found secondarily to the molybdenum-dependent form [108]. Finally, because of the oxygen-sensitivity of nitrogenases, the use of top surface soil samples in direct contact with the Earth's atmosphere may have restricted the *nifH* diversity present [108].

Although it might be hypothesized that the long continental isolation and unique Antarctic environmental conditions might lead to a selection of Antarctic gene types, no particular psychrophilic associations could be detected for cbbL and nifH sequences retrieved here. All the most closely related sequences originated from samples taken all over the world (Table S3). For cbbL types IA and IC, environmental sequences from a variety of Chinese soil samples grouped closely with ours, although sequences from other ecosystems were also

present. It should be noted, however, that the grouping of Antarctic cbbL with mainly Chinese cbbL sequences may result from database limitations. To date, most reference data originates from these Chinese soil samples, possibly biasing the associations based on habitat metadata and new research with samples from a variety of other ecosystems should, in time, provide new insights from more comprehensive comparisons.

Analysis of the 16S rRNA gene sequencing data revealed the main phyla retrieved here were generally the same as those found in other studies focusing on terrestrial Antarctic samples, [23, 216, 302-305]. Furthermore, similar to a previous report from Victoria Land [303], the most abundantly retrieved phylum Acidobacteria displayed a very low diversity with only 41 OTUs (Table 5).

Finally, it is worth noting that the DNA extraction protocol can greatly influence sequencing results. Three of the samples investigated here were previously analyzed in a study by Tytgat et al. [275] who investigated the bacterial community composition of terrestrial samples from the Sør Rondane Mountains. The relative abundances of both OTUs and reads showed remarkable differences for several phyla (*e.g.* Acidobacteria, Bacteroidetes) (Fig. S1, Table S2). The large differences between these abundances clearly show the influence of the DNA extraction protocol on the results and call for caution when comparing bacterial community compositions from different studies.

3.5. Conclusion

The results presented here are among the first to report on genes involved in primary production and diazotrophy in a terrestrial Antarctic environment, using deep sequencing. They call for a shift in the idea that only Cyanobacteria are the principal primary producers in Antarctic soils, to include non-cyanobacteria as potential contributors, which can modify how biogeochemical cycling is modeled in this extreme environment. Deep sequencing revealed mainly type IC actinobacterial and proteobacterial cbbL, confirming the presence of multiple non-cyanobacterial microorganisms that may potentially contribute to primary production in these exposed high-altitude oligotrophic terrestrial Antarctic bacterial communities. Based on the *nifH* primer set used here, nitrogen fixation on the other hand, seems restricted to Nostocales Cyanobacteria. Finally, sequencing of 16S rRNA genes revealed a broad variety of bacterial phyla with greatly varying abundances between the separate samples.

3.6. Acknowledgements

This work was supported by the Fund for Scientific Research – Flanders [project G.0146.12]. Additional support was provided by the Belgian Science Policy Office [project CCAMBIO]. The computational resources [Stevin supercomputer Infrastructure] and services used in this work were provided by the Flemish Supercomputer Center (VSC) funded by Ghent University, the Hercules Foundation and the Flemish Government – department EWI.

3.7. Supplementary information

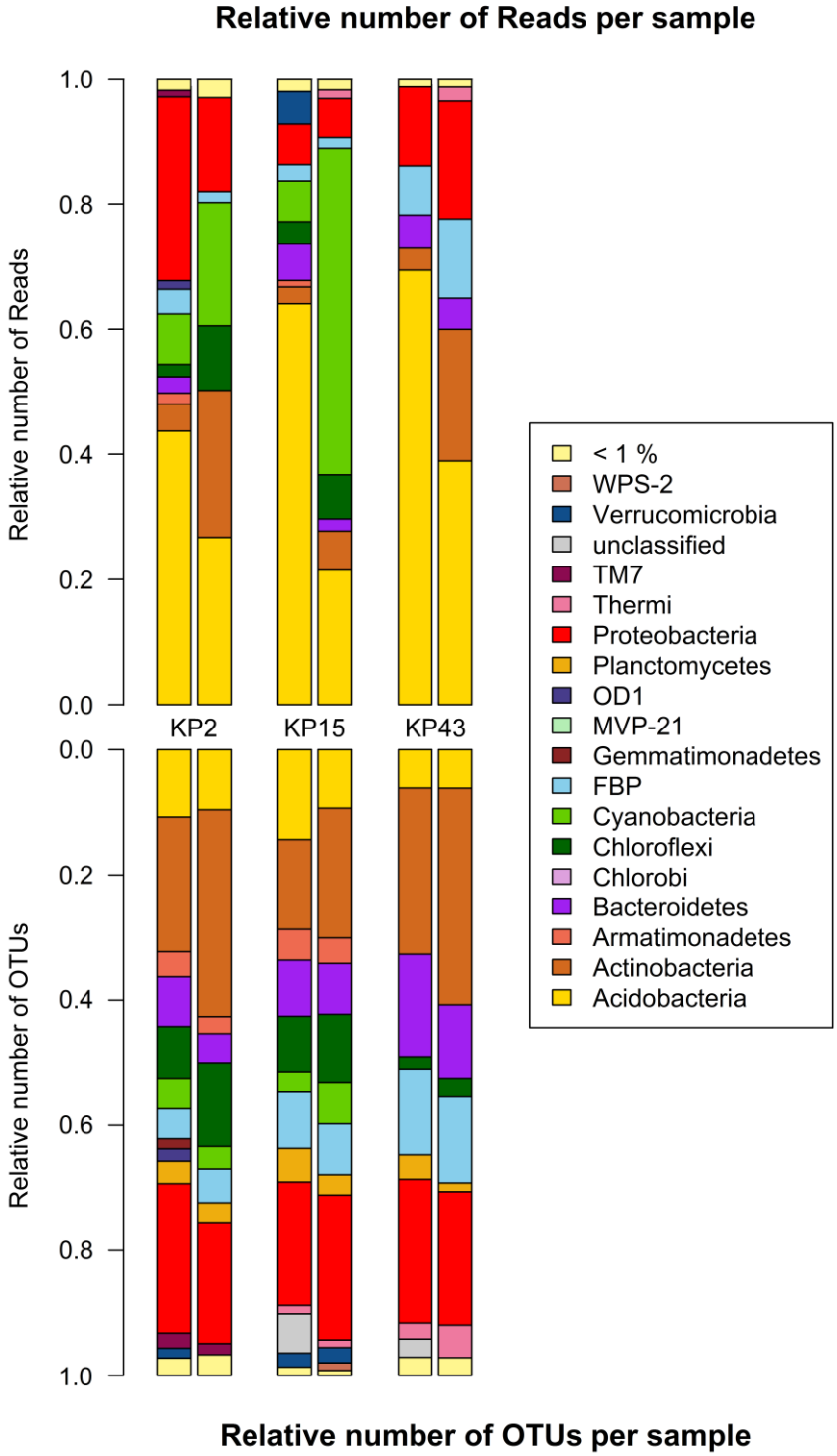

Fig. S1 Graphical comparison between relative abundances of 16S rRNA gene reads and OTUs recovered with different extraction protocols. The comparison is based on the values supplied in Table S2. Bar charts show the relative abundances of reads (top) and OTUs (bottom) recovered with the DNA extraction protocol used by Tytgat et al. [275] (combination of [292] and [293]) and the extraction method used in the current study (PowerLyzer® PowerSoil® DNA isolation kit). For every sample, the left bar chart shows the relative abundances as obtained by Tytgat et al. [275], whereas the right bar chart shows those obtained by use of the PowerLyzer® PowerSoil® DNA isolation kit. Phyla with relative abundances of reads or OTUs below 1.00 % are grouped together in the < 1 % group

Table S1 Composition of ORUs, ONUs and OTUs. Each worksheet is named after the analyzed gene and shows the number of sequences per sample for each ORU (cbbL IAIC), ONU (nifH) or OTU (16S rRNA). For cbbL IAIC, the visual cluster the ORU groups in is listed next to the ORU. Separate ORUs are labelled as "separate". For the 16S rRNA OTUs, classification is provided next to each OTU

Table S1 is available at:
http://users.ugent.be/~gktahon/Chapter 2/

Table S2 Comparison between relative abundances of reads and OTUs recovered with different DNA extraction methods. The left part of the table shows the results obtained with the DNA extraction protocol used by Tytgat et al. [275] (combination of Corinaldesi et al. 2005 [292] and Zwart et al. 1998 [293]). The right part shows results obtained with the DNA extraction method used in the current study (PowerLyzer® PowerSoil® DNA isolation kit)

		Corinaldesi, Danovaro and Dell'Anno 2005, Zwart et al. 1998)				**PowerLyzer® PowerSoil® DNA isolation kit**			
	Phylum	**KP2**	**KP15**	**KP43**	**All**	**KP2**	**KP15**	**KP43**	**All**
Reads	Acidobacteria	43,71 %	64,03 %	69,41 %	64,40 %	26,71 %	21,47 %	38,91 %	27,36 %
	Actinobacteria	4,30 %	2,68 %	3,51 %	3,50 %	23,48 %	6,26 %	21,05 %	16,99 %
	Armatimonadetes	1,76 %	1,04 %	0,01 %	0,46 %	0,67 %	0,70 %	0,00 %	0,54 %
	Bacteroidetes	2,61 %	5,84 %	5,32 %	4,97 %	0,78 %	1,94 %	4,95 %	2,03 %
	Chlorobi	0,67 %	0,00 %	0,00 %	0,11 %	0,00 %	0,00 %	0,00 %	0,00 %
	Chloroflexi	2,00 %	3,58 %	0,09 %	0,96 %	10,33 %	7,04 %	0,74 %	7,24 %
	Cyanobacteria	8,03 %	6,50 %	0,03 %	2,37 %	19,72 %	52,15 %	0,04 %	27,02 %
	FBP	3,94 %	2,62 %	7,84 %	6,36 %	1,74 %	1,73 %	12,70 %	3,96 %
	Gemmatimonadetes	0,15 %	0,06 %	0,04 %	0,06 %	0,14 %	0,00 %	0,00 %	0,06 %
	MVP-21	0,06 %	0,00 %	0,00 %	0,01 %	0,06 %	0,00 %	0,00 %	0,03 %
	OD1	1,36 %	0,00 %	0,06 %	0,26 %	0,14 %	0,00 %	0,04 %	0,07 %
	Planctomycetes	0,48 %	0,98 %	0,39 %	0,50 %	0,32 %	0,31 %	0,11 %	0,27 %
	Proteobacteria	29,34 %	6,44 %	12,60 %	14,28 %	14,94 %	6,20 %	18,78 %	12,68 %
	Thermi	0,03 %	0,45 %	0,48 %	0,40 %	0,18 %	1,40 %	2,26 %	1,03 %
	TM7	1,06 %	0,18 %	0,16 %	0,31 %	0,77 %	0,08 %	0,28 %	0,43 %
	Unclassified	0,21 %	0,42 %	0,08 %	0,16 %	0,00 %	0,00 %	0,14 %	0,03 %
	Verrucomicrobia	0,24 %	5,18 %	0,00 %	0,89 %	0,00 %	0,58 %	0,00 %	0,20 %
	WPS-2	0,03 %	0,00 %	0,00 %	0,00 %	0,02 %	0,14 %	0,00 %	0,06 %
OTUs	Acidobacteria	10.76 %	14.35 %	6.15 %	10.42 %	9.61 %	9.35 %	6.16 %	8.37 %
	Actinobacteria	21.51 %	14.35 %	26.54 %	20.80 %	33.03 %	20.73 %	34.60 %	29.45 %
	Armatimonadetes	3.98 %	4.93 %	0.32 %	3.08 %	2.70 %	4.07 %	0.00 %	2.26 %
	Bacteroidetes	7.97 %	8.97 %	16.50 %	11.15 %	4.80 %	8.13 %	11.85 %	8.26 %
	Chlorobi	0.40 %	0.00 %	0.00 %	0.13 %	0.00 %	0.00 %	0.00 %	0.00 %
	Chloroflexi	8.37 %	8.97 %	1.94 %	6.43 %	13.21 %	10.98 %	2.84 %	9.01 %
	Cyanobacteria	4.78 %	3.14 %	0.97 %	2.96 %	3.60 %	6.50 %	0.47 %	3.53 %
	FBP	4.78 %	8.97 %	13.59 %	9.11 %	5.41 %	8.13 %	13.74 %	9.09 %
	Gemmatimonadetes	1.59 %	0.90 %	0.32 %	0.94 %	0.90 %	0.00 %	0.00 %	0.30 %
	MVP-21	0.80 %	0.00 %	0.00 %	0.27 %	0.60 %	0.00 %	0.00 %	0.20 %
	OD1	1.99 %	0.00 %	0.32 %	0.77 %	0.60 %	0.00 %	0.47 %	0.36 %
	Planctomycetes	3.59 %	5.38 %	3.88 %	4.28 %	3.30 %	3.25 %	1.42 %	2.66 %
	Proteobacteria	23.90 %	19.73 %	22.98 %	22.20 %	19.22 %	23.17 %	21.33 %	21.24 %
	Thermi	0.40 %	1.35 %	2.59 %	1.44 %	0.90 %	1.22 %	5.21 %	2.44 %
	TM7	2.39 %	0.45 %	0.97 %	1.27 %	1.80 %	0.81 %	0.95 %	1.19 %
	Unclassified	0.80 %	6.28 %	2.91 %	3.33 %	0.00 %	0.00 %	0.95 %	0.32 %
	Verrucomicrobia	1.59 %	2.24 %	0.00 %	1.28 %	0.00 %	2.44 %	0.00 %	0.81 %
	WPS-2	0.40 %	0.00 %	0.00 %	0.13 %	0.30 %	1.22 %	0.00 %	0.51%

Table S3 Habitat origin of, and similarity with nearest neighbors per cluster or separate ORU with a close cultured representative. For environmental studies, only one representative was used per cluster. Type IA cbbL clusters are indicated with an asterisk. ENV: environmental sequence

Gene	Cluster/Separate ORU	Closest neighbor(s)		
		Accession no.	Origin	Similarity (%)
cbbL	*Nocardia*/*Solirubrobacter*-like	NZ_AUIK00000000	*Solirubrobacter soli* DSM 22325; Korea; Ginseng field soil	78.0-92.0
		CP006850	*Nocardia nova* SH22a; Brazil; Amazon rain forest	
		NZ_CM001852	*Nocardioides* sp. CF8; USA; Aquifer solids	
		JQ836518	ENV; China; Soil	
		JQ964837	ENV; China; Soil	
		HG941299	ENV; China; Soil	
		DQ683667	ENV; USA; Damariscotta lake sediment	
		KF523933	ENV; China; Soil	
		AY422913	ENV; Hawaii; 210-yr old Kilauea Caldera Rim volcanic deposit	
		KJ720456	ENV; China; Soil	
		HQ174590	ENV; China; Cropland soil	
	UT 3	KJ720473	ENV; China; Soil	76.9-93.3
	Mycobacterium-like	CCBF010000001	*Mycobacterium mageritense*; Spain; Human	84.0-93.3
		JQ836511	ENV; China; Soil	
		JQ964846	ENV; China; Soil	
		HG941325	ENV; China; Soil	
		HQ388555	ENV; China; Paddy soil	
		JF922974	ENV; China; Cropland soil	
		HQ174614	ENV; China; Cropland soil	
		HQ997122	ENV; China; Cropland soil	
		HM208333	ENV; China; Agricultural soil	
		KJ720348	ENV; China; Soil	
	UT 1	KJ720481	ENV; China; Soil	72.3-82.7
	Nitrosospira multiformis-like	CP000103	*Nitrosospira multiformis* ATCC25196; Suriname; Soil	95.3/97.3
		JQ964913	ENV; China; Soil	
	Burkholderiales-like	CP002878	*Cupriavidus necator* N-1; USA; Soil	91.3-94.0
		CP003915	*Advenella mimigardefordensis* DPN7; Germany; Human	
		JQ964917	ENV; China; Soil	
		KJ720488	ENV; China; Soil	
	Rhodospirillum centenum-like	CP000613	*Rhodospirillum centenum* SW; USA; Hot Spring	86.7-94.7
		JQ964986	ENV; China; Soil	
	UT 2	HQ174566	ENV; China; Soil	83.0-88.0
	Salinisphaera shabanensis-like	AFNV02000018	*Salinisphaera shabanensis* E1L3A; Saoudi Arabia; Brine-seawater	93.2-94.0
		KF523930	ENV; China; Soil	
		HG941095	ENV; China; Soil	
		HQ174674	ENV; China; Cropland soil	
		JQ836507	ENV; China; Soil	
		HQ996963	ENV; China; Cropland soil	

Gene	Cluster/Separate ORU	Closest neighbor(s)		
		Accession no.	Origin	Similarity (%)
	Mesorhizobium ciceri-like	CP002447	*Mesorhizobium ciceri* biovar *biserrulae* WSM1271; Italy; Soil	96.0-97.3
		HG941007	ENV; China; Soil	
	Rubrivivax-like	AP012320	*Rubrivivax gelatinosus* IL144;Japan; Waste water	84.7
		NZ_ATUI00000000	*Rubrivivax benzoatilyticus* JA2; India; Rice paddy soil	
		HQ388721	ENV; China; Paddy soil	
		HQ941135	ENV; China; Soil	
	Methyloversatilis universalis-like	AFHG01000057	*Methyloversatilis universalis* FAM5; USA; Lake Washington sediment	90.7-91.3
		KJ720437	ENV; China; Soil	
	Nitrobacter hamburgensis-like	CP000319	*Nitrobacter hamburgensis* X14; Soil	88.7
		HQ174639	ENV; China; Cropland soil	
		HQ997127	ENV; China; Cropland soil	
		KJ720381	ENV; China; Soil	
	Nitrosospira sp. APG3-like	CAUA01000016	*Nitrosospira* sp. APG3; USA; Freshwater lake sediment	92.7-94.7
		HG941039	ENV; China; Soil	
	Bradyrhizobium sp. BTAi1-like*	CP000494	*Bradyrhizobium* sp. BTAi1; USA; Stem nodules of *Aeschynomene indica*	84.8-95.3
		JQ964849	ENV; China; Soil	
		KJ740620	ENV; China; Soil	
		DQ149753	ENV; USA; Pine forest and agricultural soils	
		HG940774	ENV; China; Soil	
		JQ836283	ENV; China; Soil	
nifH	*Nostocales*-like	CP001037	*Nostoc punctiforme* PCC 73102; Australia; *Macrozamia* sp. root	88.9-100
		CP003943	*Calothrix* sp. PCC 7507; Switzerland; Sphagnum bog	
		CP007203	*Nodularia spumigena* CCY9414; Baltic Sea; Brackish water	
		CP003642	*Cylindrospermum stagnale* PCC 7417; Sweden; Greenhouse soil	
		DQ995890	ENV; Australia; Alkaline soil	
		DQ140640	ENV; Bahamas; Salt pond microbial mat	
		KC140445	ENV; Baltic Sea; Water	
		KC667460	ENV; USA; Holm-oak Rhizosphere soil	
		GU196863	ENV; Indian ocean; *Ornithocercus steinii*	
		HM140755	ENV; Antarctica; Microbial mat	
		AY819602	ENV; USA; Conifer forest soil	
		EU915059	ENV; Antarctica; Orange pond microbial mat	

Chapter 4

Diversity of Phototrophic Genes Suggests Multiple Bacteria May Be Able to Exploit Sunlight in Exposed Soils from the Sør Rondane Mountains, East Antarctica

Redrafted from:

Tahon G., Tytgat B. and Willems A. (2016). Diversity of Phototrophic Genes Suggests Multiple Bacteria May Be Able to Exploit Sunlight in Exposed Soils from the Sør Rondane Mountains, East Antarctica. Frontiers in Microbiology 7: doi: 10.3389/fmicb.2016.02026.

Author's contributions:

Conceived and designed the experiments: GT, AW. Performed the experiments: GT. Analyzed the data: GT, BT. Contributed analysis tools: BT. Wrote the paper: GT, AW. All authors approved the final manuscript.

Summary

Microbial life in exposed terrestrial surface layers in continental Antarctica is faced with extreme environmental conditions, including scarcity of organic matter. Bacteria in these exposed settings can therefore be expected to use alternative energy sources such as solar energy, abundant during the austral summer. Using Illumina MiSeq sequencing, we assessed the diversity and abundance of four conserved protein encoding genes involved in different key steps of light-harvesting pathways dependent on (bacterio)chlorophyll (*pufM*, *bchL*/*chlL* and *bchX* genes) and rhodopsins (actinorhodopsin genes), in exposed soils from the Sør Rondane Mountains, East Antarctica. Analysis of *pufM* genes, encoding a subunit of the type 2 photochemical reaction center found in anoxygenic phototrophic bacteria, revealed a broad diversity, dominated by *Roseobacter*- and *Loktanella*-like sequences. The *bchL* and *chlL*, involved in (bacterio)chlorophyll synthesis, on the other hand, showed a high relative abundance of either cyanobacterial or green algal trebouxiophyceael *chlL* reads, depending on the sample, while most *bchX* sequences belonged mostly to previously unidentified phylotypes. Rhodopsin-containing phototrophic bacteria could not be detected in the samples. Our results, while suggesting that Cyanobacteria and green algae are the main phototrophic groups, show that light-harvesting bacteria are nevertheless very diverse in microbial communities in Antarctic soils.

4.1. Introduction

Antarctica is nearly completely covered by ice, with only ~0.32 % of its surface ice-free. Although most of the ice-free regions are found in the Antarctic Peninsula and the Transantarctic Mountains, inland mountain ranges, such as the Sør Rondane Mountains (Dronning Maud Land), also represent an important fraction of the exposed surface area [4, 17]. The absence of vascular plants in continental Antarctica combined with the extreme environmental conditions have led to depleted soils with low availability of nutrients, especially of organic carbon and nitrogen, and water [17, 40, 275]. As a result, the mainly microscopic life in these areas [34, 54, 55] may thus be expected to use alternative energy sources to overcome these limitations. Sunlight, abundantly available during the austral summer, may be an important resource for certain members of the bacterial communities inhabiting exposed continental environments, and this should be reflected in the diversity of key genes for light-harvesting functions.

As life on Earth evolved, microorganisms developed different ways to harvest solar energy. Two main mechanisms have been described, either using rhodopsins or complex photochemical reaction centers that contain (bacterio)chlorophyll [137]. Early phototrophic prokaryotes (~3.5 Giga annum ago) used reductants such as H_2, Fe^{2+} or H_2S for bacteriochlorophyll-dependent anaerobic anoxygenic phototrophy and did not involve oxygen [143, 306, 307]. Later on (at least ~2.4 Giga annum ago), oxygenic chlorophyll-dependent phototrophy, using H_2O, arose in Cyanobacteria and played a key role in oxygenating the Earth's atmosphere [306, 307]. Under these new atmospheric conditions, many of the anaerobic anoxygenic phototrophic bacteria may have disappeared from the now oxygenated habitats, although some groups adapted and embarked on an aerobic lifestyle [146]. These aerobic anoxygenic phototrophic bacteria (AAP) were first reported in 1978 [150] and are defined as aerobic species that synthesize bacteriochlorophyll and use light energy as an auxiliary energy source for their mostly heterotrophic metabolism [146, 149]. Moreover, they do not contain carbon fixation enzymes [152]. Since their discovery nearly four decades ago, numerous AAP, predominantly belonging to the Proteobacteria, have been described from various habitats [146]. Some species capable of aerobic anoxygenic phototrophy have also been found in the Gemmatimonadetes, Acidobacteria and Chloroflexi. Phototrophic species belonging to the latter phylum are, however, not included in the AAP, as are many other aerobic bacteria that synthesize Bchl and perform anoxygenic phototrophy under aerobic

conditions (*e.g.* phototrophic methylotrophs) [146, 147, 152]. The majority of all aforementioned anoxygenic phototrophs rely on a heterodimeric type 2 reaction center with *pufL* and *pufM* encoding the conserved proteins. Hence, these two *puf* genes are frequently used and convenient markers to study the diversity of anoxygenic phototrophic bacteria [146, 174, 175].

Additionally, several other genes encoding subunits of key enzymes in the (bacterio)chlorophyll synthesis pathway are also well conserved among phototrophic microorganisms. All oxygenic and anoxygenic phototrophic bacteria use the dark-operative protochlorophyllide oxidoreductase (DPOR) enzyme complex, encoded by the *chlLNB* and *bchLNB* genes, respectively. Apart from in bacteria, DPOR is also found in green algae and lower land plants [308]. The complex plays a key role in the biosynthesis of (bacterio)chlorophyll, converting protochlorophyllide to chlorin [162, 164]. Whereas in Cyanobacteria, green algae and lower land plants, chlorin is converted immediately to chlorophyll (Chl), in anoxygenic phototrophic bacteria (APB) a second enzyme complex, chlorin oxidoreductase (COR), encoded by *bchXYZ* genes, reduces chlorin to bacteriochlorophyllide, the direct precursor for bacteriochlorophyll (Bchl) [159, 309]. DPOR and COR exhibit a high degree of structural similarity. Interestingly, the amino acid sequences of the different DPOR and COR subunits (BchLNB/ChlLNB and BchXYZ, respectively), exhibit significant similarity (~15-30 %) to those of the nitrogenase enzyme complex (NifHDK), leading to the hypothesis that these three enzyme complexes all evolved from the same common ancestor [162-164].

Besides phototrophy using photochemical reaction centers, a second type of phototrophy, employing rhodopsins also evolved [137], although little is known about its origin in time. Microbial rhodopsins have been described in various groups, mostly in aquatic habitats, performing a range of functions, including light-driven ion pumping [181, 190]. Although previously detected in Siberian permafrost [194], Antarctic sea ice, sea water and continental lakes [193, 208, 310-314], little rhodopsin data are available for terrestrial Antarctica (based on metagenome data available on MG-RAST [207] and IMG [208]). Recently, a new family of proton pumping microbial rhodopsins, actinorhodopsins, has been discovered in freshwater Actinobacteria, [199]. To our knowledge, their occurrence in Antarctica has not been reported yet.

A first cloning survey of genes for phototrophic mechanisms in samples from the oligotrophic high-altitude soils near the Belgian Princess Elisabeth Station in the Sør Rondane Mountains revealed a high diversity of *pufLM* genes, whereas proteorhodopsin genes could not be amplified from any of the samples [277]. In this study, we aimed to more comprehensively assess the diversity of bacteria capable of exploiting sunlight as an alternative energy source. To further test the hypothesis that sunlight may be a very important resource for certain members of the bacterial communities inhabiting these exposed oligotrophic soils, an Illumina MiSeq paired-end 300 bp sequencing approach was used with primers targeting *pufM*, actinorhodopsin and *bchL*/*chlL*/*bchX* genes. For the latter, sequence data obtained in a previous study into the diversity of *nifH* genes [Tahon et al. Submitted (Chapter 3)], that were originally discarded because they lacked multiple of the conserved NifH amino acid residues, were revisited [108, 162].

4.2. Materials and Methods

Study site and sample collection

Four samples, previously used in the pilot survey [277], were studied (Table 1), to allow comparison. During the Antarctic summer of 2009, top surface samples - mostly consisting of weathered granite parent material - were collected aseptically in the vicinity of the Belgian Princess Elisabeth Station (71° 57' S, 23° 20' E) at Utsteinen, Dronning Maud Land, East Antarctica. All samples were frozen at -20 °C upon collection. Sample KP2 was collected ~1.3 km south of the research station. The three other samples were collected on the Utsteinen ridge, ~500 m north of the Belgian base.

DNA extraction

From each homogenized sample, 400 mg subsamples were taken in triplicate. Total genomic DNA was extracted and purified using the PowerLyzer® PowerSoil® DNA isolation kit (MoBio Laboratories) and a modified lysis protocol as instructed by the manufacturer. This extraction protocol was previously identified as the one yielding most bacterial diversity [277]. Following extraction, DNA was quantified using the Qubit® 2.0 fluorometer (Life Technologies) and stored at -20 °C until processing.

PCR and preparation for Illumina sequencing

A Veriti thermal cycler (Life Technologies) was used to amplify partial actinorhodopsin, *pufM*, *bchL*/*chlL* and *bchX* genes. Primer selection was based on two criteria: 1) to amplify a broad diversity of the gene and 2) to produce an amplicon size suitable for Illumina MiSeq 300 bp paired-end sequencing (Table 2). To complement the Nextera XT index kit (Illumina), primers were extended with an adapter.

For each of the soil samples, PCR was performed in triplicate on all three DNA extracts, for each primer set, resulting in a total of nine PCR products per sample per gene. PCRs were performed in 25 µl reaction mixtures containing 3 µl of genomic DNA (> 6 ng μl^{-1}), 1x Qiagen PCR buffer (Qiagen), 0.2 mM of each deoxynucleotide triphosphate, 0.625 U of Qiagen *Taq* polymerase (Qiagen), 100 mM bovine serum albumin and forward and reverse

primer with final concentrations as shown in Table 2. All nine PCR products (three DNA extracts x three PCRs) were pooled and purified using the Ampure beads XT (Agencourt) protocol with slight modifications. Briefly, only 0.8 reaction volume of beads was used and DNA was resuspended in MilliQ water. Tagging of pooled PCR products was performed using the Nextera XT indices (Illumina) during an eight cycle version of the amplicon PCR with the indices replacing the primers. Afterwards, PCR products were purified as described above, with resuspension in Tris buffer (0.1 M, pH 8.5). Integrity and amplicon sizes of the PCR products were checked using a BioAnalyzer (Agilent), following quantification using a Qubit, as described above. Afterwards, samples were pooled equimolarly and sequenced on an Illumina MiSeq 300 bp paired-end platform (GATC). PhiX was spiked at 20 % per lane.

Sequence data processing

For all genes, the forward and reverse sequencing reads were merged using the *fastq_mergepairs* command in USEARCH [282] allowing a minimum overlap length of 8 nucleotides and a maximum of six mismatches in the overlapping region. For *bchL*/*chlL*/*bchX* and *pufM*, merged sequences shorter than 370 and 200, and longer than 470 and 350, were removed using the *fastq_minmergelen* and *fastq_maxmergelen* commands, respectively. Primer sequences were removed from the merged sequences using cutadapt v1.8 [283], resulting in sequences with a minimum length of 193 and 321 bp, and a maximum length of 225 and 369 bp for *pufM* and *bchL*/*chlL*/*bchX*, respectively. Subsequently, during quality filtering using USEARCH, sequences with one or more nucleotides beneath the Phred Q20 threshold score and a maximum error >0.5 were removed from further analyses. Afterwards, all sequences were placed in reading frame +1, followed by removal of sequences showing no similarity to our genes of interest or containing stop codons and/or indels resulting in a frameshift. Detection of putative chimeric sequences was done using the Uchime model (default parameters) [247] in Mothur [246]. Finally, all remaining sequences were translated to proteins using MEGA 6 using the bacterial genetic code [245].

Table 1 Parameters associated with analyzed samples

Sample	Sample Coordinates	Altitude (m)	Description of sample area	Conductivity (µS/cm)	pH	Water Content	TOC
KP2	71° 57' 28.6" S, 23° 19' 45.8" E	1320	Small gravel particles in between the rocks, Utsteinen nunatak	19	6.54	6.28 %	0.08 %
KP15	71° 56' 45.8" S, 23° 20' 43.6" E	1366	Brown soil under lichen, East part of Utsteinen ridge	33	5.57	3.38 %	0.33 %
KP43	71° 56’ 47.3” S, 23° 20’ 44.6” E	1362	Brown soil with dark green fragments, East part of Utsteinen ridge	520	6.22	0.91 %	2.57 %
KP53	71° 56’ 45.3” S, 23° 20’ 42.4” E	1362	Grey soil on East part of Utsteinen ridge	312	6.34	0.23 %	0.21 %

Table 2 PCR primers and conditions used for screening different genes

Gene	Target	Primer	Sequence 5’-3’	Final concentration	Region	Amplicon size	Program[g]
pufM	Universal	pufM_uniF[a]	GGN AAY YTN TWY TAY AAY CCN TTY CA	1.0 µM	584-825[d]	± 240 bp	94°C (4 min); 35x 94°C (40 s), 49°C (30 s), 72°C (30 s); 72°C (7 min)
		pufM_WAW[a]	AYN GCR AAC CAC CAN GCC CA	0.5 µM			
actinorhodopsin	Clade LG1 & LG2	LG-for[b]	TAY MGN TAY GTN GAY TGG	0.4 µM	283-614[e]	± 330 bp	95°C (7 min), 45x 94°C (30 s), 51.5°C (1 min 30 s), 72°C (30 s) ; 72°C (10 min)
		LG1A-for[b]	MGN TAY ATH GAY TGG YT	0.4 µM			
		LG2-for[b]	TAY MGN TAY GCN GAY TGG	0.4 µM			
		LG-rev[b]	ATN GGR TAN CAN CCC CA	0.8 µM			
nifH, bchL, chlL, bchX	Universal	IGK3[c]	GCI WTH TAY GGI AAR GGI GGI ATH GGI AA	1.0 µM	19-413[f]	395 bp	95°C (10 min); 40x 95°C (45 s), 52°C (30 s), 72°C (40 s); 72°C (10 min)
		DVV[c]	ATI GCR AAI CCI CCR CAI ACI ACR TC	1.0 µM			

[a] From [232], [b] From [201], [c] From [279], [d] Based on the *pufM* sequence of *Sphingomonas sanxanigenens* DSM 19645 (CP006644), [e] Based on the actinorhodopsin sequence of *Leifsonia rubra* CMS 76R (ATIA01000023), [f] Based on the *nifH* sequence of *Azotobacter vinelandii* (M20568), [g] All programs were optimized in this study

Sequence analyses

For *pufM*, an updated version of our previously described database containing publicly available sequences [277] was used. For *bchL*/*chlL* and *bchX* a new database was assembled to contain all related sequence records from NCBI and IMG (https://img.jgi.doe.gov/) [244] available per November 15th 2015. Newly obtained nucleotide sequences and their derived protein sequences were added to the databases using the import module of BioNumercs 7.5 (Applied Maths). For the sequences obtained with the primer set IGK3/DVV, to retain only BchL/ChlL (L subunit of DPOR) and BchX (X subunit of COR) sequences, NifH sequences were separated based on the presence of the conserved amino acid residues Ala43, the dipeptide Glu93-Pro94, Arg101, the dipeptide Ile104-Thr105 and Glu111 (positions based on the NifH protein sequence of *Azotobacter vinelandii*, accession number M20568) [162]. For phylogenetic analyses, all Illumina PufM sequences were clustered at a 95 % cut-off using CD-HIT [289, 290], grouping them into operational puf units (OPUs) of which one representative was used to construct the phylogenetic tree. BchL/ChlL and BchX sequences were grouped in operational BchL/ChlL units (OLUs) and operational BchX units (OXUs), respectively, at 95 % cut-off. BchL/ChlL and BchX sequences were processed together for phylogenetic analyses.

When analyzing sequencing data, there is no consensus on inclusion or removal of singletons. Although many authors choose to remove them, particularly when the focus is on dominant community members [315], we have opted not to do so, for several reasons. Previous research has shown that singletons may be informative and valuable in reflecting rare and/or unique lineages of dormant or inactive bacteria that may grow when the suitable conditions are met [316]. Removal of singletons would thus lead to a loss of power to detect these rare lineages in communities and lead to an underestimation of biodiversity levels [317-320]. On the other hand, singletons may represent erroneous sequences and therefore the use of a very stringent quality control is required to accurately sort informative low-abundance sequence reads from errors and artifacts. We therefore implemented a very stringent quality control. The length of the *pufM*, *bchX* and *bchL*/*chlL* amplicons (Table 2) allowed a large or even complete overlap between the forward and reverse sequencing reads. In the overlapping region, only six mismatches were allowed and every sequence with one or more nucleotides with a base call accuracy lower than 99 % was discarded. Furthermore, since we studied protein-encoding genes rather than 16S rRNA genes, a number of additional quality control steps could be performed. The gene sequences were placed in frame +1, translated into amino acids and

these were analyzed for the presence of stop codons, indels that result in a frameshift and presence of conserved sites. These steps allowed additional removal of erroneous sequences so that the number of singleton sequences was reduced a thousand fold and the final data are of much higher quality and likely represent real, though perhaps rare, diversity. Indeed, the vast majority of leftover singletons were found to group within the named clusters or close to sequences of named species (Figures 3, 4a, 4b and Tables 3, S1). As the goal of our study was to explore the (nearly) whole diversity of several protein encoding genes, including rare types, the number of singletons retrieved also aids to get more insight in the diversity coverage of the approach used. To estimate total diversity in a sample different parameters (*e.g.* Chao1, ACE) can be used [321] (Table 3). Calculation of these parameters takes the number of singletons into account and removal of singletons would confound the ability to estimate alpha diversity.

A first alignment was made with all sequences present in our databases, using Clustal Omega [248, 249]. Afterwards, alignments were trimmed to the size of our sequenced amplicons and visually inspected, excluding from further analysis all non-overlapping reference sequences. Remaining sequences were realigned, after which the alignment was used to construct a maximum likelihood (ML) phylogenetic tree (1000 bootstrap replicates), by using the FastTree tree building software [250] with the Whelan and Goldman evolutionary model and the discrete gamma model with 20 rate categories. From the resulting phylogenetic trees, closest relatives of our newly obtained sequences as well as representative sequences from the entire tree were selected to prepare a smaller tree representing the initial complete tree, following the same protocol. Sequences from uncultured bacteria were not included in the final tree. Trees were visualized using the iTOL software [251, 252] and related OPUs, OLUs or OXUs were grouped into visual clusters that were named after cultivated bacteria that grouped in or close to the cluster. In the absence of cultivated members, the clusters were given an Utsteinen (UT) cluster number designation.

Statistical analyses

For statistical analyses of PufM, BchL/ChlL and BchX sequences, the Vegan package [253] in R (https://cran.r-project.org) was used. A non-normalized table of the total number of protein sequences per OPU, OLU or OXU was used to perform rarefaction analyses and determine the total number of expected OPUs/OLUs/OXUs per sample and for the four samples combined. Parameters calculated for each dataset include species richness (Chao1) and evenness (Pielou). A normalized table (consensus of 10,000 iterations) of the total number of protein sequences per OPU, OLU or OXU was also generated, to assess the impact on relative abundances.

Accession numbers

Raw sequences were submitted to the NCBI sequence read archive under accession number SRP067116.

4.3. Results

Actinorhodopsin genes could not be amplified from the samples and were therefore not included in the Illumina run. Sequencing and thorough quality control of sequence data resulted in 678940 high-quality *pufM* sequences (length 193-198 bp) and 119822 and 4950 high-quality *bchL*/*chlL* and *bchX* sequences (length 321-348 bp). At a 95 % cut-off, the PufM protein sequences constituted 925 OPUs of which 248 were singletons. For BchL/ChlL and BchX, the sequences grouped in 207 OLUs and 48 OXUs, respectively. A total of 48 OLUs and 18 OXUs were singletons (Table 3).

As there is no consensus on the use of normalization which may lead to the omission of valuable diversity data [322], we assessed how much diversity would be lost by normalizing. For each gene, the original OTU table was normalized 10000 times. Based on these results, box plots were generated showing the large variation in the OTU diversity recovered after standardization (Figure S1). Therefore we created a consensus OTU table out of 10,000 standardizations and used this for calculating the relative abundances of reads and OTUs (Figure S2). Surprisingly, differences with relative abundances calculated using non-normalized data (Figure 1) are limited (<0.05 % for most of the OTUs and clusters). Normalization does, however, involve removal of sequence data and as a result leads to the discarding of (rare) diversity. This is most clear in the BchX dataset where the effect of normalization on sample KP43 is more pronounced as this sample had a very large number of reads and rare OTUs, many of which were removed. To avoid potential loss of rare sequence types, non-normalized data was used in the diversity assessments.

PufM

For PufM, of the 925 OPUs, 358 were detected in one sample only (Figure S3, Table S1). The number of OPUs in each sample varied from 463 to 644 (Table 3). Of 228 shared OPUs, two contained the majority of PufM sequences and both grouped with PufM of heterotrophic alphaproteobacterial AAPs. OPU C1 contained 80.11 % of all sequences and was 98.48 % similar to the PufM sequence of *Roseobacter denitrificans* OCh 114 (Figure 1, Table S1). OPU C2 comprised 12.03 % of all reads and had the PufM sequence of *Loktanella* sp. RCC2403 as closest match (96.97 % amino acid similarity) (Figure 1, Table S1). Rarefaction analysis showed that, although the graphs started to flatten, saturation was not yet reached

(Figure 2). This was corroborated by the number of estimated OPUs (Chao1) that was much higher than the number of observed OPUs (Table 3).

After ML analysis, nearly all 925 OPUs grouped into 28 clusters (Figure 3), mostly containing reads from all four terrestrial samples as well as PufM sequences originating from cultured bacteria (Table 4). The *Loktanella*-like and *Roseobacter*-like clusters were the largest, containing 273 and 386 of the OPUs and, since they contained OPUs C2 and C1, they also contained most of the reads: 12.70 % and 82.39 % respectively. It should be noted that the cluster defined as *Roseobacter*-like also contains some PufM sequences of *Tateyamaria*, *Erythrobacter* and *Jannaschia* (< 0.1 % of reads). The third largest cluster (PufM UT 5) grouped among less related alphaproteobacterial AAP PufM sequences and contained only 1.59 % of the sequence data, mostly originating from sample KP43 (Figures 1 and 3, Table 4). The other clusters and the separate OPUs each contained less than 0.85 % of the reads. In the phylogenetic ML tree, our OPUs grouped with a broad variety of known PufM sequences originating from Alpha-, Beta- and Gammaproteobacteria, and even Chloroflexi (Figure 3). Affiliations with alphaproteobacterial AAP PufM sequences, however, were most frequent. Several clusters (PufM UT 1 – PufM UT 5) and separate OPUs did not group closely with known diversity, suggesting that many organisms harboring *pufM* genes still remain unreported.

The PufM sequences appeared to be cosmopolitan: a broad diversity of cultured and uncultured sequences originating from habitats from all over the world, including polar regions, was found grouping with the new PufM sequences (Table S2). Some clusters, however, represented new PufM phylotypes, as no cultured or environmental PufM sequence was found grouping with them (*e.g.* cluster PufM UT 3).

Table 3 Overview of sequence data characteristics. Diversity indices were calculated on the basis of derived protein sequences, binned at 95 % similarity

		Illumina data					Normalized data				
		KP2	**KP15**	**KP43**	**KP53**	**All**	**KP2**	**KP15**	**KP43**	**KP53**	**All**
PufM	No. of sequences	192517	246511	95721	144191	678940	11715	11715	11715	11715	46860
	No. of OPUs	509	644	502	463	925	72	85	110	71	171
	No. of singleton OPUs	44	97	76	31	248	6	18	35	15	74
	No. of identified singleton OPUs[a]	43	92	76	31	242	6	16	35	15	72
	evenness (H/H_{max})	0.106	0.128	0.261	0.100	0.136	0.141	0.171	0.331	0.130	0.188
	Chao1	666.51	884.32	781.40	634.92	1227.39	110.75	214.00	157.83	166.14	329.88
BchL/ChlL	No. of sequences	72910	18836	11715	16361	119822	11715	11715	11715	11715	46860
	No. of OLUs	115	113	75	66	207	82	113	75	66	192
	No. of singleton OLUs	11	13	17	7	48	7	22	17	9	55
	No. of identified singleton OLUs[a]	11	13	16	7	47	7	22	16	9	54
	evenness (H/H_{max})	0.267	0.243	0.252	0.214	0.380	0.286	0.244	0.252	0.214	0.395
	Chao1	136.11	130.65	120.11	85.25	277.50	124.86	158.56	120.11	154.00	279.35
BchX	No. of sequences	53	345	4241	311	4950					
	No. of OXUs	7	19	32	8	46					
	No. of singleton OXUs	3	5	10	0	18					
	No. of identified singleton OXUs[a]	3	5	10	0	18					
	evenness (H/H_{max})	0.623	0.315	0.168	0.216	0.231					
	Chao1	13.00	30.25	45.75	8.33	67.86					

[a] The number of identified singleton OPUs, OLUs or OXUs corresponds to singletons that show a high similarity to, respectively, PufM, BchL/ChlL or BchX sequences of cultured organisms or that belong to the Utsteinen (UT) clusters (details provided in Table S1)

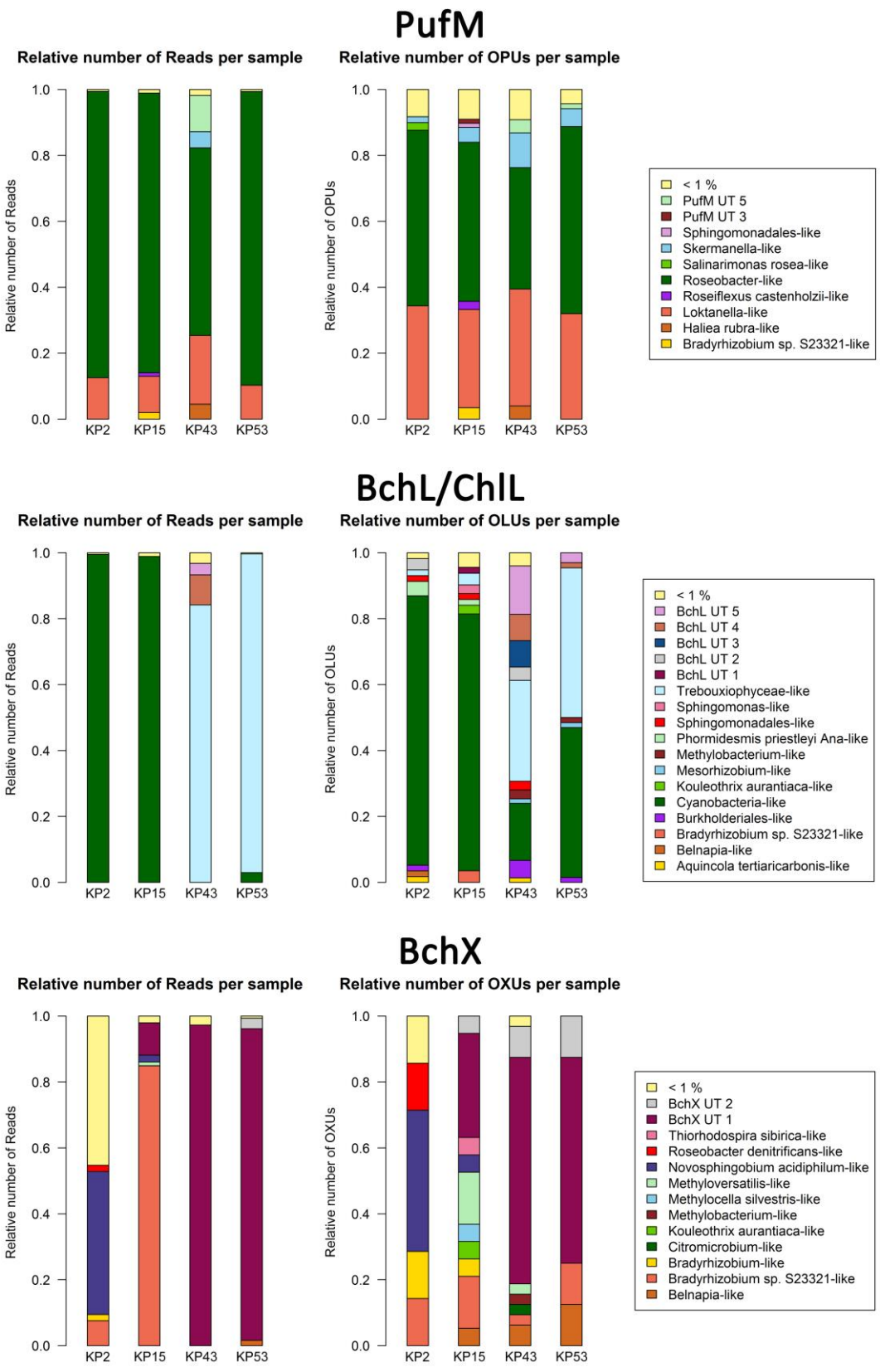

Figure 1 Bar plots showing relative numbers of reads and OPUs (PufM), OLUs (BchL/ChlL) or OXUs (BchX) per cluster. Data were not normalized (normalized bar plots are shown in Figure S2). Clusters or separate OPUs/OLUs/OXUs containing less than 1 % of the data were grouped together in the < 1 % group

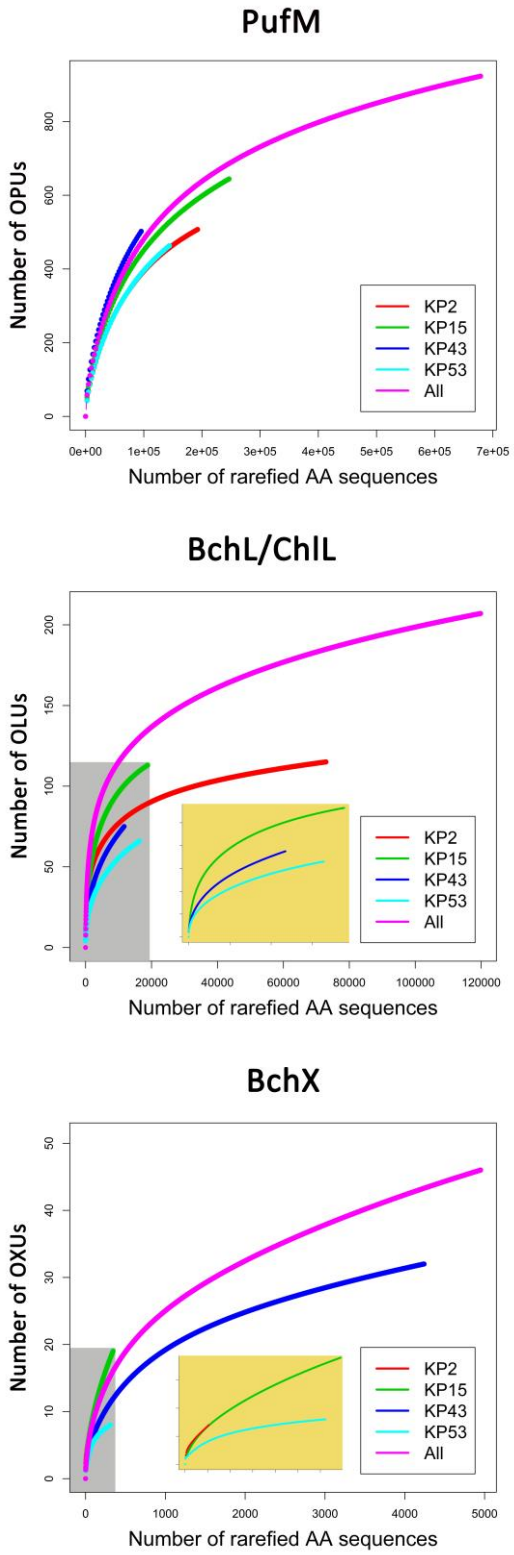

Figure 2 Rarefaction curves based on grouping protein sequences that have 95 % similarity. Analysis was performed using the Vegan package in R. Embedded figures (beige background in BchL/ChlL and BchX) show a more detailed view of rarefaction curves completely enclosed in the gray area

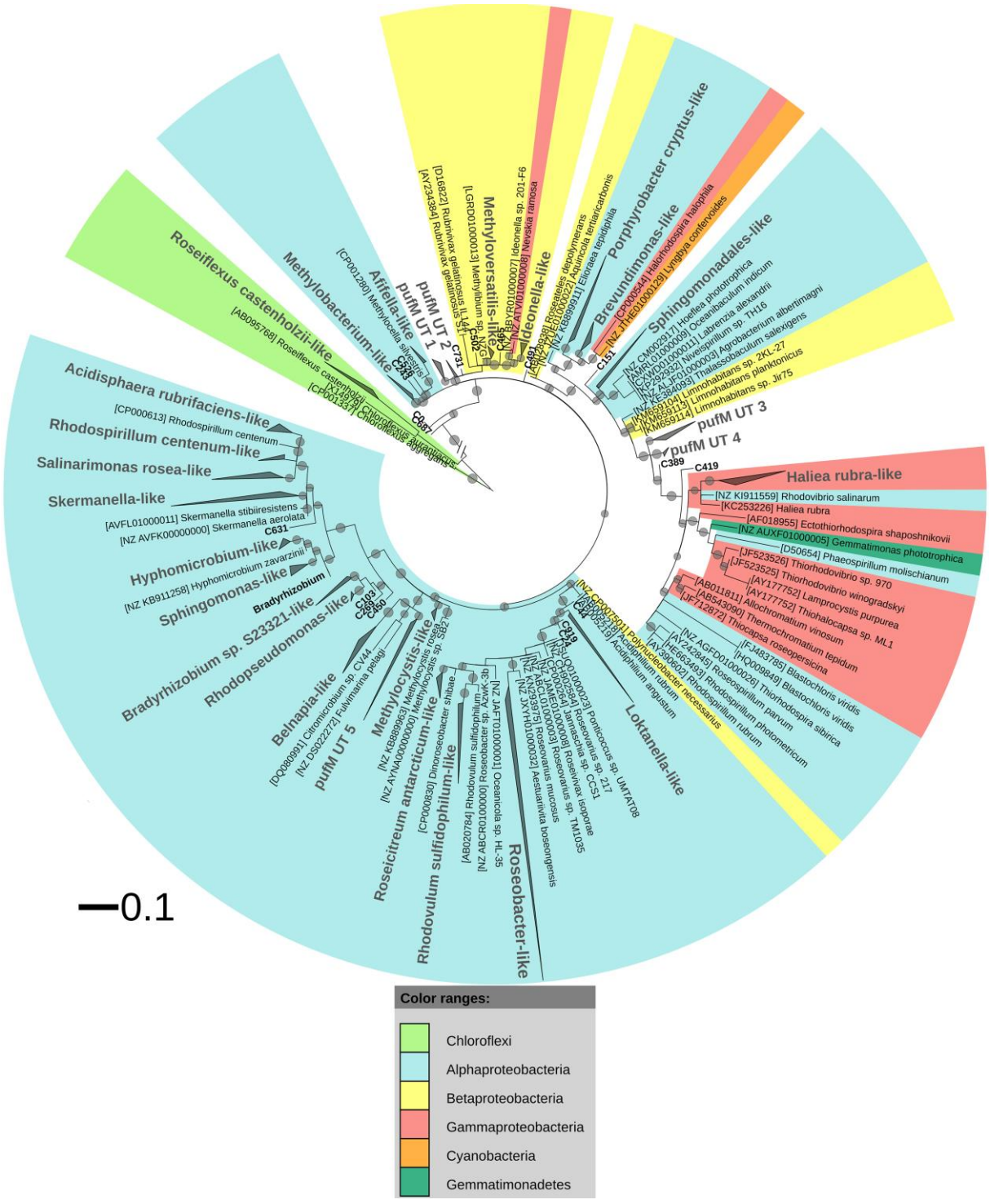

Figure 3 ML phylogenetic tree (1000 bootstrap replicates) of PufM sequences. Scale bar indicates 0.1 substitutions per amino acid position. OPU clusters (larger font size) were named after cultivated bacteria that grouped in or close to the cluster. In the absence of cultivated members, clusters were given an Utsteinen (UT) cluster number designation. For clusters, the total branch lengths to the closest and farthest leaf of the cluster were used as sides of the triangle. OPUs not enclosed in clusters are labelled in bold. For reference data, taxon name and accession number is listed. Bootstrap values are displayed as circles with a diameter reflecting the height of the bootstrap value. Smallest circles represent the lower cut-off of 70 %. Chloroflexi PufM sequences were used as an outgroup. PufM sequences originating from cultured *Bradyrhizobium* species were grouped in the cluster defined as '*Bradyrhizobium*' to simplify the topology of the tree

BchL/ChlL

The 119822 sequences that grouped with BchL/ChlL reference sequences constituted 207 OLUs, with the number of OLUs per sample varying from 66 to 114 (Table 3). A total of 48 OLUs contained only one sequence, whereas 96 OLUs were restricted to sequences from one sample (Figure S3, Table S1). Nine OLUs were common to all four samples (Figure S3). Of these, three OLUs (C0, C1 and C13) together represented 79.07 % of all BchL/ChlL sequences (Table S1). OLU C13 (47.37 % of sequences, mostly from KP2) and OLU C1 (13.38 % of sequences, mostly from KP15) were ~97.7 % similar to the ChlL sequences of *Synechococcus* sp. PCC 6312 (order *Chroococcales*, Cyanobacteria) and *Microcoleus vaginatus* (order *Oscillatoriales*, Cyanobacteria), respectively (Figure 1). Interestingly, OLU C0, containing 18.33 % of the sequences and retrieved mostly from samples KP43 and KP53, displayed a very high similarity (~99.1 %) to Trebouxiophyceae ChlL (Figures 1 and S4, Table S1). All remaining OLUs contained less than 3.07 % of the sequences.

Rarefaction analysis (Figure 2) showed saturation was not yet reached for the single, or the pooled samples, although the graphs started to flatten. Indeed, the estimated number of OLUs (Chao1 values) was only slightly higher than the number of observed OLUs, indicating most diversity was retrieved from the samples (Table 3). Furthermore, all samples showed a similar, very uneven distribution (Table 3).

After ML analysis, the 207 OLUs could be grouped into 17 clusters, 12 of which grouped with BchL/ChlL sequences from known microorganisms, and three separate OLUs (C71, C107 and C115) (Figures 1 and 4a). The clusters were dispersed all over the BchL/ChlL phylogeny although the majority of OLUs and reads grouped with ChlL from oxygenic photosynthetic organisms (Cyanobacteria and Trebouxiophyceae green algae) (Tables 4 and S1, Figures 1, 4a and S4). Notably, a small number of cyanobacterial reads from samples KP2 and KP15 (cluster *Phormidesmis priestleyi* Ana-like), together with sequences mainly obtained from marine unicellular Cyanobacteria (Figure 4a and Table 4), grouped among proteobacterial BchL. This aberrant grouping was previously ascribed to the occurrence of several shared conserved signature indels, absent from ChlL of other Cyanobacteria [164]. The other clusters grouped with BchL of Chloroflexi and Proteobacteria or belonged to four clusters (BchL UT 1 – 4) that could not be assigned to a named phylum (Figure 4a). OLUs grouping with Chloroflexi were only retrieved from sample KP15 (Figure 4a and Table 4). Of the clusters grouping among proteobacterial BchL, most grouped with Alphaproteobacteria,

some with Betaproteobacteria, and none was found grouping with Gammaproteobacteria (Figure 4a).

No clear psychrophilic association could be deduced from the habitat metadata of the nearest neighbors. Most BchL/ChlL sequences grouped together with sequences retrieved from samples taken in a variety of ecosystems worldwide (Table S2).

BchX

A total of 4950 BchX sequences were obtained, many from sample KP43. They were binned at 95 % protein similarity into 46 OXUs of which 31 were unique to one of the samples and 18 were singletons (Figure S3, Tables 3 and S1). No OXU was found common between all four terrestrial samples (Figure S3). A single OXU, OXU C3, shared between samples KP15, KP43 and KP53, contained 82.10 % of all BchX sequences and made up most of the reads retrieved from samples KP43 and KP53 (Figure 1, Table S1). However, phylogenetic analysis revealed that it did not group together with BchX sequences of known bacteria (Figure 4b). The less abundant OXUs C2 (3.68 %) and C32 (2.61 %) grouped in the same cluster as OXU C3 (BchX UT 1) and contained nearly exclusively sequences from sample KP43 (Table S1). OXU C1 (5.72 %), containing sequences from samples KP2, KP15 and KP43, grouped with the BchX sequence originating from *Bradyrhizobium* sp. S23321 (Table S1). All other OXUs represented less than 1 % of BchX sequences.

Rarefaction analysis (Figure 2) indicated that saturation was not reached. The graph of sample KP53, however, started to flatten. This was also confirmed by the number of estimated OXUs, which was nearly identical to the number of observed OXUs for sample KP53 (Table 3). Evenness analysis resulted in very low values, except for sample KP2, which showed a more even distribution, although this could be explained by the very low number of sequences retrieved from this sample, grouping in seven OXUs (Table 3).

After ML analysis, 27 OXUs, of which 22 grouped in cluster BchX UT 1, could not be associated with a named phylum. The remaining 19 OXUs grouped with BchX sequences of 11 known bacterial taxa (Figure 4b) of Chloroflexi and Alpha-, Beta- and Gammaproteobacteria, although the latter was represented by only one sequence (OXU C30) (Figure 4b, Table S1). For samples KP43 and KP53, however, most of the sequences and OXUs grouped in two clusters (BchX UT 1 and BchX UT 2) or separate OXUs that could not be associated with a known BchX reference sequence (Figure 1, Tables 4 and S1).

Similar to PufM and BchL/ChlL, no clear grouping of BchX sequences from cold habitats could be seen, as our sequences grouped with sequences retrieved from a broad diversity of terrestrial and aquatic ecosystems worldwide (Table S2).

Table 4 Distribution of OTUs (95 % protein similarity) and reads per PufM, BchL/ChlL or BchX cluster or separate OTU. Values in brackets refer to the OTUs and reads obtained from the individual samples KP2, KP15, KP43 and KP53, respectively. Separate OTUs without a close known cultured representative are not included. For the Cyanobacteria/Trebouxiophyceae-like ChlL cluster Cyanobacteria-like and Trebouxiophyceae-like sequences are listed separately

	Cluster/Separate OTU	**No. of OTUs**	**OTUs (%)**	**No. of reads**	**Reads (%)**
PufM	*Roseobacter*-like	386 (271, 311, 185, 263)	41.73 %	559360 (167181, 209233, 54446, 128500)	82.39 %
	Loktanella-like	273 (175, 192, 178, 148)	29.51 %	86210 (24202, 27177, 20025, 14806)	12.70 %
	Skermanella-like	74 (9, 29, 53, 25)	8.00 %	5796 (153, 600, 4713, 330)	0.85 %
	Haliea rubra-like	23 (1, 4, 20, 2)	2.49 %	4348 (6, 29, 4300, 13)	0.64 %
	PufM UT 5	22 (4, 6, 20, 7)	2.38 %	10789 (29, 80, 10500, 180)	1.59 %
	Bradyrhizobium sp. S23321-like	22 (1, 22, 1, 1)	2.38 %	4945 (50, 4887, 5, 3)	0.73 %
	Roseiflexus castenholzii-like	16 (2, 16, 3, 1)	1.73 %	2572 (29, 2536, 6, 1)	0.38 %
	Salinarimonas rosea-like	13 (12, 4, 2, 0)	1.41 %	315 (289, 22, 4, 0)	0.046 %
	Sphingomonadales-like	10 (3, 8, 4, 0)	1.08 %	64 (6, 46, 12, 0)	0.0094 %
	Porphyrobacter cryptus-like	9 (3, 3, 5, 3)	0.97 %	671 (82, 17, 453, 119)	0.099 %
	PufM UT 3	8 (3, 8, 2, 0)	0.86 %	647 (164, 463, 20, 0)	0.095 %
	PufM UT 2	6 (1, 5, 5, 1)	0.65 %	1209 (8, 775, 419, 7)	0.18 %
	Methylobacterium-like	6 (1, 2, 5, 2)	0.65 %	529 (44, 20, 455, 10)	0.078 %
	Belnapia-like	6 (1, 0, 4, 4)	0.65 %	131 (20, 0, 88, 23)	0.019 %
	Rhodovulum sulfidophilum-like	6 (1, 5, 1, 1)	0.65 %	35 (1, 32, 1, 1)	0.0052 %
	Roseicitreum antarcticum-like	5 (4, 3, 1, 2)	0.54 %	185 (31, 91, 14, 49)	0.027 %
	PufM UT 1	5 (0, 1, 4, 0)	0.54 %	5 (0, 1, 4, 0)	0.00074 %
	PufM UT 4	3 (3, 1, 0, 0)	0.32 %	66 (58, 8, 0, 0)	0.0097 %
	Methylocystis-like	2 (1, 2, 1, 0)	0.22 %	276 (2, 177, 97, 0)	0.041 %
	Hyphomicrobium-like	2 (2, 1, 1, 0)	0.22 %	71 (64, 4, 3, 0)	0.010 %
	Ideonella-like	2 (2, 1, 1, 0)	0.22 %	65 (5, 4, 56, 0)	0.0096 %
	Rhodospirillum centenum-like	2 (1, 2, 0, 0)	0.22 %	32 (15, 17, 0, 0)	0.0047 %
	Methylibium-like	2 (0, 2, 1, 0)	0.22 %	10 (0, 8, 2, 0)	0.0015 %
	Acidisphaera rubrifaciens-like	1 (0, 0, 1, 0)	0.11 %	13 (0, 0, 13, 0)	0.0019 %
	Methyloversatilis-like	1 (0, 1, 0, 0)	0.11 %	5 (0, 5, 0, 0)	0.0007 %
	Brevundimonas-like	1 (1, 0, 0, 0)	0.11 %	3 (3, 0, 0, 0)	0.00044 %
	Afifella-like	1 (0, 1, 0, 0)	0.11 %	2 (0, 2, 0, 0)	0.00029 %
	Rhodopseudomonas-like	1 (1, 0, 0, 0)	0.11 %	2 (2, 0, 0, 0)	0.00029 %
	Sphingomonas-like	1 (0, 1, 0, 0)	0.11 %	1 (0, 1, 0, 0)	0.00015 %
BchL/ChlL	Cyanobacteria-like	111 (94, 88, 13, 30)	56.62 %	91746 (72566, 18627, 68, 485)	76.57 %
	Trebouxiophyceae-like	32 (2, 4, 23, 30)	15.46 %	25715 (7, 22, 9867, 15819)	21.46 %
	BchL UT 5	11 (0, 0, 11, 2)	5.31 %	440 (0, 0, 412, 28)	0.37 %
	BchL UT 4	6 (0, 1, 6, 1)	2.90 %	1066 (0, 1, 1063, 2)	0.89 %
	BchL UT 3	6 (0, 0, 6, 0)	2.90 %	102 (0, 0, 102, 0)	0.085 %
	Phormidesmis priestleyi Ana-like	5 (5, 2, 0, 0)	2.42 %	270 (265, 5, 0, 0)	0.23 %
	Burkholderiales-like	5 (2, 0, 4, 1)	2.42 %	103 (5, 0, 97, 1)	0.086 %

	Cluster/Separate OTU	No. of OTUs	OTUs (%)	No. of reads	Reads (%)
	Bradyrhizobium sp. S23321-like	4 (0, 4, 0, 0)	1.93 %	146 (0, 146, 0, 0)	0.12 %
	BchL UT 2	4 (4, 1, 3, 0)	1.93 %	77 (6, 2, 69, 0)	0.064 %
	Sphingomonadales-like	4 (2, 2, 2, 0)	1,93 %	59 (40, 5, 14, 0)	0.049 %
	Sphingomonas-like	3 (1, 3, 0, 0)	1.45 %	10 (2, 8, 0, 0)	0.0084 %
	Kouleothrix aurantiaca-like	3 (0, 3, 0, 0)	1.45 %	8 (0, 8, 0, 0)	0.0067 %
	Methylobacterium-like	2 (0, 1, 2, 1)	0.97 %	12 (0, 2, 6, 4)	0.010 %
	Aquincola tertiaricarbonis-like	2 (2, 0, 1, 0)	0.97 %	12 (10, 0, 2, 0)	0.010 %
	Belnapia-like	2 (2, 0, 0, 0)	0.97 %	8 (8, 0, 0, 0)	0.0067 %
	BchL UT 1	2 (0, 2, 0, 0)	0.97 %	3 (0, 3, 0, 0)	0.0025 %
	Mesorhizobium-like	1 (1, 0, 1, 1)	0.48 %	27 (1, 0, 4, 22)	0.023 %
	Rhodovulum sulfidophilum-like	1 (0, 1, 0, 0)	0.48 %	1 (0, 1, 0, 0)	0.00083 %
BchX	BchX UT 1	22 (0, 6, 22, 5)	47.83 %	4453 (0, 34, 4125, 294)	89.96 %
	Methyloversatilis-like	4 (0, 3, 1, 0)	8.70 %	14 (0, 4, 10, 0)	0.28 %
	Bradyrhizobium sp. S23321-like	3 (1, 3, 1, 1)	6.52 %	300 (4, 293, 1, 2)	6.06 %
	BchX UT 2	3 (0, 1, 3, 1)	6.52 %	37 (0, 1, 26, 10)	0.75 %
	Novosphingobium acidiphilum-like	3 (3, 1, 0, 0)	6.52 %	30 (23, 7, 0, 0)	0.61 %
	Belnapia-like	2 (0, 1, 2, 1)	4.35 %	39 (0, 2, 32, 5)	0.79 %
	Methylobacterium-like	1 (0, 0, 1, 0)	2.17 %	42 (0, 0, 42, 0)	0.85 %
	Citromicrobium-like	1 (0, 0, 1, 0)	2.17 %	2 (0, 0, 2, 0)	0.040 %
	Bradyrhizobium-like	1 (1, 1, 0, 0)	2.17 %	2 (1, 1, 0, 0)	0.040 %
	Kouleothrix aurantiaca-like	1 (0, 1, 0, 0)	2.17 %	1 (0, 1, 0, 0)	0.020 %
	Methylocella silvestris-like	1 (0, 1, 0, 0)	2.17 %	1 (0, 1, 0, 0)	0.020 %
	Thiorhodospira sibirica-like	1 (0, 1, 0, 0)	2.17 %	1 (0, 1, 0, 0)	0.020 %
	Roseobacter denitrificans-like	1 (1, 0, 0, 0)	2.17 %	1 (1, 0, 0, 0)	0.020 %

4.4. Discussion

Given the distance to the ocean (~200 km) and the extent of the surrounding ice cover, bacteria in exposed soils of the Sør Rondane Mountains, East Antarctica, are faced with very low availability of organic matter [38, 45, 275] and might thus be expected to use alternative energy sources such as sunlight. During an initial survey using libraries of ~100 clones, phototrophy genes were investigated for the first time in this terrestrial Antarctic location [277]. Analysis of partial PufM sequences revealed diversity predominantly associated with phylotypes from aerobic anoxygenic photoheterotrophic Alphaproteobacteria. The most abundant ion pumping microbial rhodopsin gene family, proteorhodopsin, however, could not be amplified from the samples [277]. In the present study we used high-throughput Illumina MiSeq paired-end 300 bp sequencing to more comprehensively study the presence and diversity of genes involved in light-harvesting, in the same samples.

To target a wider diversity of anoxygenic phototrophs using the photosynthetic type 2 reaction center, the more universal pufM_uniF/pufM_WAW primer set [232] was used. Using our *pufLM* database assembled from publicly available sequences, an *in silico* comparison of these primers with other less degenerate primers (*i.e.* pufM_557F, pufMR and pufM_750R) used in other studies [228, 241] clearly showed that the Yutin et al. primer set [232] targets a much wider *pufM* diversity (Figure S5). In addition, contrary to the clone library results [277], the primer set used here gave successful amplification in all samples. As observed in the clone libraries, alphaproteobacterial-like PufM sequences were most frequently recovered (98.60 %), followed distantly by gammaproteobacterial-like sequences (0.64 %). Deep sequencing also revealed presence of some betaproteobacterial-like (0.012 %) and even chloroflexi-like (0.38 %) PufM (Figure 3). The dominance of alphaproteobacterial PufM sequences has been previously observed in the Arctic [205] and Antarctic [175, 233], whereas chloroflexi-like PufM sequences have not previously been reported from polar or most other environments studied so far. Also in Arctic soils [205], gammaproteobacterial-like PufM sequences were found to contribute less to the general diversity.

The relative abundance of *Roseobacter*-like and *Loktanella*-like PufM in our samples is remarkably high (Figure 1). PufM from these two AAP-containing taxa have previously nearly exclusively been reported from marine and saline lake environments from polar and non-polar regions [174, 175, 226, 229, 230, 272, 323-326]. The high relative abundance of

Roseobacter-like PufM sequences (82.39 %) in our terrestrial samples is therefore striking. *Roseobacter* has been found to be important in sulfur cycling in aquatic environments [327-329]. However, the absence of *Roseobacter* 16S rRNA gene sequences in this study area [275] and in terrestrial Antarctic systems in general (based on metagenome data available in NCBI, MG-RAST [207] and IMG/M [208]) suggests the presence of other microorganisms, containing PufM highly similar to that of *Roseobacter*, in our samples. The remainder of the PufM diversity in our samples, although recovered in small relative numbers, also mainly related to aquatic photoheterotrophic taxa. However, several of the PufM sequences recovered (*e.g. Methylobacterium*-like, *Rhodopseudomonas*-like) were highly similar to PufM reported from Arctic soils [205] or Chinese paddy soils [149, 257, 258, 268]. Thus although aerobic anoxygenic phototrophy is frequently studied in aquatic environments, our data strongly suggest that this lifestyle may potentially be important in terrestrial ecosystems.

The primers [279] previously used to amplify a broad diversity of *nifH* sequences also amplified structurally similar oxidoreductase subunits involved in (bacterio)chlorophyll synthesis (*bchL*/*chlL* encoding for the L subunit of DPOR in APB, Cyanobacteria, green algae and lower land plants, and *bchX* encoding for a COR subunit of APB). The high relative abundance (98.26 %) of cyanobacterial plus Trebouxiophyceae green algal ChlL suggested an important role for oxygenic photosynthetic organisms in our samples (Table 4, Figures 4a and S4). Notably, an inverse pattern was observed: nearly all of the ChlL reads recovered from samples KP2 (99.89 %) and KP15 (98.92 %) grouped with *Chroococcales* and *Oscillatoriales* Cyanobacteria, respectively (Table S1), with very few Trebouxiophyceae-like sequences (<0.12 %). KP43 and KP53, on the other hand, contained far less cyanobacterial reads (0.58 and 2.96 % respectively) and a very high relative abundance of Trebouxiophyceae-like ChlL (84.23 and 96.69 % respectively) (Figure 1, Table S1). A similar pattern was observed previously [277] [Tahon et al. Submitted (Chapter 3)]: a high relative abundance of both cyanobacterial cbbL type IB (RuBisCO) and 16S rRNA gene sequences, grouping with *Chroococcales* and *Oscillatoriales* Cyanobacteria, was recorded from the KP2 and KP15 samples, respectively, and much less in KP43 and KP53. A high relative number of trebouxiophyceael chloroplast 16S rRNA gene sequences was recorded from the latter samples [Tahon et al. Submitted (Chapter 3].

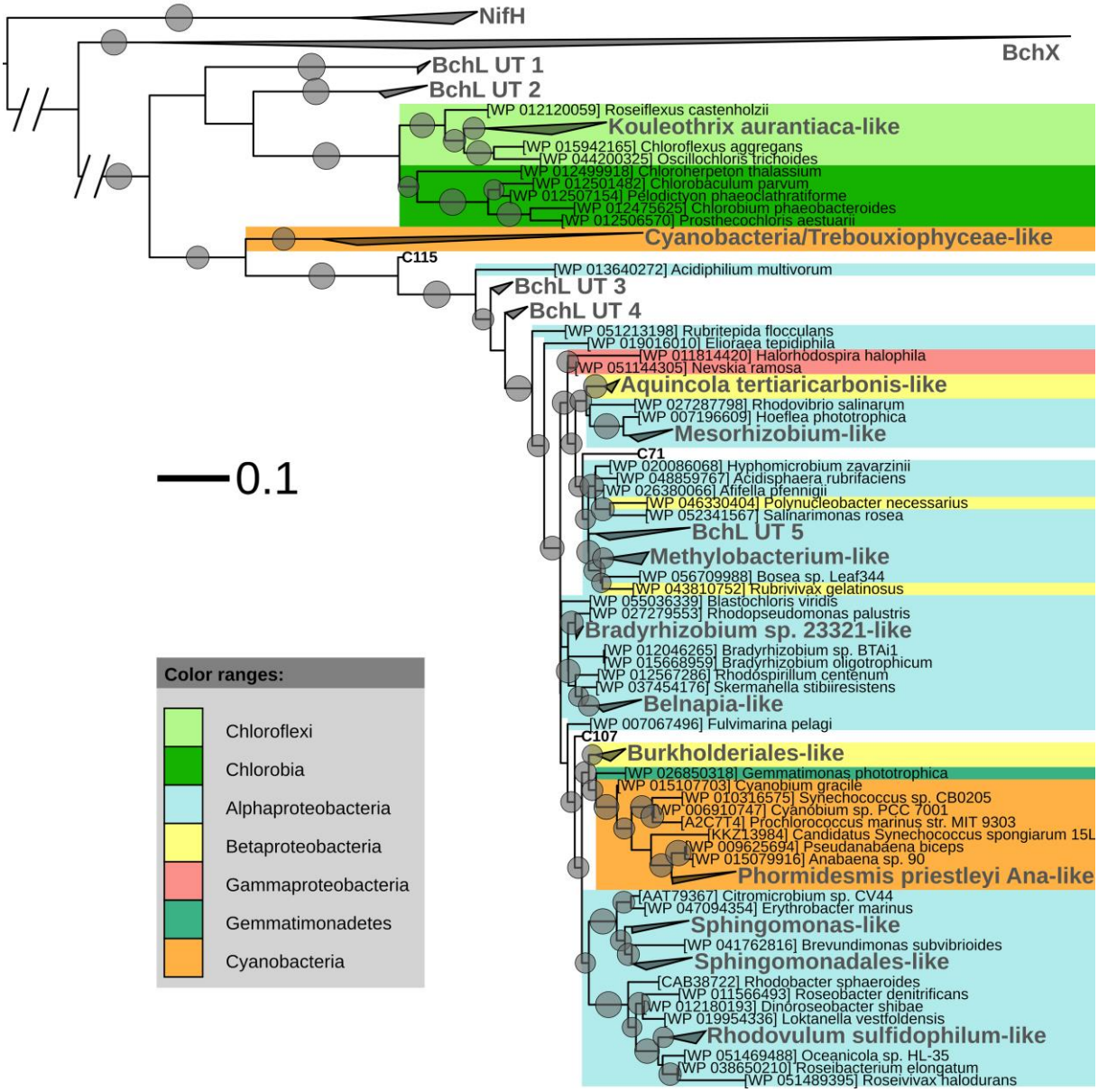

Figure 4a ML phylogenetic tree (1000 bootstraps) of BchL/ChlL sequences. NifH sequences were used as an outgroup. Scale bar indicates 0.1 substitutions per amino acid position. OLU clusters (larger font size, labelled in grey) were named after cultivated bacteria that grouped in or close to the cluster. In the absence of cultivated members, clusters were given an Utsteinen (UT) cluster number designation. For clusters, the total branch lengths to the closest and farthest leaf of the cluster were used as sides of the triangle. OLUs not enclosed in clusters are labelled in bold. For reference data, taxon name and accession number is listed. Bootstrap values of at least 70 % are displayed as circles with a diameter reflecting the height of the bootstrap value. BchX sequences are shown as a single cluster (details given in Figure 4b)

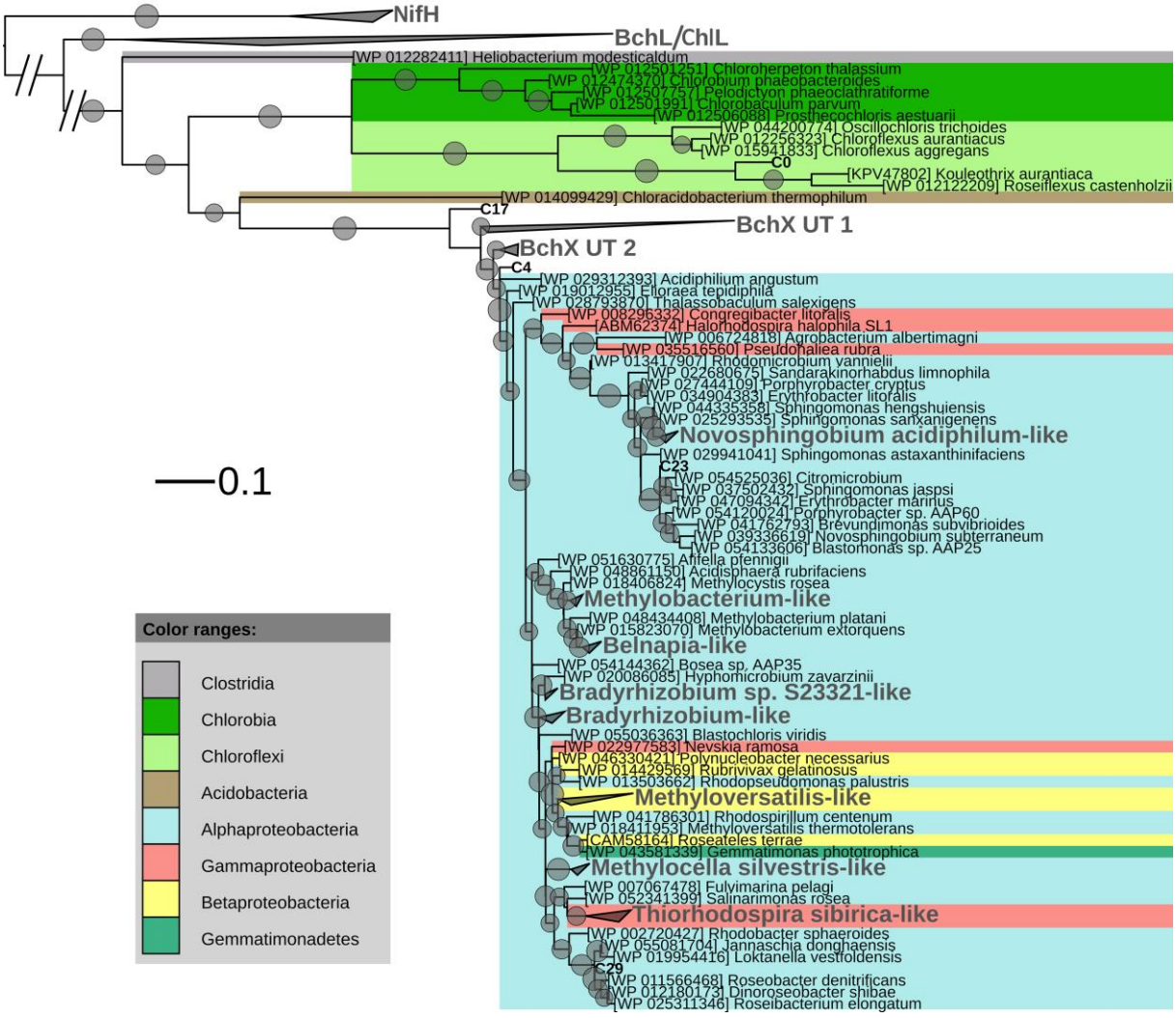

Figure 4b BchX ML phylogenetic tree (1000 bootstraps). NifH sequences were used as an outgroup. Scale bar indicates 0.1 substitutions per amino acid position. OXUs not enclosed in a cluster are labelled in bold. OXU clusters (larger font size, labelled in grey) were named after cultivated bacteria that grouped in or close to the cluster. In the absence of cultivated members, clusters were given an Utsteinen (UT) cluster number designation. For clusters, the total branch lengths to the closest and farthest leaf of the cluster were used as sides of the triangle. For reference data, taxon name and accession number is listed. Bootstrap values of at least 70 % are displayed as circles. Smallest circles represent the lower cut-off of 70 % with a diameter reflecting the height of the bootstrap value. BchL/ChlL sequences are shown as a single cluster (details given in Figure 4a)

However, as the primers were originally designed to amplify a broad diversity of *nifH* [279], and not *bchL*/*chlL*, it is conceivable that they might show a bias towards particular groups. Therefore, an *in silico* analysis, using a broad diversity of publicly available sequences, was performed to investigate possible primer bias. The IGK3 and DVV primers [279] (Table 2) generally showed one or two mismatches, located at the 5' primer end, with *bchL*/*chlL* reference data of different phyla (Figure S5). Chlorobi sequences, however, showed most mismatches (four and one in DVV and IGK3, respectively). The absence of clear differences

in primer specificity for different groups suggests there is unlikely to be extensive bias in the primers. Thus the high relative abundance of oxygenic photosynthetic microorganisms in ChlL suggests they indeed appear to be an important phototrophic group in the investigated samples.

In addition to the highly recovered cyanobacterial and green algal ChlL, a small number of non-cyanobacterial BchL sequences (1,74 % - 2091 sequences, including 25 singletons) grouped with a broad diversity of mainly aerobic anoxygenic phototrophic bacteria primarily belonging to alphaproteobacterial taxa and to lesser extent betaproteobacterial and chloroflexi taxa (Figure 4a).

The IGK3 and DVV primers also amplify *bchX*. Because this gene is not present in Cyanobacteria or in Trebouxiophyceae, the BchX dataset was relatively small (4950 sequences). The alternative explanation that primer mismatch might have reduced the number of sequences recovered, is less likely. Indeed, *in silico* analysis showed that the number of mismatches with a set of representative *bchX* sequences was limited (Figure S5). The low number of sequences precludes firm conclusions regarding BchX diversity. The BchX sequences mainly grouped with alphaproteobacterial taxa and to a lesser extent with Betaproteobacteria, Gammaproteobacteria and Chloroflexi (Figure 4b), however, with greater sequencing depth, relative abundances may change. Most of the BchX reads grouped into two clusters (BchX UT 1 and 2) without any close known representative, indicating the existence of multiple not yet cultured or recognized APB. Indeed many anoxygenic phototrophs can grow heterotrophically and it is thus possible that some taxa, originally described as regular heterotrophs, may have phototrophic capacities that have not been noticed. Indeed, *Salinarimonas rosea* DSM 21201 originally tested negative for bacteriochlorophyll *a* synthesis [330] and the phototrophic capacities of *Skermanella stibiiresistens* SB22 were originally not reported [331], whereas more recent analysis of their genomic sequences (accession numbers AUBC01000000 and AVFL01000000) revealed their phototrophic potential.

Finally, since BchL/ChlL, BchX and NifH exhibit a high degree of protein sequence similarity [108], and their genes can be retrieved using the same primer set, it may be a challenge to correctly annotate these sequences. During our analyses, we noticed that public databases contain several BchL/ChlL and BchX sequences annotated as NifH, and *vice versa*.

When studying these genes and including reference data, it is therefore important to take into account gene specific conserved amino acid positions to ensure correct interpretation of data. Comparing the datasets for different genes is difficult as datasets obtained with different primers cannot be compared directly because of differences in primer specificity, PCR efficiency or bias. Nevertheless, because the large sampling depth of PufM complicates evaluation, we have normalized PufM and BchL/ChlL datasets (Table 3) to tentatively allow a rough comparison. This shows that pufM diversity is somewhat lower though of similar magnitude than that of BchL/ChlL (171 OPUs – 192 OLUs). It should be noted that horizontal gene transfer may cause discrepancies between phylogenies of 16S rRNA or *cbbL*, and photosynthesis genes [332]. However, tentatively, it can be noted that several of the APB taxa retrieved here were previously reported from 16S rRNA gene (*e.g. Bradyrhizobium*, *Sphingomonas*, *Afifella*, *Methylibium*) and *cbbL* (RuBisCO) (*e.g. Mesorhizobium*, *Bradyrhizobium*, *Methyloversatilis*, *Rhodospirillum centenum*) clone library and Illumina sequencing results from the same samples [277] [Tahon et al. Submitted (Chapter 3)]. Interestingly, the relative abundance of *Bradyrhizobium*-related sequences from sample KP15 was much higher than from the other samples in the three photosynthetic datasets (Table 4), as well as the 16S rRNA and cbbL gene sequence datasets [277] [Tahon et al. Submitted (Chapter 3)]. Furthermore, the genome of *Bradyrhizobium* sp. S23321 – the closest neighbor to most of our *Bradyrhizobium*-related sequences – revealed a gene content adapted to survival in a broad range of environments [333]. The combination of these data thus suggests photoautotrophic bradyrhizobia may be present in sample KP15.

In the pilot study proteorhodopsin could not be detected [277]. Actinorhodopsin, a similar light-driven proton pump, was originally retrieved from aquatic Actinobacteria [199] and little is known about its presence in terrestrial environments. Because Illumina data of partial 16S rRNA genes previously showed the samples investigated here to contain diverse Actinobacteria (6.26 – 23.48 % of reads and 20.73 – 34.60 % of OTUs) [275] [Tahon et al. Submitted (Chapter 3)], we used several primer sets (Table 2) to amplify actinorhodopsin genes from our samples. While our attempts failed, this does not necessarily imply that these systems are absent in the terrestrial Antarctic bacterial communities. Currently available primers may be unsuitable to capture all actinorhodopsin diversity, as most reference data originates from aquatic systems [199, 201, 202, 334, 335]. Future metagenome datasets may resolve this question.

4.5. Conclusion

We studied the presence of bacterial phototrophic pathways in a terrestrial Antarctic environment. While we could not detect actinorhodopsin genes, our analysis of other genes showed that a broad variety of oxygenic and anoxygenic phototrophs is present in soils of the Sør Rondane Mountains, East Antarctica. The high relative abundance of oxygenic photosynthetic microorganisms, however, suggests they are an important phototrophic group. Sequencing results of BchL, ChlL and BchX, involved in (bacterio)chlorophyll synthesis, were dominated either by Cyanobacteria- or Trebouxiophyceae-related sequences. Moreover, the presence of currently unknown non-cyanobacterial phylotypes suggests the existence of multiple not yet cultured or recognized anoxygenic phototrophic bacteria. Illumina Miseq sequencing of PufM, typical for light-harvesting bacteria with a type 2 reaction center, revealed a very high relative abundance of two groups of sequences, *i.e. Roseobacter*-like and *Loktanella*-like, and a large diversity of other less abundant taxa from Alpha-, Beta- and Gammaproteobacteria, Chloroflexi and several unassigned groups. Although photoheterotrophic bacterial light-harvesting is nearly exclusively studied in aquatic environments, our results suggest the potential relevance of this mechanism in terrestrial ecosystems.

4.6. Acknowledgements

This work was supported by the Fund for Scientific Research – Flanders (project G.0146.12). Additional support was obtained from the Belgian Science Policy Office (project CCAMBIO). The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the Flemish Supercomputer Center (VSC) funded by Ghent University, the Hercules Foundation and the Flemish Government – department EWI. This study is a contribution to the State of the Antarctic Ecosystem (AntEco) research program of the Scientific Committee on Antarctic Research (SCAR).

4.7. Supplementary information

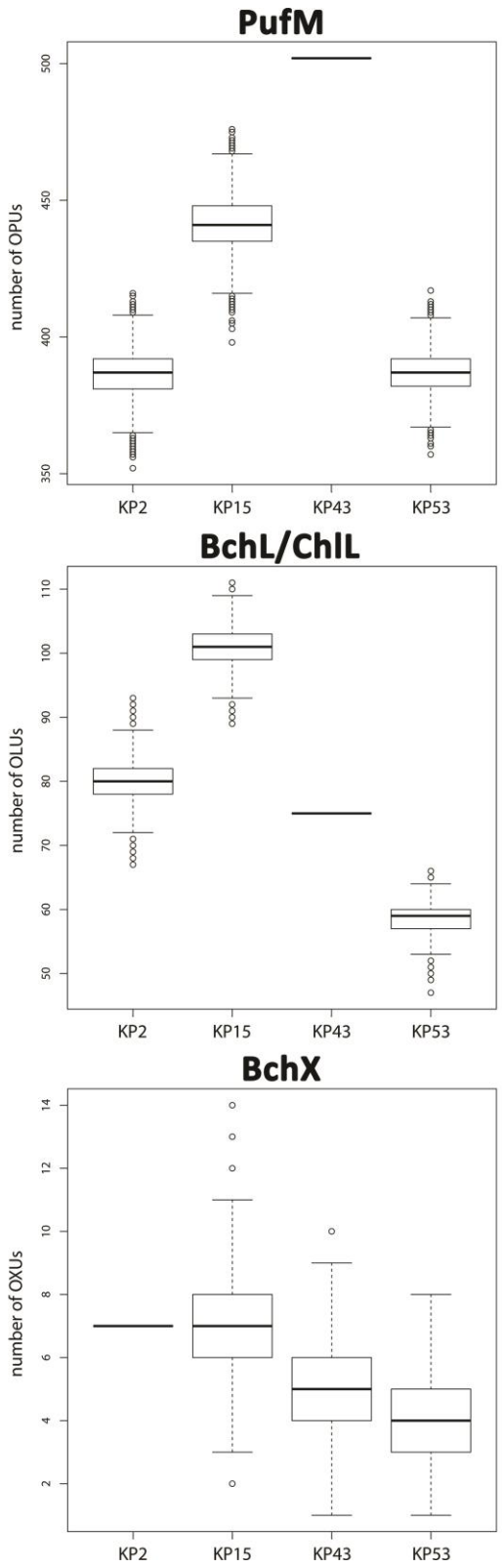

Figure S1 Box plots showing variation in OPUs/OLUs/OXUs recovered after standardization. For each gene, box plots were generated in R (https://cran.r-project.org/) using 10,000 iterations of the non-normalized OTU table

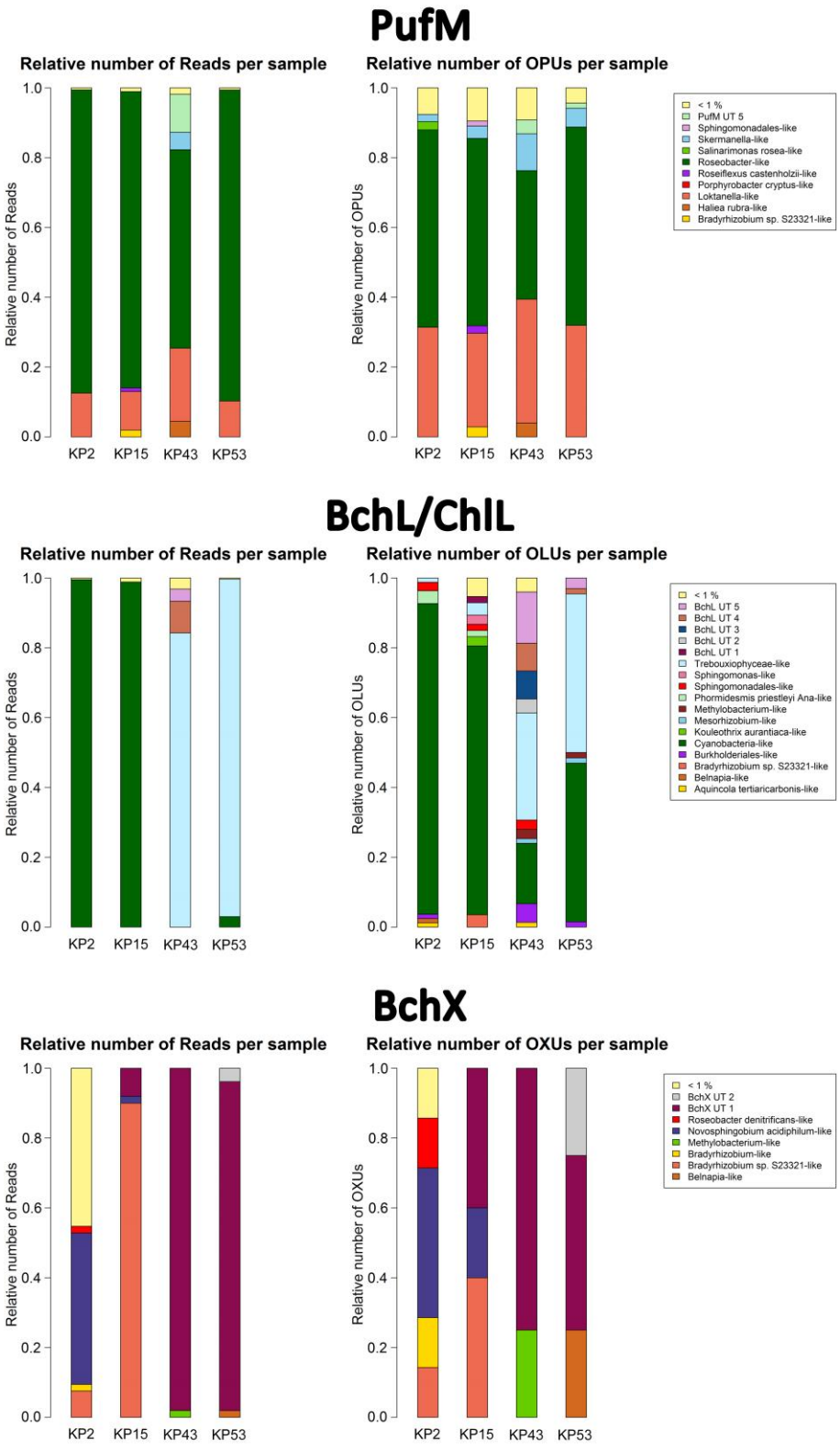

Figure S2 Bar plots showing relative numbers of reads and OPUs (PufM), OLUs (BchL/ChlL) or OXUs (BchX) per cluster. Data were normalized (non-normalized bar plots are shown in Figure 1). Clusters or separate OPUs/OLUs/OXUs containing less than 1% of the data were grouped together in the < 1% group

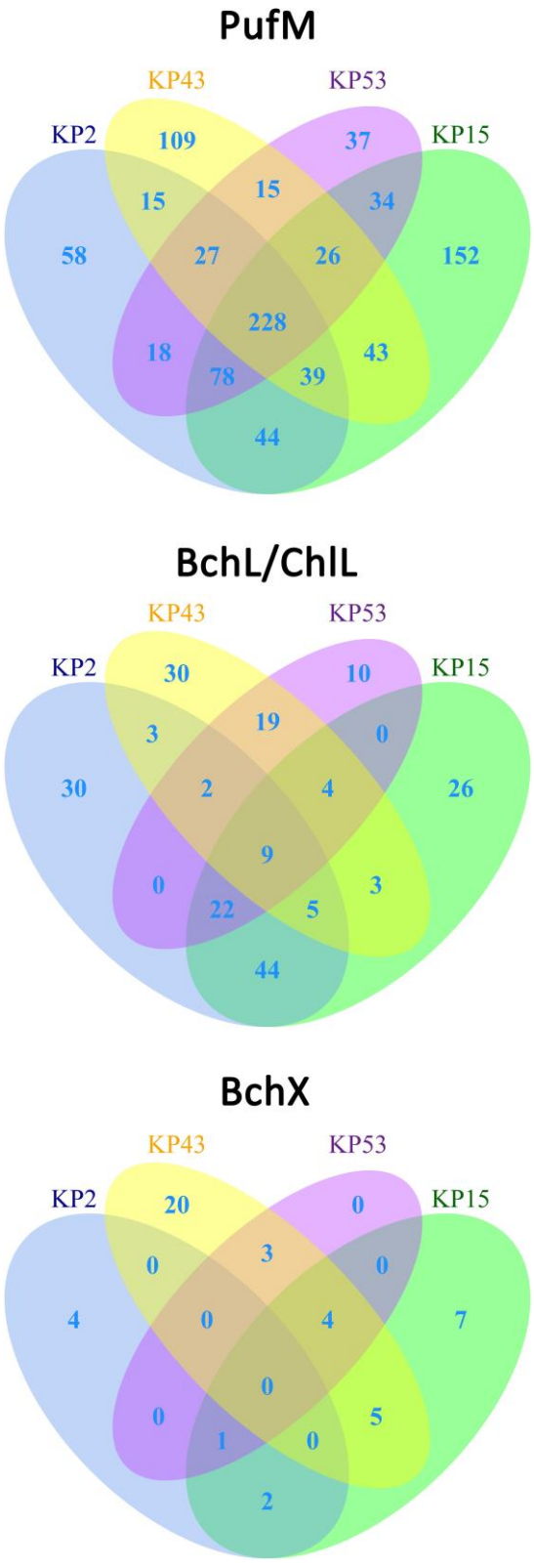

Figure S3 Unweighted Venn diagrams of OPUs/OLUs/OXUs. Venn diagrams were calculated using the VennDiagram package (http://cran.r-project.org/web/packages/VennDiagram/index.html) in R. Colored areas show the number of OPUs/OLUs/OXUs present in the terrestrial samples or shared between multiple samples

Figure S4 Detailed view of the Cyanobacteria-like ChlL cluster from Figure 4a. Scale bar indicates 0.01 substitutions per amino acid position. Bootstrap values of at least 70% are displayed as circles. Smallest circles represent the lower cut-off of 70% with a diameter reflecting the height of the bootstrap value

Figure S4 is available at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.02026/full

Figure S5 Alignment of mismatches of *pufM*, *bchL*/*chlL* and *bchX* genes with primers used in this study. Sequences of a selection of different phototrophic organisms are included. Accession numbers are given between brackets. Taxon (phylum or class) is given next to each organism name. Primers used are given in blue; for *pufM*, primers frequently used in literature are shown in black. Reverse primers are shown as reverse complementary sequences. Mismatches are labelled in red. For *pufM* primer pufM 750R, additional mismatches on position 768, compared to primer pufMR, are labelled in green. Nucleotide positions are relative to *Loktanella vestfoldensis* (NZ_KB907976)

Table S1 Composition of OPUs, OLUs and OXUs. Each worksheet is named after the gene analyzed and shows the number of sequences per sample for each OPU (PufM), OLU (ChlL/BchL) or OXU (BchX), as well as the cluster the OPU/OLU/OXU groups in. OPUs/OLUs/OXUs not enclosed in a cluster are labeled as Separate. For cyanobacterial OLUs, the order is given between brackets

Table S1 is available at:

http://journal.frontiersin.org/article/10.3389/fmicb.2016.02026/full#supplementary-material

Table S2 Habitat origin of nearest neighbors per named cluster, or separate OTU (95 % amino acid similarity) with a close cultured representative. For environmental studies, only one representative was used per cluster. ENV = environmental. For the Cyanobacteria/Trebouxiophyceae-like ChlL cluster Cyanobacteria-like and Trebouxiophyceae-like sequences are listed separately

	Cluster/Separate OTU	**Closest neighbor(s)**		
		Accession no.	**Origin**	**Similarity (%)**
PufM	*Acidisphaera rubrifaciens*-like	NZ_BANB01000656	*Acidisphaera rubrifaciens*; Japan; Hot spring surface water and sediment	90.8
	Afifella-like	NZ_JHZK01000006	*Afifella pfennigii*; French Polynesia; Benthic microbial mat	95.4
	Belnapia-like	NZ_KN676113	*Belnapia* sp. F-4-1; China; Soil	85.9-93.8
		NZ_JQKB01000015	*Belnapia moabensis*; USA; Biological soil crust	
	Bradyrhizobium sp. S23321-like	AP012279	*Bradyrhizobium* sp. S23321; Japan; Paddy field soil	92.3-100
		JN248466	ENV; Arctic ocean; Water (60 m deep)	
		KC768186	ENV; Antarctica; Surface water	
		KF008147	ENV; Antarctica; Surface water	
		AB510451	ENV; China; Paddy soil	
		KM654597	ENV; Norwegian and Barents sea; Water	
	Brevundimonas-like	NC_014375	*Brevundimonas subvibrioides*; USA; Freshwater	95.4
		NZ_JNIX01000000	*Brevundimonas bacteroides*; USA; Freshwater	
	Haliea rubra-like	KC253226	*Haliea rubra*; France; Surface water	90.8-95.3
	Hyphomicrobium-like	NZ_KB911258	*Hyphomicrobium zavarzinii*; Russia; Soil	93.8-95.4
	Ideonella-like	NZ_BBYR01000007	*Ideonella* sp. 201-F6; Japan; Sediment	93.8
	Loktanella-like	NZ_KB907976	*Loktanella vestfoldensis*; Antarctica; Lake microbial mat	90.0-100
		JN378830	*Loktanella* sp. RCC2403; Arctic ocean; Water (3 m deep)	
		JN248513	ENV; Arctic ocean; Water (5 m deep)	
		KC900127	ENV; Antarctica; Surface water	
		JF421745	ENV; Mediterranean seawater	
		GU079560	ENV; USA; Seawater	
		KF008169	ENV; Antarctica; Surface water	
		GQ468949	ENV; Mediterranean seawater	
	Methylibium-like	LGRD01000013	*Methylibium* sp. NZG; China; *Bangia atropurpurea*	96.9

Cluster/Separate OTU	Closest neighbor(s)		
	Accession no.	Origin	Similarity (%)
	KC900142	ENV; Antarctica; Surface water	
Methylobacterium-like	NZ_KI912577	*Methylobacterium* sp. 10; USA; Lake water	92.3-98.5
	NZ_KB910516	*Methylobacterium* sp. 77; USA; Lake water	
	NZ_AGJK00000000	*Methylobacterium extorquens*; Finland; *Pinus sylvestris*	
	AP014705	*Methylobacterium aquaticum*; Spain; Drinking water	
	NZ_JTHG01000063	*Methylobacterium platani*; South Korea; *Platanus orientalis* leaf	
	JN248518	ENV; Arctic ocean; Water (80 m deep)	
Methylocystis-like	NZ_KB889963	*Methylocystis rosea*; Norway; Wetland soil	93.8-95.2
	NZ_AYNA01000017	*Methylocystis* sp. SB2; USA; Spring bog	
	KF494876	ENV; Norway; Seawater	
	AB510462	ENV; China; Paddy soil	
Methyloversatilis-like	NZ_AFHG00000000	*Methyloversatilis universalis*; USA; Lake water	95.4
	NZ_ARVV01000001	*Methyloversatilis discipulorum*; USA; Lake water	
	NZ_KB889967	*Methyloversatilis thermotolerans*; USA; Lake water	
Porphyrobacter cryptus-like	NZ_KE383966	*Porphyrobacter cryptus*; Portugal; Hot spring	89.2 -93.8
	AB510465	ENV; China; Paddy soil	
Rhodopseudomonas-like	NZ_LJIC01000210	*Rhodopseudomonas* sp. AAP120; China; Freshwater lake	98.5
Rhodospirillum centenum-like	CP000613	*Rhodospirillum centenum*; USA; Hot springs	92.2-93.8
Rhodovulum sulfidophilum-like	AB020784	*Rhodovulum sulfidophilum*; Holland; Marine mud	89.2-95.3
Roseicitreum antarcticum-like	JQ995753	*Roseicitreum antarcticum*; Antarctica; Sandy intertidal sediment	96.9-100
	KF487001	ENV; Norway; Water	
	KC900116	ENV; Antarctica; Surface water	
Roseiflexus castenholzii-like	AB095768	*Roseiflexus castenholzii*; Japan; Microbial mat in hot spring	70.3-78.1
Roseobacter-like	CP000362	*Roseobacter denitrificans*; Japan; *Enteromorpha linza*	69.2-100
	CP002624	*Roseobacter litoralis*; Seaweed	
	X57597	*Erythrobacter* sp. OCH114	
	NZ_JWLL01000010	*Tateyamaria* sp. ANG-S1; USA; *Euprymna scolopes*	
	NZ_JYFE01000060	*Jannaschia aquimarina*; South Korea; Seawater	
	JF421742	ENV; Mediterranean seawater	
	JQ340526	ENV; Pacific Ocean; Surface water	
	KM654595	ENV; Norway; Seawater	
	AB510458	ENV; China; Paddy soil	
Salinarimonas rosea-like	NZ_AUBC00000000	*Salinarimonas rosea*; China; Salt mine soil	81.3-92.2
	KC900120	ENV; Antarctica; Surface water	
Skermanella-like	AVFL01000001	*Skermanella stibiiresistens*; China; Iron mine soil	81.5-93.8
	NZ_AVFK00000000	*Skermanella aerolata*; South Korea; Air	
Sphingomonadales-like	AYSC01000048	*Blastomonas* sp. CACIA14H2; Brazil; Water	90.8-96.9
	NZ_LJYW01000001	*Prosthecomicrobium hirschii*; USA; Freshwater pond	
	NZ_ATVO01000008	*Sandarakinorhabdus limnophila*; Germany; Freshwater lake	

	Cluster/Separate OTU	Closest neighbor(s)		
		Accession no.	**Origin**	**Similarity (%)**
		JRVC01000019	*Novosphingobium subterraneum*; USA; Subsurface core (180 m deep)	
		NZ_KK073876	*Sphingomonas jaspsi*; Japan; Freshwater	
		JQ340545	ENV; Pacific Ocean; Surface water	
	Sphingomonas-like	NZ_JPJC00000000	*Sphingomonas* sp. FUKUSWIS1; Germany; Lake water	95.3
		NZ_JPDP00000000	*Sphingomonas* sp. STIS6.2; Germany; Lake water	
	PufM UT 1	JN248492	ENV; Arctic ocean; Surface water	89.2-92.2
	PufM UT 2	JN248492	ENV; Arctic ocean; Surface water	92.2-96.8
	PufM UT 4	JQ340684	ENV; Pacific Ocean; Surface water	93.8
		GQ468982	ENV; Mediterranean seawater	
	PufM UT 5	AB510461	ENV; China; Paddy soil	90.7-93.8
BchL/ChlL	Cyanobacteria-like	ALF54602	*Nostoc piscinale* CENA21; Brazil; Solimoes river	75.9-98.1
		Q8DGH0	*Thermosynechococcus elongatus* BP-1; Japan; Hot spring	
		WP_012410753	*Nostoc punctiforme*; Australia; *Macrozamia* sp. root	
		WP_011142366	*Gloeobacter violaceus*; Switzerland; Calcereous rock	
		WP_006634709	*Microcoleus vaginatus*; USA ; Desert soil crust	
		WP_015204277	*Crinalium epipsammum*; The Netherlands; Sandy crust	
		AIQ80516	ENV; Glacier foreland soil	
		AIC84974	ENV; USA; Strip mine lake	
		CEK40168	ENV; Pakistan; Rhizosphere of *Pinus roxburghii*	
	Trebouxiophyceae-like	ABX82598	chloroplast *Trebouxia aggregata*	96.2-100
		YP_009104867	chloroplast *Myrmecia israelensis*	
		AGZ19374	chloroplast *Chlorella* sp. ArM0029B; Arctic	
		NP_045884	chloroplast *Chlorella vulgaris*; water	
		YP_009104838	chloroplast *Symbiochloris reticulata*;	
	BchL UT 5	AIQ80532	ENV; Glacier foreland soil	86.1-92.6
	Phormidesmis priestleyi Ana-like	KPQ34041	*Phormidesmis priestleyi* Ana; USA; Microbial mat	91.7-99.1
	Burkholderiales-like	WP_056902525	*Pseudorhodoferax* sp. Leaf274; Switzerland; *Arabidopsis* leaf	96.3-100
		AGS08013	*Aquincola tertiaricarbonis*; Germany; contaminated groundwater	
		AIC84916	ENV; USA; Strip mine lake	
		AIQ80467	ENV; Glacier foreland soil	
	Bradyrhizobium sp. S23321-like	WP_015684485	*Bradyrhizobium* sp. S23321; Japan; Paddy field soil	100
		ERF83826	*Bradyrhizobium* sp. DFCI-1; USA; Human gut	
		AIQ80568	ENV; Glacier foreland soil	
	Sphingomonadales-like	WP_022680698	*Sandarakinorhabdus limnophila*; Germany, Lake water	93.5-96.3
		WP_017667988	*Sandarakinorhabdus* sp. AAP62; China; Lake water	
		WP_054135752	*Blastomonas* sp. AAP25; Czech Republic; Lake water	
	Sphingomonas-like	WP_010405207	*Sphingomonas echinoides*; Germany	96.3-97.2
		WP_031439826	*Sphingomonas* sp. FUKUSWIS1; Germany; Lake water	
		AIQ80498	ENV; Glacier foreland soil	

	Cluster/Separate OTU	Closest neighbor(s)		
		Accession no.	**Origin**	**Similarity (%)**
	Kouleothrix aurantiaca-like	KPV49864	*Kouleothrix aurantiaca*; Japan; Industrial waste water sludge	81.5-89.8
	Methylobacterium-like	WP_003603754	*Methylobacterium extorquens*; Finland; *Pinus sylvestris*	94.4-99.1
		WP_048432787	*Methylobacterium platani*; South Korea; *Platanus orientalis* leaf	
		WP_019903624	*Methylobacterium* sp. 77; USA; Lake water	
		AIQ80515	ENV; Glacier foreland soil	
	Aquincola tertiaricarbonis-like	WP_046114963	*Aquincola tertiaricarbonis*; Germany; contaminated groundwater	97.2
	Belnapia-like	WP_043338848	*Belnapia moabensis*; USA; Biological soil crust	95.3-95.4
		WP_043360866	*Belnapia* sp. F-4-1; Tibet; Soil	
	Mesorhizobium-like	WP_027033752	*Mesorhizobium loti*; Brazil; Zinc mine soil	95.4
		AAY59338	ENV; Canada; Soil sample near *Salix arctica*	
		AAY59364	ENV; Canada; Rhizosphere of *Dryas integrifolia*	
	Rhodovulum sulfidophilum-like	WP_042458897	*Rhodovulum sulfidophilum*; Holland; Marine mud	93.5
		ALK25180	ENV; Marine sample	
BchX	*Methyloversatilis*-like	WP_054124830	betaproteobacterium AAP99; China; Freshwater lake water	91.7-97.3
		WP_020165691	*Methyloversatilis discipulorum*; USA; Lake water	
		WP_008064756	*Methyloversatilis universalis*; USA; Lake sediment	
	Bradyrhizobium sp. S23321-like	WP_015684462	*Bradyrhizobium* sp. S23321; Japan; Paddy field soil	97.2-99.1
		CEL12566	ENV; Pakistan; Rhizosphere of wheat	
		AIQ80536	ENV; Glacier foreland soil	
	Novosphingobium acidiphilum-like	WP_028640508	*Novosphingobium acidiphilum*; Germany; Lake water	97.2
	Belnapia-like	WP_052389091	*Belnapia moabensis*; USA; Biological soil crust	96.3
		AIC84959	ENV; USA; Strip mine lake	
	Methylobacterium-like	WP_056467683	*Methylobacterium* sp. Leaf104; Switzerland; *Arabidopsis thaliana*	97.3
		WP_056479496	*Methylobacterium* sp. Leaf117; Switzerland; *Arabidopsis thaliana*	
		WP_019903647	*Methylobacterium* sp. 77; USA; Lake water	
	Citromicrobium-like	WP_054525036	*Citromicrobium*; China; Seawater	98.2
	Bradyrhizobium-like	WP_015668980	*Bradyrhizobium oligotrophicum*; Japan; Rice field soil	94.5
		WP_012046288	*Bradyrhizobium* sp. BTAi1; USA; Stem nodules of *Aeschynomene indica*	
	Kouleothrix aurantiaca-like	KPV47802	*Kouleothrix aurantiaca*; Japan; Industrial waste water sludge	82.5
	Methylocella silvestris-like	WP_012591062	*Methylocella silvestris*; Germany; Acidic forest cambisol	96.3
	Thiorhodospira sibirica-like	WP_006787504	*Thiorhodospira sibirica*; Russia; Lake water	89.0-90.8
		AIQ80570	ENV; Glacier foreland soil	
	Roseobacter denitrificans-like	WP_011566468	*Roseobacter denitrificans*; Japan; Fresh water	98.2
		ALK25195	ENV; Marine sample	
		AIC84928	ENV; USA; Strip mine lake	
		AAT77530	ENV; Mediterranean seawater	

Chapter 5

Isolation and characterization of aerobic anoxygenic phototrophs from exposed soils from the Sør Rondane Mountains, East Antarctica

Redrafted from:

Tahon G. and Willems A. (2017). Isolation and characterization of aerobic anoxygenic phototrophs from exposed soils from the Sør Rondane Mountains, East Antarctica. Systematic and Applied Microbiology, *under review*.

Author's contributions:

Conceived and designed the experiments: GT, AW. Performed the experiments: GT. Analyzed the data: GT. Wrote the paper: GT, AW. All authors approved the final manuscript.

Summary

We investigated the culturable aerobic phototrophic bacteria present in soil samples collected in the proximity of the Belgian Princess Elisabeth Station in the Sør Rondane Mountains, East Antarctica. Until recently, only oxygenic phototrophic bacteria (Cyanobacteria) were well known from Antarctic soils. More recent non-cultivation-based studies demonstrated the presence of anoxygenic phototrophs, and particularly aerobic anoxygenic phototrophic bacteria in these areas. We studied ~1000 isolates obtained after prolonged incubation in different growth conditions. They were characterized by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Representative strains were identified by sequence analysis of 16S rRNA genes. Over half of the isolates grouped among known aerobic anoxygenic phototrophic taxa, particularly with Sphingomonadaceae, *Methylobacterium* and *Brevundimonas*. A total of 330 isolates were tested for presence of key phototrophy genes. While rhodopsin genes were not detected, multiple isolates possessed key genes of the bacteriochlorophyll synthesis pathway. The majority of these potential aerobic anoxygenic phototrophic strains grouped with Alphaproteobacteria (*Sphingomonas*, *Methylobacterium*, *Brevundimonas*, *Polymorphobacter*, *Roseomonas*, *Aureimonas*), and, surprisingly, also with Actinobacteria (*Aquipuribacter*, *Nocardioides*, *Rhodococcus*) and Bacteroidetes (*Spirosoma*, *Hymenobacter*). Laboratory growth experiments revealed that light indeed benefits the growth of several of the AAP isolates, whereas others demonstrated better growth in the dark.

5.1. Introduction

Phototrophy represents one of the oldest and most important bacterial processes on Earth for which two mechanisms have been described. The simplest mechanism involves ion-pumping rhodopsin proteins [181, 184]. The last decade, environmental studies have revealed the enormous diversity of microbial rhodopsins. Although they comprise a diverse group of photoactive transmembrane proteins, the proteo- and actinorhodopsin proton pumping families, predominantly found in aquatic environments all over our planet, are by far the most abundant and widespread [189-191, 193, 196, 199, 201, 202, 336, 337].

The second, less widespread but more efficient bacterial phototrophy mechanism, relies on (bacterio)chlorophyll-containing photochemical reaction centers. Chlorophyll-dependent species are found solely in the Cyanobacteria. Anoxygenic phototrophic bacteria (APB) (*i.e.* those relying on bacteriochlorophyll (Bchl)) are found in the Acidobacteria, Chlorobi, Chloroflexi, Firmicutes, Gemmatimonadetes and Proteobacteria [137, 147]. Most known APB are aerobic anoxygenic phototrophs (AAP). These AAP do not contain carbon fixation enzymes [152] and use light as an auxiliary energy source for their mostly heterotrophic metabolism [146, 149].

Several genes encoding subunits of key enzymes in the (bacterio)chlorophyll synthesis pathway are well conserved among phototrophic bacteria. The dark-operative protochlorophyllide oxidoreductase enzyme complex is present in all known phototrophic bacteria using photochemical reaction centers. In Cyanobacteria the complex is encoded by *chlLNB* genes, whereas APB rely on the homologous *bchLNB* genes. Additionally, APB contain a second enzyme complex involved in the Bchl synthesis pathway: chlorophyllide oxidoreductase, encoded by *bchXYZ* genes [159, 162, 164]. For light-harvesting, the majority of APB relies on a type 2 photochemical reaction center. These reaction centers have a heterodimeric structure, with *pufL* and *pufM* encoding the conserved proteins. Hence, these genes have proven excellent markers to study APB diversity [146, 174, 175].

Previously, we reported the diversity of key protein encoding genes involved in (bacterio)chlorophyll- and rhodopsin-dependent phototrophy in exposed areas of the Sør Rondane Mountains (SRM), East Antarctica. This 220 km long wedge-shaped mountain chain (71° 30’ – 72° 40’ S, 22 – 28° E), located ~200 km inland from the King Haakon VII Sea, consists mainly of groups of mountains and individual nunataks (*i.e.* isolated mountain tops projecting above the surrounding ice layer), and contains ~900 km² of exposed surface area

[38, 45]. The investigated ice-free areas appear to be suitable habitats for phototrophic microorganisms, and especially AAP, due to the availability of sunlight, oxygen and a minimum of organic nutrients [277, 338]. Our results suggested the presence of a diverse AAP community, including novel AAP bacteria. However, since most of these bacteria are yet to be cultivated, their characteristics and biochemical potential remain unknown. Amplicon sequencing of 16S rRNA genes only provides insights into who is present, whereas such inventories of protein encoding genes can, in general, not be linked to certain bacteria due to possible horizontal gene transfer and gene duplication events [263, 339, 340]. Although metagenome sequencing may reveal functional potential, recreating and closing genomes from such data is hard due to genomic microheterogeneity. Furthermore, a considerable portion of genes cannot be assigned a function [341, 342]. Thus, while culture-independent methods can describe the functional capacities of whole microbial communities, isolation and characterization of Antarctic bacteria, and microorganisms in general, is of great scientific relevance for investigating ecophysiology and adaptive strategies and linking function to identity. Cultured microorganisms not only permit testing for certain functions (*e.g.* phototrophy) or expression of genes, but also the sequencing of their genome. In addition they may help extend identifications obtained from metagenome data [341, 343-345].

Previous research using deep sequencing of 16S rRNA genes revealed that oxygenic phototrophs (Cyanobacteria) are sometimes only present in relatively small numbers in soils of the SRM [275] [Tahon et al. Submitted (Chapter 3)]. In contrast, these environments seem to contain a broad diversity of anoxygenic phototrophs, whereas rhodopsins could not be detected [277, 338]. Therefore, in the present study, we focused on the isolation of APB, and more specifically AAP, from the same exposed samples used previously for non-cultivation based studies. Isolates were identified at the molecular level, screened for the presence of different phototrophy genes and functionally tested in laboratory experiments to assess influence of light on growth. This data, contributing to the bacterial characterization of exposed surface soils of the SRM, represents one of the first reports on cultivation of aerobic anoxygenic bacteria from continental Antarctica.

5.2. Materials and Methods

Sampling site and sampling method

As previously described [277, 338], top surface soil samples were collected aseptically from four exposed sites in the proximity of the Belgian Princess Elisabeth Station, Utsteinen, East Antarctica (71° 57' S, 23° 20' E, elevation 1382 m). Three samples (KP15, KP43 and KP53) were taken on the East part of Utsteinen nunatak, ~500 m north of the research station. For sample KP2 this was ~1.3 km south of the Belgian base, on the East part of Utsteinen ridge. Samples were stored in sterile polypropylene containers and stored at -20 °C upon collection until returned to the Laboratory for Microbiology (Ghent University, Belgium) where they were stored in a cold room facility (-20 °C).

Media and isolation of bacterial strains

For isolation of aerobic phototrophic microorganisms, two defined low nutrient media, one selective for aerobic photoautotrophs (PA) and one for aerobic photoheterotrophs (PH), were prepared. Media compositions were based on media previously used for the isolation of phototrophic bacteria [230, 346-351]. Both media contained 3.50 mM $K_2HPO_4.3H_2O$, 1.47 mM KH_2PO_4, 0.81 mM $MgSO_4.7H_2O$, 3.42 mM NaCl, 0.58 mM $CaSO_4.2H_2O$, 25 µM $Fe_2(SO_4)_3$, 69.6 nM $ZnSO_4.7H_2O$, 0.252 µM $MnCl_2.4H_2O$, 25.2 nM $CoCl_2.6H_2O$, 10 nM $CuCl_2.2H_2O$ and 25 nM $NiCl_2.H_2O$. No carbon sources were added to the PA medium. The PH medium was enriched with a mix of six different carbon sources (glucose, sucrose, sodium succinate, sodium pyruvate, sodium acetate and malate) frequently used for isolating phototrophic bacteria. Concentrations of carbon sources were set at 0.5 mM each to mimic the oligotrophic Antarctic environment. Nitrogen traces in the aforementioned components mimicked the low *in situ* Antarctic nitrogen conditions [220, 352]. Therefore no additional nitrogen source was added. To support growth of photodiazotrophs, 24.32 µM MoO_3 and 1 µM V_2O_5 were added to the media [353-356]. For solid media, 15 g L^{-1} Bacto agar (BD) was added. The final pH of both media was set to 7.0.

The isolation of phototrophic bacteria was performed as follows: a ten-fold dilution series (10^{-1} – 10^{-5}) was prepared for each sample, starting from one gram of aseptically weighed soil homogenized in 9 ml sterile liquid growth medium using a vortex. Finally, 100 µl of each

dilution was plated. For liquid enrichments, sterile 120 ml glass vials containing 20 ml of dilutions 10^{-2} – 10^{-5} were set up and sealed with a gas permeable seal (4titude).

Culture plates and liquid enrichments were set up in four replicates and incubated under aerobic atmosphere at 4 and 15 °C (two replicates each). For bacteria in Antarctic surface samples, residing in a dark freezer for several years may be similar to a long austral winter. To aid recovery of dormant bacteria, we devised an isolation procedure that would mimic the gradual light transition from winter to summer: Incubation was performed in illuminated incubators (13-15 µmol photons m^{-2} s^{-1}), starting at 0 hours of light per day for one week, and increasing by two hours day^{-1} every week until a maximum of 18 hours day^{-1} by week 10 to simulate the increasing day length over the transition from winter to summer at the sampling locations. From solid media, distinct colonies were purified from week 10 on. For liquid enrichments, one vial of every condition was non-continuously shaken (aeration once per day by gentle manual shaking during 10 s) and the other vial was not aerated. After 21 weeks, 1/100 dilutions were plated and incubated for up to seven months. For each condition, all colonies with distinct morphology were purified by seeding single colonies on new plates filled with the same medium.

MALDI-TOF mass spectrometry

Due to the very slow growth of the isolates and limited amount of available biomass, a modified version of the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) sample preparation protocol, as previously described by Wieme and colleagues [357] was used, using cell suspensions rather than cell extracts. One loop (1 µl yellow Looplast® loop, LP Italiana) of bacterial cells was suspended in 10 µl of sterile MilliQ water and mixed. If an isolate did not produce biomass equivalent to one loop, all biomass from a 4.5 cm ø petri dish was used. One µl of freshly prepared suspension was spotted in duplicate onto a 384 OptiTOF stainless steel MALDI plate (AB Sciex) and dried at room temperature. Afterwards, one µl of 0.5 % (w/v) α-cyano-4-hydroxycinnamic acid in 50:48:2 acetonitrile:water:trifluoroacetic acid solution was added to each spot and allowed to dry.

Acquisition of the bacterial fingerprints was done using a 4800 Plus MALDI TOF/TOF™ Analyzer (Applied Biosystems) in linear positive ion mode, as previously described [357]. Raw mass spectra were extracted as t2d files from the analyzer, imported in to the Data Explorer 4.0 software (Applied Biosystems) and converted to text files. Subsequently, these

text files were imported in BioNumerics 7.5 (Applied Maths) and transformed to fingerprints. The similarity between spectra was determined using Pearson's product moment correlation. Afterwards, spectra were clustered using unweighted pair group with arithmetic mean.

Identification and characterization of bacterial strains

Bacterial DNA was extracted using the alkaline lysis protocol [358] and stored at -20 °C until further processing. Gene fragments were amplified (Table 1) using a Veriti thermal cycler (Life Technologies). All amplification reactions were carried out in 25 µl reaction mixtures containing 1x Qiagen PCR buffer (Qiagen), 0.2 mM of each deoxynucleotide triphosphate, 0.625 units of Qiagen *Taq* polymerase (Qiagen) and forward and reverse primers (Table 1). Each reaction mixture received 2.5 µl of template DNA.

PCR products were purified using a Nucleofast 96 PCR cleanup membrane system (Macherey-Nagel) and Tecan Genesis Workstation 200 (Tecan). Sequencing of the protein encoding genes was performed using the amplification primers (Table 1). For 16S rRNA genes, sequencing primers used are listed by Coenye et al. [359] and Cleenwerck et al. [281]. Sequencing was carried out using a BigDye Xterminator™ purification kit (Applied Biosystems) and an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems).

Table 1 PCR primers and conditions used for screening different genes

Gene	Target	Primer	Sequence 5'-3'	Final concentration	Region	Amplicon size	Program
16S rRNA	Universal	pA[a]	AGA GTT TGA TCC TGG CTC AG	0.1 µM	8-1541[k]	± 1500 bp	95 °C (5 min); 3x 95 °C (1 min), 55 °C (2 min 15 s), 72 °C (2 min 15 s); 30x 95 °C (35 s), 55 °C (1 min 15 s), 72 °C (1 min 15 s); 72 °C (10 min)
		pH[a]	AAG GAG GTG ATC CAG CCG CA	0.1 µM			
pufLM	AAP	pufLF[b]	CTK TTC GAC TTC TGG GTS GG	0.2 µM	64-1612[l]	± 1500 bp	94 °C (3 min); 30x 94 °C (1 min), 60 °C (1 min), 72 °C (2 min); 72 °C (10 min) [236]
		pufMR[c]	CCA TSG TCC AGC GCC AGA A	0.2 µM			
pufM	Universal	pufM_uniF[d]	GGN AAY YTN TWY TAY AAY CCN TTY CA	1.0 µM	584-825[m]	± 240 bp	94 °C (4 min); 35x 94 °C (40 s), 49 °C (30 s), 72 °C (30 s); 72 °C (7 min)
		pufM_WAW[d]	AYN GCR AAC CAC CAN GCC CA	0.5 µM			
proteorhodopsin	Universal	PR-1aF[e]	GAT CGA GCG NTA YRT HGA RTG G	1.87 µM	340-665[n]	± 335 bp	94 °C (2 min), 30x 94 °C (30 s), 52 °C (30 s), 72 °C (30 s) ; 72 °C (7 min) [189]
		PR-1aR[e]	GAT CGA GCR TAD ATN GCC CAN CC	1.87 µM			
actinorhodopsin	Clade LG1 & LG2	LG-for[f]	TAY MGN TAY GTN GAY TGG	0.4 µM	283-614[o]	± 330 bp	95 °C (7 min), 45x 94 °C (30 s), 51.5 °C (1 min 30 s), 72 °C (30 s) ; 72 °C (10 min)
		LG1A-for[f]	MGN TAY ATH GAY TGG YT	0.4 µM			
		LG2-for[f]	TAY MGN TAY GCN GAY TGG	0.4 µM			
		LG-rev[f]	ATN GGR TAN CAN CCC CA	0.8 µM			
nifH, *bchL*, *bchX*	Universal	IGK3[g]	GCI WTH TAY GGI AAR GGI GGI ATH GGI AA	1.0 µM	19-413[p]	395 bp	95 °C (10 min); 40x 95 °C (45 s), 52 °C (30 s), 72 °C (40 s); 72 °C (10 min)
		DVV[g]	ATI GCR AAI CCI CCR CAI ACI ACR TC	1.0 µM			
nifH	Universal	F2[h]	TGY GAY CCI AAI GCI GA	1.0 µM	115-473[p]	359 bp	95 °C (5 min); 35x 95 °C (45 s), 51 °C (45 s), 72 °C (45 s); 72 °C (7 min)
		R6[h]	GCC ATC ATY TCI CCI GA	1.0 µM			
cbbL	RuBisCO IA & IB	RubIgF[i]	GAY TTC ACC AAR GAY GAY GA	0.4 µM	571-1382[q]	± 800 bp	95°C (3 min); 3x 95°C (1 min), 49°C (2 min 15 s), 72°C (2 min 15 s); 30x 95°C (35 s), 49°C (1 min 15 s), 72°C (1 min 15 s); 72°C (7 min)
		RubIgR[i]	TCR AAC TTG ATY TCY TTC CA	0.4 µM			
cbbL	RuBisCO IA & IC	K2f[j]	ACC AYC AAG CCS AAG CTS GG	0.2 µM	496-990[q]	492-495 bp	95 °C (3 min); 35x 95 °C (1 min), 62 °C (1 min), 72 °C (1 min 30 s); 72 °C (10 min) [235]
		V2r[j]	GCC TTC SAG CTT GCC SAC CRC	0.2 µM			

Gene	Target	Primer	Sequence 5'-3'	Final concentration	Region	Amplicon size	Program
cbbM	RuBisCO II	cbbM343F[c]	GGY AAY AAC CAR GGY ATG GG	0.1 µM	343-1126[c]	700-800 bp	95 °C (3 min); 30x 95 °C (1 min), 50 °C (2 min), 72 °C (3 min); 72 °C (7 min) [59]
		cbbM1126R[c]	CGY ARB GCR TTC ATR CCR CC	0.1 µM			

[a] From [280], [b] From [241], [c] From [242], [d] From [232], [e] From [189], [f] From [201], [g] From [279], [h] From [360], [i] From [239], [j] From [240], [k] Based on the 16S rRNA gene sequence of *Escherichia coli* (A14565), [l] Based on the *pufLM* sequence of *Sphingomonas sanxanigenens* DSM 19645 (CP006644), [m] Based on the *pufM* sequence of *Sphingomonas sanxanigenens* DSM 19645 (CP006644), [n] Based on the *proteorhodopsin* sequence of *Vibrio campbellii* BAA-1116 (FJ985782), [o] Based on the *actinorhodopsin* sequence of *Leifsonia rubra* CMS 76R (ATIA01000023), [p] Based on the *nifH* sequence of *Azotobacter vinelandii* (M20568), [q] Based on the *cbbL* IA sequence of *Bradyrhizobium* sp. ORS278 (CU234118)

Analyses of sequences

Sequences were assembled using the BioNumerics 7.5 software (Applied Maths). To ensure high quality data, every position of the assembly was the consensus of a minimum of two separate sequences. Afterwards, sequences were manually curated to eliminate all low-quality sequences containing indels, stop codons and ambiguous bases. The 16S rRNA genes were initially partially sequenced (V1-V3 region). Partial sequences were grouped at 98 % similarity using CD-HIT [289, 290]. Partial sequences with ≥98 % similarity were considered to represent one phylotype. Afterwards, the full 16S rRNA gene of one representative of each phylotype was sequenced and identified using the EzTaxon database (http://www.ezbiocloud.net/identify) [361].

For analysis of the overlap between the diversity picked up by cultivation and by a culture-independent approach, partial 16S rRNA gene data (V1-V3 region) from the same samples, previously obtained using Illumina MiSeq 2x 300 bp sequencing [Tahon et al. Submitted (Chapter 3)], was compared with 16S rRNA gene data from the isolates. Sequences were compared using CD-HIT. This comparison also allowed verification of taxonomy previously inferred on the Illumina sequence data using the May 2013 GreenGenes training set [Tahon et al. Submitted (Chapter 3)] Additionally, one representative of each phylotype together with one representative of each Illumina OTU (binned at 97 % similarity), was used to construct a phylogenetic tree. Sequences were aligned using Clustal Omega [248, 249] on the Stevin supercomputer at UGent. A total of 20 iterations were performed to optimize the alignment. Afterwards, the alignment was trimmed to the size of the Illumina sequences and visually inspected. The resulting alignment was used to construct a maximum likelihood phylogenetic tree (1000 bootstrap replicates) using the FastTree tree building software [250] with the general time-reversible model and the discrete gamma model with 20 rate categories. The tree was visualized using the iTOL software [251, 252].

Laboratory-based growth experiments to test the effect of light on growth

To investigate the influence of light on growth in the presence of limited amounts of carbon sources, selected strains were grown in liquid PH medium containing different concentrations of the six carbon sources. All strains were grown in PH medium containing a mix of the six carbon sources at 0.25 mM (medium A) and 5 mM (medium E) of each carbon source, respectively. A subset was additionally grown in PH medium containing a mix of 0.5, 1.0 and

2.5 mM of each carbon source (media B, C and D, respectively). All media were enriched with 4 mM ammonium sulfate as a nitrogen source because none of the isolates had yielded a *nifH* amplicon.

Growth under either illumination or darkness was monitored by following the increase of the carbon dioxide concentration in the headspace of liquid cultures as a proxy for growth. In order to track these changes, sterile 120 ml serum vials containing 19.8 ml PH medium were inoculated in six replicates with 200 µl cell suspension of OD 1 (600 nm). For slow growing and low biomass producing isolates OD 0.1 or 0.05 was used. Subsequently, vials were sealed with black butyl stoppers and aluminum crimps. For each medium type, blanks, a positive control (*Loktanella vestfoldensis* LMG 22003) and a negative control (*Cupriavidus metallidurans* LMG 1195) were included. The vials were kept in a temperature controlled room at 15 °C and exposed to either continuous light or darkness, in three replicates each. Dark vials were completely covered in aluminum foil.

Because of the very slow growth, measurements were performed once per week: carbon dioxide concentrations were measured every 7 days using a CompactGC (Global Analyze Solutions) gas chromatograph, for up to 10 weeks. The change in pressure due to carbon dioxide production was measured with an infield 7 pressure meter (UMS). Values obtained by performing gas chromatography were converted to µmol gas L_{Liquid}^{-1} by compensating for change in gas pressure and taking the solubility of carbon dioxide into account.

Statistical analyses

Data was plotted using the Grofit package v1.1.1-1 [362] in R 3.2.3 (https://cran.r-project.org/). Afterwards, a non-parametric fit of the growth data was made for each replicate (1000 bootstraps). Generated maximum slope (*i.e.* growth rate in µM CO_2 produced per hour) and maximal CO_2 concentration values in stationary phase were used for significance testing using a one-way ANOVA.

Accession numbers

The sequences determined in this study have been deposited in the National Center for Biotechnology GenBank database under accession numbers KY386300 - KY386629 (16S rRNA) and KY437105 to KY437155 (*pufLM*).

5.3. Results

Isolation of bacteria

While all media showed a good yield, the use of oligotrophic media and incubation under low temperatures (4 and 15 °C) resulted in growth only after prolonged periods of time. From solid media, purification of distinct colonies was performed from week 10 on, whereas for the liquid setups – originally enriched for 21 weeks – this was after up to seven months. From the various conditions, a total of 1552 colonies were randomly picked up. The number of isolates picked up per sample and per incubation condition is listed in Table S1.

The majority of the colonies were extremely small (<0.5 mm ø) and many showed no visual pigmentation, except upon concentration of biomass. Furthermore, the liquid enrichments resulted in a large number of cyanobacterial liquid cultures. Samples KP2 and KP15 resulted in most cyanobacterial growth and diversity. For samples KP43 and KP53 cyanobacterial growth was absent from most of the setups. These samples, however, displayed growth of green microalgae that were identified as *Stichococcus* sp. by 18S rRNA gene sequencing.

Because our focus was on AAP, the cyanobacterial enrichment cultures which may harbor a broad variety of potentially novel cyanobacterial strains were donated to the Cyanobacteria culture collection at the University of Liège for further isolation and characterization. As determined by microcopy, these contained a large variety of *Nostoc* sp., *Stigonema* sp., as well as cyanobacterial growth that could not be identified (A. Wilmotte, personal communication).

MALDI-TOF MS

For the dereplication, identification and characterization of the isolates, an approach combining MALDI-TOF MS fingerprinting and sequence analysis of 16S rRNA and phototrophy genes was used. First, MALDI-TOF MS was performed to characterize all 1552 isolates. Only spectra of good quality were considered for further analysis: quality mass spectrometry profiles (minimal highest peak intensity of 200 and slope <35 % of the maximal peak intensity) could be generated for 1038 isolates. For the other isolates, even on repetition, no good quality profile could be obtained, mostly due to the very low amount of biomass available. As a result, these isolates were not used in further analyses. Subsequently, cluster analysis (Pearson correlation) of MALDI-TOF MS profiles was performed. Visual inspection

resulted in 141 clusters (data not shown). Approximately 15 % of the isolates did not group in these clusters and were characterized by unique profiles often of a somewhat lower quality. Prior to further analyses, the initial number of isolates was reduced to a subset of isolates representing all MALDI-TOF MS clusters as well as good quality unique profiles (330 isolates total) (Table S2).

Identification based on 16S rRNA gene sequences

After partial 16S rRNA gene sequencing, the 330 representative isolates were binned in 77 phylotypes (98 % DNA similarity). Based on the full 16S rRNA gene sequence, taxon assignment with the EzTaxon cultured database allocated 63 phylotypes (295 isolates) to 29 genera (95 % cutoff [363]) belonging to the Actinobacteria, Proteobacteria, Firmicutes, Bacteroidetes and Deinococcus-Thermus (Table 2). Of the remaining phylotypes, the 16S rRNA gene sequence was less than 95 % similar to that of their closest cultured neighbor. Of these, 12 grouped with the aforementioned phyla and potentially make up new genera or families (90-95 % 16S rRNA similarity to closest cultured neighbor) (Table 2). Interestingly, two isolates displayed less than 80 % similarity to named species and thus may represent the first cultured isolates of novel or uncultured taxa at a less detailed phylogenetic level. The sequence of isolate R-68168 (phylotype 64) was most similar (78.70 %) to that of *Streptacidiphilus rugosus* AM-16 (Actinobacteria), whereas the sequence of R-68213 (phylotype 76) was most similar (79.07 %) to that of *Hippea maritima* DSM 10411 (Deltaproteobacteria). Therefore, these sequences, 89.99 % similar to each other, were additionally compared with the GreenGenes database. Results grouped these isolates with environmental sequences of the candidate phylum FBP (96.26 and 97.83 % highest similarity, respectively) [364]. They are currently being studied in detail and will be reported on separately.

Table 2 Overview of phylotypes recovered from the samples. For each phylotype, the number of representative strains (330 total) enclosed in the group are listed as well as the nearest phylogenetic neighbor and the number of isolates testing positive for different phototrophy genes. Types previously [Tahon et al. Submitted (Chapter 3)] retrieved using Illumina MiSeq sequencing (≥97 % similarity) are indicated with a *

Phylotype	No. of	Identification	Representative	Accession	Nearest phylogenetic neighbor			*puLM*	*bchL*/*bchX*
					Type strain	Accession no.	Sequence similarity (%)		
Proteobacteria									
Alphaproteobacteria									
3	1	*Aureimonas* sp.	R-68373	KY386318	*Aureimonas ferruginae* CC-CFT023	JQ864240	96.64	0	1
4*	10	*Brevundimonas variabilis*	R-68295	KY386469	*Brevundimonas variabilis* ATCC15255	AJ227783	99.79	9	6
1*	10	*Methylobacterium* sp.	R-68173	KY386380	*Methylobacterium iners* 5317S-33	EF174497	98.52	9	2
6	1	*Polymorphobacter* sp.	R-68699	KY386562	*Polymorphobacter multimanifer* 262-7	AB649056	95.99	0	0
2*	1	*Rhodopseudomonas* sp.	R-67878	KY386506	*Rhodopseudomonas pseudopalustris* DSM123	AB498818	98.67	0	0
18*	1	*Roseomonas aquatica*	R-68475	KY386403	*Roseomonas aquatica* TR53	AM231587	99.93	1	0
19*	1	*Roseomonas* sp.	R-68165	KY386462	*Roseomonas tokyonensis* K-20	AB297501	98.60	0	1
5	2	Sphingomonadaceae bacterium	R-67883	KY386496	*Polymorphobacter multimanifer* 262-7	AB649056	94.94	1	1
9*	53	*Sphingomonas* sp.	R-68274	KY386532	*Sphingomonas aerolata* NW12	AJ429240	99.09	3	7
12	12	*Sphingomonas* sp.	R-68304	KY386332	*Sphingomonas cynarae* SPC-1	HQ439186	98.24	6	3
7	18	*Sphingomonas* sp.	R-67984	KY386379	*Sphingomonas faeni* MA-olki	AJ429239	99.62	4	4
16	12	*Sphingomonas* sp.	R-68260	KY386460	*Sphingomonas hunanensis* JSM 083058	FJ527417	98.52	10	4
13	3	*Sphingomonas* sp.	R-68524	KY386357	*Sphingomonas mucosissima* CP173-2	AM229669	98.27	0	0
8	1	*Sphingomonas* sp.	R-68700	KY386453	*Sphingomonas oligophenolica* JCM 12082	AB018439	98.24	0	0
17	1	*Sphingomonas* sp.	R-67954	KY386585	*Sphingomonas pituitosa* EDIV	AJ243751	95.58	0	0
15	7	*Sphingomonas* sp.	R-68270	KY386304	*Sphingomonas yantingensis* 1007	JX566547	96.79	7	5
11*	1	*Sphingomonas* sp.	R-68222	KY386535	*Sphingomonas yunnanensis* YIM 003	AY894691	97.52	0	0
Betaproteobacteria									
20*	2	*Noviherbaspirillum* sp.	R-67997	KY386345	*Noviherbaspirillum psychrotolerans* PB1	JN390675	98.62	0	0
21*	1	*Noviherbaspirillum suwonensis*	R-68579	KY386599	*Noviherbaspirillum suwonensis* 54105-65	JX275858	98.95	0	0
22*	2	*Massilia* sp.	R-67978	KY386448	*Massilia eurypsychrophila* B528-3	KJ361504	97.70	0	0
23	5	*Variovorax* sp.	R-67932	KY386486	*Variovorax boronicumulans* BAM-48	AB300597	97.55	0	0
24	2	*Variovorax* sp.	R-67871	KY386393	*Variovorax ginsengisoli* Gsoil 3165	AB245358	99.26	0	0
Actinobacteria									
43*	2	Actinomycetales bacterium	R-67786	KY386350	*Jatrophihabitans endophyticus* S9650	JQ346802	93.6	0	0
44*	1	Actinomycetales bacterium	R-68701	KY386494	*Jatrophihabitans soli* KIS75-12	KP017569	94.37	0	0
45*	7	Actinomycetales bacterium	R-68183	KY386330	*Frankia alni* ACN14A	CT573213	94.11	0	0
46*	1	Actinomycetales bacterium	R-67810	KY386542	*Cryptosporangium minutisporangium* IFO15962	AB037007	94.41	0	0

Phylotype	No. of	Identification	Representative	Accession	Nearest phylogenetic neighbor			*puLM*	*bchL/bchX*
					Type strain	Accession no.	Sequence similarity (%)		
47*	3	Actinomycetales bacterium	R-67836	KY386609	*Modestobacter versicolor* CP153-2	AJ871304	94.28	0	0
48*	2	Actinomycetales bacterium	R-68223	KY386537	*Sporichthya brevicatena* IFO 16195	AB006164	93.78	0	0
77*	1	*Angustibacter* sp.	R-68259	KY386618	*Angustibacter aerolatus* 7402J-48	JQ639056	96.64	0	0
54*	4	*Aquipuribacter* sp.	R-67807	KY386530	*Aquipuribacter hungaricus* IV-75	FM179321	97.88	0	1
57*	22	*Arthrobacter* sp.	R-67818	KY386623	*Arthrobacter agilis* DSM 20550	X80748	99.46	0	0
58*	2	*Arthrobacter flavus*	R-67793	KY386372	*Arthrobacter flavus* TB 23	ALPM01000	99.25	0	0
75	1	*Arthrobacter* sp.	R-68384	KY386384	*Arthrobacter oxydans* DSM 20119	X83408	99.25	0	0
55*	5	*Arthrobacter pityocampae*	R-68518	KY386301	*Arthrobacter pityocampae* Tp2	EU885749	99.34	0	0
42*	1	*Auraticoccus monumenti*	R-68201	KY386505	*Auraticoccus monumenti* MON 2.2	FN552748	99.58	0	0
59*	3	Dermacoccaceae bacterium	R-68253	KY386608	*Calidifontibacter indicus* PC IW02	EF187228	94.71	0	0
41*	1	*Friedmanniella* sp.	R-68221	KY386527	*Friedmanniella luteola* FA1	AB445453	98.22	0	0
40	5	*Friedmanniella* sp.	R-67749	KY386408	*Friedmanniella sagamiharensis* FB2	AB445456	96.09	0	0
10	1	*Geodermatophilus* sp.	R-68085	KY386308	*Geodermatophilus terrae* PB261	JN033773	99.02	0	0
53*	1	*Knoellia* sp.	R-68061	KY386418	*Knoellia sinensis* DSM 12331	AVPJ010000	98.70	0	0
33*	2	*Marmoricola* sp.	R-67804	KY386524	*Marmoricola aquaticus* CBMAI	JN615437	97.75	0	0
32	2	*Marmoricola* sp.	R-67781	KY386333	*Marmoricola korecus* Sco-A36	FN386723	98.67	0	0
51	7	*Modestobacter* sp.	R-68230	KY386550	*Modestobacter lapidis* MON 3.1	LN810544	97.28	0	0
49*	2	*Nakamurella* sp.	R-68216	KY386516	*Nakamurella lactae* DLS-10	AM778124	98.07	0	0
50*	3	*Nakamurella* sp.	R-68216	KY386516	*Nakamurella lactae* DLS-10	AM778124	98.57	0	0
34	9	*Nocardioides* sp.	R-67800	KY386498	*Nocardioides antarcticus* M-SA3-94	KM347967	97.76	0	0
27*	1	*Nocardioides aquaticus*	R-68162	KY386458	*Nocardioides aquaticus* EL-17K	X94145	99.38	0	0
25	2	*Nocardioides* sp.	R-68562	KY386518	*Nocardioides ginsengagri* BX5-10	GQ339904	97.38	0	0
28	3	*Nocardioides plantarum*	R-68307	KY386526	*Nocardioides plantarum* NCIMB 12834	AF005008	100.00	0	0
26*	1	*Nocardioides* sp.	R-68482	KY386444	*Nocardioides salarius* CL-Z59	DQ401092	97.00	0	0
29*	14	*Nocardioides* sp.	R-68154	KY386423	*Nocardioides terrigena* DS-17	EF363712	97.79	0	1
30*	1	*Nocardioides* sp.	R-67827	KY386592	*Nocardioides terrigena* DS-17	EF363712	96.06	0	0
31*	7	*Nocardioides* sp.	R-68145	KY386386	*Nocardioides terrigena* DS-17	EF363712	97.52	0	1
52*	9	*Phycicoccus* sp.	R-68264	KY386392	*Phycicoccus ochangensis* L1b-b9	GQ344405	98.09	0	0
35	2	*Rhodococcus* sp.	R-68019	KY386523	*Rhodococcus aerolatus* PAMC 27367	KM044053	96.34	0	0
36*	1	*Rhodococcus* sp.	R-68187	KY386338	*Rhodococcus aerolatus* PAMC 27367	KM044053	95.64	0	1
37*	1	*Rhodococcus* sp.	R-68509	KY386574	*Rhodococcus aerolatus* PAMC 27367	KM044053	95.92	0	0
38*	2	*Rhodococcus* sp.	R-67872	KY386395	*Rhodococcus aerolatus* PAMC 27367	KM044053	95.85	0	0
39	7	*Rhodococcus* sp.	R-68273	KY386321	*Rhodococcus fascians* LMG 3623	JMEN010000	98.77	0	2
14*	1	Solirubrobacterales bacterium	R-68159	KY386450	*Conexibacter arvalis* KV-962	AB597950	94.02	0	0

Phylotype	No. of	Identification	Representative	Accession	Nearest phylogenetic neighbor			*puLM*	*bchL/bchX*
					Type strain	Accession no.	Sequence similarity (%)		
Deinococcus-Thermus									
60*	3	*Deinococcus* sp.	R-68561	KY386437	*Deinococcus marmoris* DSM12784	JNIV0100023	100.00	0	0
61*	3	*Deinococcus saxicola*	R-68514	KY386612	*Deinococcus saxicola* AA-1444	AJ585984	99.93	0	0
Firmicutes									
56	1	*Paenibacillus* sp.	R-68670	KY386433	*Paenibacillus frigoriresistens* YIM 016	JQ314346	97.66	0	0
FBP									
64*	1	FBP bacterium	R-68168	KY386300	*Nocardioides echinoideorum* CC-CZW004	KM085325	78.70	0	0
76*	1	FBP bacterium	R-68213	KY386500	*Hippea Maritima* DSM 10411	CP002606	79.07	0	0
Bacteroidetes									
70*	1	*Adhaeribacter* sp.	R-68225	KY386541	*Adhaeribacter aquaticus* DSM16391	AXBK01000	96.37	0	0
67*	1	*Hymenobacter* sp.	R-67758	KY386449	*Hymenobacter arcticus* R2-4	KC213491	98.05	0	0
62	4	*Hymenobacter* sp.	R-68471	KY386441	*Hymenobacter roseosalivarius* AA718	Y18833	98.01	0	0
63	1	*Hymenobacter* sp.	R-68402	KY386432	*Hymenobacter roseosalivarius* AA718	Y18833	97.94	0	0
65	1	*Hymenobacter* sp.	R-68403	KY386435	*Hymenobacter roseosalivarius* AA718	Y18833	98.89	0	0
66*	2	*Hymenobacter* sp.	R-68243	KY386583	*Hymenobacter roseosalivarius* AA718	Y18833	94.75	0	0
68*	8	*Hymenobacter* sp.	R-68178	KY386311	*Hymenobacter terrae* DG7A	KF862488	93.96	1	0
69	1	*Hymenobacter* sp.	R-68030	KY386555	*Hymenobacter terrae* DG7A	KF862488	91.29	0	0
74*	10	*Spirosoma* sp.	R-68079	KY386376	*Spirosoma rigui* WPCB118	EF507900	97.94	0	1
72	1	*Pedobacter* sp.	R-67967	KY386410	*Pedobacter duraquae* WB2.1-25	AM491368	98.36	0	0
71	2	*Pedobacter* sp.	R-68289	KY386572	*Pedobacter panaciterrae* Gsoil 042	AB245368	97.59	0	0
73	2	*Pedobacter* sp.	R-68191	KY386353	*Pedobacter ruber* W1	HQ882803	95.43	0	0

The identifications obtained were compared to the MALDI-TOF MS dendrogram. A total of 126 of the 141 clusters and several unique profiles, accounting for 892 of the 1038 isolates enclosed in the MALDI-TOF MS dendrogram, were well defined and contained isolates belonging to the same genus or species, based on the 16S rDNA identification of the closest EzTaxon hit. Other clusters were taxonomically heterogenous and contained isolates belonging to different genera. This may be explained by profiles of a somewhat lower quality caused by the low biomass obtained for these strains. The comparison also revealed that several phylotypes were represented by multiple MALDI-TOF MS clusters and unique spectra, indicating that several distinct strains were isolated within these groups. Most diversity was retrieved from the solid media and less from the liquid enrichments. The distribution of the recovered taxa across the different terrestrial samples and cultivation setups is shown in Table 3. Several genera, including *Sphingomonas*, *Nocardioides*, *Rhodococcus* and *Hymenobacter*, were retrieved from all samples and nearly all setups, whereas others (*e.g. Rhodopseudomonas*, *Polymorphobacter*, *Knoellia*), were retrieved only from one sample and cultivation setup (Table 3). The most abundantly represented genera over all isolation conditions were *Sphingomonas* (397 isolates), *Nocardioides* (83 isolates) and *Arthrobacter* (83 isolates) (Table 3).

In a previous study [Tahon et al. Submitted (Chapter 3)] the bacterial communities present in the samples had been investigated by sequencing partial 16S rRNA genes (V1-V3 region) using Illumina MiSeq 2x 300 bp sequencing. Grouping at 97 % similarity resulted in a total of 703 OTUs. Comparison of these sequences with those of the isolates allowed more insight in the overlap in diversity retrieved between cultivation and the culture-independent approach. Many of the isolates' 16S rRNA genes grouped together at high similarities (≥97 %) with the environmental sequences (Table 2, Figures S1-S9). However, for 30 of the 77 cultured phylotypes no sequence sharing more than 97 % similarity was present in the culture-independent dataset.

Although deep sequencing revolutionized our knowledge of the bacterial world, these techniques also have weaknesses. Currently, for the widely used Illumina MiSeq platform, the maximum amplicon length after merging reads is restricted to ~550 bp. For 16S rRNA gene genes, these partial sequences encompass only one third of the complete gene and as a result their identification at more detailed taxonomic levels (*i.e.* genus and species) may prove difficult. The identity of the isolates based on full 16S rRNA gene sequences, therefore allowed tentative confirmation or improvement of the previous identification of the Illumina sequences obtained from the same samples. Maximum likelihood analysis clearly showed that

the neighboring Illumina sequences of the 327 representative isolates (74 phylotypes) grouping with Actinobacteria, Proteobacteria, Bacteroidetes and Deinococcus-Thermus had previously all been assigned a correct taxonomy at phylum, class and order level (Figures S1 and S3-S9). For 63 phylotypes the neighboring sequences had also been assigned a correct family. For the others, related Illumina sequences had been unclassified at family level. For phylotypes 29, 49 and 77, the identity of the isolates made it possible to assign a tentative genus and species identity to the highly similar Illumina sequences that were previously identified only to family level. In a few other cases, there were discrepancies in the identification of isolates and the OTUs they grouped with. For example, the V1-V3 region of isolates grouping in phylotypes 4 and 20 (identified as *Brevundimonas variabilis* and *Noviherbaspirillum* sp.) was (nearly) identical to sequences recovered using Illumina (Table 2, Figures S3 and S7) that had been identified as *Mycoplana* and *Collimonas* using the May 2013 GreenGenes training set [285-287]. Repeating the identification of these sequences, now with the current EZTaxon and GreenGenes databases led to the same identification, as *Brevundimonas variabilis* and *Noviherbaspirillum* sp. Thus, although differences sometimes may be due to using a short sequence for identification, they might also be explained by differences in sequence databases used, with the ongoing addition of new sequence data and novel taxa to the databases, improving identifications.

Table 3 Recovery of isolates from the different isolation conditions. Identifications based on the MALDI-TOF MS dendrogram combined with 16S rRNA gene sequencing. PH: medium with carbon sources (glucose, sucrose, sodium succinate, sodium pyruvate, sodium acetate and malate, 0.5 mM each). PA: medium without carbon sources. A: agar, S: shaken liquid setup, NS: non-shaken liquid setup

Identification	**Together**												
	4 °C						**15 °C**						
	PH			**PA**			**PH**			**PA**			
	A	**S**	**NS**	**A**	**S**	**NS**	**A**	**S**	**NS**	**A**	**S**	**NS**	**Total no. of isolates**
Proteobacteria													
Alphaproteobacteria													
Aureimonas sp.	0	2	2	0	0	1	0	0	0	0	0	0	5
Brevundimonas variabilis.	0	1	3	0	3	5	0	0	14	1	0	0	27
Methylobacterium sp.	0	0	5	0	14	6	1	0	0	0	0	0	26
Polymorphobacter sp.	0	0	0	0	0	0	0	0	0	0	0	1	1
Rhodopseudomonas sp.	0	0	0	0	0	0	0	0	0	0	0	1	1
Roseomonas sp.	0	0	0	0	0	1	1	0	0	0	0	0	2
Sphingomonadaceae bacterium	0	0	0	0	0	0	0	0	0	0	0	4	4
Sphingomonas sp.	35	62	51	25	34	30	26	40	73	5	7	9	397

Identification	Together												
	4 °C						15 °C						
	PH			PA			PH			PA			
	A	S	NS	A	S	NS	A	S	NS	A	S	NS	Total no. of isolates
Betaproteobacteria													
Noviherbaspirillum sp.	0	0	0	1	2	1	0	0	0	0	0	0	4
Massilia sp.	0	0	0	3	0	0	0	0	0	0	0	0	3
Variovorax sp.	0	2	0	1	1	2	0	0	0	0	4	3	13
Actinobacteria													
Actinomycetales bacterium	0	0	0	0	0	2	13	0	0	6	0	3	24
Angustibacter sp.	0	0	0	0	0	0	1	0	0	0	0	0	1
Aquipuribacter sp.	0	0	0	0	0	0	1	0	0	6	0	0	7
Arthrobacter sp.	12	1	4	19	5	6	17	0	0	19	0	0	83
Auraticoccus monumenti	0	0	0	0	0	0	2	0	0	0	0	0	2
Dermacoccaceae bacterium	0	0	0	0	0	0	4	0	0	0	0	0	4
Friedmanniella sp.	0	0	2	0	1	0	2	0	0	5	0	0	10
Geodermatophilus sp.	1	0	0	0	0	0	0	0	0	0	0	0	1
Knoellia sp.	3	0	0	0	0	0	0	0	0	0	0	0	3
Marmoricola sp.	0	0	0	0	0	1	2	0	0	8	0	0	11
Modestobacter sp.	1	3	0	0	3	0	3	0	0	6	0	0	16
Nakamurella sp.	0	0	0	0	0	0	1	0	0	8	0	0	9
Nocardioides sp.	11	5	1	5	9	11	11	3	5	20	0	2	83
Phycicoccus sp.	3	3	0	7	0	0	1	4	0	0	5	0	23
Rhodococcus sp.	0	0	0	9	1	5	3	5	0	1	0	2	26
Solirubrobacterales bacterium	0	0	0	0	0	0	1	0	0	0	0	0	1
Deinococcus-Thermus													
Deinococcus sp.	0	0	0	0	9	4	2	0	0	0	0	0	15
Firmicutes													
Paenibacillus sp.	0	0	0	0	0	0	1	0	0	0	0	0	1
FBP													
FBP bacterium	0	0	0	0	0	0	2	0	0	0	0	0	2
Bacteroidetes													
Adhaeribacter sp.	0	0	0	0	0	0	1	0	0	0	0	0	1
Hymenobacter sp.	4	5	3	3	9	11	3	0	0	1	1	0	40
Pedobacter sp.	1	0	0	3	0	0	4	1	0	0	1	0	10
Spirosoma sp.	2	2	4	5	3	3	0	0	0	1	10	6	36

Protein coding gene analyses

For all 330 representative isolates phototrophy potential was tested by amplification of key genes involved in rhodopsin- and (bacterio)chlorophyll-dependent light-harvesting. Proteorhodopsin and actinorhodopsin genes could not be amplified from any of the isolates. Amplification of *pufLM* resulted in 51 positive reactions. The more universal Yutin et al. *pufM* primers [232] did not provide additional positive isolates. Positive isolates belonged to *Sphingomonas* (30 isolates), *Methylobacterium* (9 isolates), *Brevundimonas* (9 isolates), Sphingomonadaceae bacterium (1 isolate), *Hymenobacter* (1 isolate) and *Roseomonas* (1 isolate) (Tables 2 and S2). Comparison of these *pufLM* sequences with *pufLM* cloned sequences from our previous work [277] revealed that only one sequence was recovered using both approaches. The sequence of isolate R-68361 (*Hymenobacter* sp.) was identical to two cloned sequences (accession no. KT154478) obtained from the same terrestrial sample (*i.e.* KP15). All other isolate *pufLM* sequences were less than 86 % similar to cloned sequences.

BchL/bchX could be amplified from 41 of the 330 representative isolates, 17 of which also tested positive for *pufLM* (Table S2). Known anoxygenic phototrophs contain both *bchL* and *bchX* [162], which co-amplify with the primers used here [279]. Therefore, these PCR products could not be sequenced directly. Since the IGK3 and DVV primers used for amplification of *bchL* and *bchX* also amplify *nifH* [279], an additional PCR was carried out with broad range *nifH* primers (F2 and R6) to verify no *nifH* was amplified (Table 2). This primer set, designed by Marusina and colleagues [360] was reported to strictly amplify *nifH* and resulted in no positive amplifications [111].

In previous research, the diversity of bacterial key protein encoding genes in the Calvin-Benson-Bassham cycle (RuBisCO) was investigated using clone libraries and Illumina MiSeq sequencing [277] [Tahon et al. Submitted (Chapter 3)]. Results revealed a large diversity of RuBisCO types IA, IB and IC (*cbbL* gene) grouping with Proteobacteria and Actinobacteria. Type II RuBisCO (*cbbM* gene) could not be amplified from the samples. Therefore, the 330 representative isolates were additionally checked for the presence of *cbbL* and *cbbM* genes (Table 1). This resulted in no positive amplifications.

Laboratory-based growth experiments to test the effect of light

To evaluate the effect of light on growth under limited availability of carbon sources, a selection of strains was grown in light and dark conditions and growth was compared. Between one and three strains were chosen from each phylotype. To take primer mismatch

into account, selected strains included both strains testing negative and positive for the different phototrophy genes. Due to the slow growth and low biomass production of the strains, only 26 strains representing 24 phylotypes could be tested (Table 4).

Eight strains grew better in light, in at least one of the media tested: R-68270 (*Spingomonas* sp.), R-68435 (*Spingomonas* sp.), R-68495 (*Sphingomonas* sp.) and R-68562 (*Nocardioides* sp.) reached a significantly higher maximum CO_2 production and R-68253 (Dermacocaccaceae bacterium), R-68375 (*Sphingomonas* sp.), R-68523 (*Spirosoma* sp.), R-68535 (*Modestobacter* sp.) as well as R-68435 and R-68562 displayed significantly faster CO_2 production in light in at least one of the media tested (Table 4). Only four of these strains had tested positive for *pufLM* or/and *bchL/bchX* (Table 4). While none of the strains grew significantly faster in the dark, four strains did reach a higher maximum CO_2 production under darkness in at least one of the growth media tested: R-68473 (*Brevundimonas variabilis*), R-68511 (Actinomycetales bacterium, distantly related to *Jatrophihabitans endophyticus* at a 16S rRNA gene sequence similarity of 93.60 %), R-68561 (*Deinococcus* sp.) and R-68669 (*Methylobacterium* sp.) (Table 4). Interestingly, of these strains, R-68669 and R-68473 had tested positive for the phototrophy genes and are highly related to known AAP genera (*Methylobacterium*) and species (*Brevundimonas subvibrioides*). For the remaining 12 strains, no significant differences in growth parameters were observed between light and dark conditions (Table 4). However, it should be noted that for several of the setups stationary phase was not reached after 10 weeks and it cannot be excluded that upon longer incubation some of these strains might show significant differences between maximum CO_2 production in light or dark.

With regard to the different concentrations of carbon sources tested, the fact that several of the setups did not reach stationary phase by the end of the 10 week experiment reduced data available for interpretation. Because lag phases differed between strains and setups (data not shown), CO_2 production after 10 weeks could not be compared. Nevertheless, for several of the strains (R-67883, R-68375, R-68552, R-68473, R-68669) there is a tentative trend of increasing maximum CO_2 production with increasing concentration of carbon source (Table 4).

The negative control LMG 1195 (*Cupriavidus metallidurans*) indeed showed no significant differences between growth in light or dark. The positive control LMG 22003 (*Loktanella vestfoldensis*) demonstrated a significantly higher CO_2 production under constant illumination in several growth media. The difference also became more significant with increasing amounts of carbon source in the growth medium (Table 4).

Table 4 Kinetic parameters for growth of different Antarctic isolates, positive control LMG 22003 and negative control LMG 1195. Significance of differences for growth between light and dark were determined via one-way ANOVA. In case of significant differences, the highest value is shown in bold. Absence of a standard deviation indicates that the results of only one replicate could be used. NG: No Growth. -: no stationary phase was reached

Strain	ID	*pufLM*	*bchL/ bchX*	Medium[a]	Maximum CO_2 production (µM)		*P*	Growth rate (µM CO_2 produced per hour)		*P*
					Light	Dark		Light	Dark	
R-67786	Actinomycetales bacterium	-	-	A	-	NG		0.033	NG	
				E	-	-		0.26	0.13 ±0.11	
R-67883	Sphingomonadaceae bacterium	-	+	A	60.25 ±6.38	47.33 ±9.12		0.077 ±0.0016	0.069 ±0.033	
				B	49.26 ±5.68	-		0.051 ±0.017	1.63 ±1.93	
				C	104.83 ±20.29	96.59 ±25.31		0.11 ±0.020	0.10 ±0.045	
				D	-	-		2.45 ±3.94	13.40 ±11.48	
				E	-	683.76 ±53.59		0.50 ±0.10	0.64 ±0.067	
R-68019	*Rhodococcus* sp.	-	-	A	NG	NG		NG	NG	
				E	NG	NG		NG	NG	
R-68061	*Knoellia* sp.	-	-	A	-	-		0.15 ±0.15	0.57	
				E	69.77 ±8.28	-		0.059 ±0.012	0.061	
R-68159	Solirubrobacterales bacterium	-	-	A	889.01 ±33.52	861.78 ±49.21		4.53 ±1.65	4.24 ±2.59	
				E	-	-		16.62 ±7.17	18.72 ±13.41	
R-68183	Actinomycetales bacterium	-	-	A	643.33 ±179.90	655.06 ±121.05		3.06 ±0.97	3.11 ±0.48	
				E	5638.50 ±346.38	6077.08 ±798.03		20.14 ±3.47	24.93 ±21.34	
R-68223	Actinomycetales bacterium	-	-	A	524.34 ±1.95	521.96 ±30.07		2.74 ±0.23	2.69 ±0.79	
				E	4022.80 ±185.22	3688.34 ±303.64		8.56 ±2.04	9.74 ±0.37	
R-68253	Dermacoccaceae bacterium	-	-	A	NG	NG		NG	NG	
				B	-	-		0.063	0.045	
				C	-	-		**3.32 ±0.21**	2.66 ±0.24	0.0212
				D	-	-		6.87 ±2.82	4.80 ±0.47	
				E	-	81.89 ±7.16		1.23 ±1.97	0.05 ±0.018	
R-68259	*Angustibacter* sp.	-	-	A	NG	NG		NG	NG	
				E	NG	NG		NG	NG	
R-68270	*Sphingomonas* sp.	+	+	A	-	-		0.17 ±0.047	0.24 ±0.25	
				E	**221.90 ±60.35**	88.21 ±35.19	<0.0010	0.49 ±0.22	0.37 ±0.30	
R-68327	*Phycicoccus* sp.	-	-	A	854.56 ±13.44	836.19 ±23.61		3.51 ±2.00	5.69 ±0.37	
				E	7312.10 ±249.73	-		16.35 ±13.71	9.06 ±9.08	
R-68375	*Sphingomonas* sp.	-	+	A	836.85 ±22.81	816.24 ±40.57		5.14 ±0.18	4.77 ±0.15	
				B	1489.97 ±40.71	1492.48 ±41.86		**4.62 ±0.099**	3.64 ±0.39	0.0129
				C	2819.94 ±59.02	2932.27 ±8.07		**7.48 ±0.15**	6.14 ±0.44	0.0075
				D	5901.99 ±265.45	5920.35 ±167.33		10.54 ±2.91	7.34 ±1.68	
				E	6859.19 ±266.01	7074.65 ±22.24		11.42 ±5.11	16.24 ±1.69	
R-68435	*Sphingomonas* sp.	+	-	A	525.63 ±34.94	535.85 ±55.37		2.90 ±0.60	2.15 ±1.75	
				E	**5822.30 ±661.88**	4950.06 ±240.32	0.0230	**18.02 ±2.21**	15.51 ±1.84	0.0171
R-68460	*Rhodococcus* sp.	-	+	A	825.83 ±24.65	879.51 ±40.04		5.03 ±0.28	5.14 ±0.22	
				E	-	7129.56 ±388.09		6.40 ±5.74	15.03 ±4.32	
R-68471	*Hymenobacter* sp.	-	-	A	691.45 ±24.57	714.32 ±28.61		2.22 ±1.30	2.27 ±1.16	
				E	6974.89 ±48.21	7275.63 ±135.29		14.39 ±4.34	17.17 ±7.05	
R-68473	*Brevundimonas variabilis*	+	+	A	NG	NG		NG	NG	

Strain	ID	*pufLM*	*bchL*/ *bchX*	Medium[a]	Maximum CO_2 production (µM)		*P*	Growth rate (µM CO_2 produced per hour)		*P*
					Light	Dark		Light	Dark	
				B	33.76 ±4.19	-		0.035 ±0.0027	0.046	
				C	-	-		0.077 ±0.063	0.073 ±0.044	
				D	-	-		0.056 ±0.053	0.099 ±0.044	
				E	272.30 ±60.35	**5995.98 ±230.32**	<0.0010	0.22 ±0.12	14.97	
R-68477	*Arthrobacter agilis*	-	-	A	NG	NG		NG	NG	
				E	-	-		0.26 ±0.35	0.77	
R-68486	*Nocardioides* sp.	-	-	A	-	891.35 ±43.68		0.85 ±0.94	1.48 ±0.0078	
				E	-	2622.53 ±3686.46		0.37 ±0.41	6.12 ±9.07	
R-68495	*Sphingomonas* sp.	-	+	A	652.56 ±211.41	804.88 ±7.46		2.44 ±0.47	3.36 ±1.54	
				E	**4181.68 ±198.86**	3701.42 ±122.40	0.0378	11.79 ±0.41	13.65 ±3.22	
R-68511	Actinomycetales bacterium	-	-	A	824.32 ±31.46	**920.88 ±37.03**	<0.0010	1.52 ±0.11	1.68 ±0.22	
				E	4010.85 ±111.51	**6523.94 ±307.18**	<0.0010	5.46 ±4.71	11.65 ±4.90	
R-68523	*Spirosoma* sp.	-	-	A	52.54 ±15.69	-		0.14 ±0.11	0.047 ±0.010	
				E	-	81.51 ±16.15		**0.25 ±0.022**	0.10 ±0.038	0.0038
R-68535	*Modestobacter* sp.	-	-	A	-	653.51 ±6.15		0.71	0.77 ±0.64	
				E	50.66 ±8.95	41.25 ±1.04		**0.12 ±0.013**	0.035 ±0.0012	<0.0010
R-68552	*Sphingomonas* sp.	+	-	A	644.49 ±259.08	425.32 ±375.43		1.26 ±0.66	0.89 ±0.62	
				B	1396.83 ±46.22	1403.64 ±55.64		5.01 ±0.17	4.15 ±1.13	
				C	2856.27 ±71.58	2889.29 ±110.72		6.21 ±0.62	5.92 ±1.17	
				D	4793.38 ±166.73	5351.29 ±441.72		5.71 ±0.57	9.02 ±3.04	
				E	-	4302.90 ±3714.21		6.32 ±4.89	6.71 ±7.95	
R-68561	*Deinococcus* sp.	-	-	A	619.09 ±3.16	**688.40 ±25.74**	0.0098	1.56 ±0.091	1.65 ±0.56	
				E	-	-		0.74	15.31 ±4.45	
R-68562	*Nocardioides* sp.	-	-	A	**747.56 ±42.60**	39.81 ±6.60	<0.0010	**1.48 ±0.23**	0.072 ±0.010	0.0040
				E	-	120.23 ±37.70		0.42 ±0.35	0.13 ±0.046	
R-68669	*Methylobacterium* sp.	+	+	A	-	-		0.12 ±0.064	0.22 ±0.16	
				B	-	352.31 ±13.20		1.19	0.61 ±0.50	
				C	-	-		1.55	1.75	
				D	-	-		0.89 ±1.45	2.28	
				E	2557.87 ±144.27	**7621.57 ±19.94**	<0.0010	1.82 ±1.69	4.99 ±6.53	
LMG 22003	*Loktanella vestfoldensis*	+	-	A	**191.52 ±50.08**	160.79 ±18.01	0.029	0.91 ±0.28	0.91 ±0.044	
				B	178.86 ±14.36	166.12 ±18.80		**0.76 ±0.0065**	0.65 ±0.028	0.0024
				C	2808.50 ±129.79	2795.66 ±70.33		4.78 ±0.99	6.22 ±0.33	
				D	**5141.38 ±341.39**	3917.59 ±30.96	0.0035	6.91 ±0.90	7.25 ±0.14	
				E	**6949.70 ±32.90**	5052.70 ±194.43	<0.0010	16.83 ±3.30	13.45 ±0.89	
LMG 1195	*Cupriavidus metallidurans*	-	-	A	318.09 ±0.85	308.88 ±3.77		2.23 ±0.0067	2.13 ±0.046	
				E	3811.06 ±171.02	3632.01 ±100.60		18.47 ±1.39	18.99 ±0.55	

[a] A, B, C, D or E indicate the PH growth medium contained 0.25, 0.5, 1.0, 2.5 or 5 mM of each carbon source (*i.e.* glucose, sucrose, sodium succinate, sodium pyruvate, sodium acetate and malate)

5.4. Discussion

AAP relying on a type 2 photochemical reaction center have predominantly been found in the Alpha- and Gammaproteobacteria, and to a lesser extent in the Betaproteobacteria. A single representative is known in the Gemmatimonadetes and in the Firmicutes [146, 147, 177]. Members of these taxa are common inhabitants of Antarctic ice-free areas [28, 29]. Given the range of extreme environmental conditions bacteria in Antarctic soils are subjected to (*e.g.* sub-zero temperatures with repeated cycles of freezing and thawing, low transient precipitation [16], very low availability of organic matter [275] and strong solar radiation at exposed sites), some bacteria may have adopted a phototrophic lifestyle, converting sunlight into chemical energy. Indeed, previous studies of environmental DNA revealed that a broad diversity of (aerobic) anoxygenic phototrophs is present in the exposed soils in the proximity of the Princess Elisabeth Station, whereas the relative abundance of oxygenic photosynthetic microorganisms was found to be low in many samples [55, 275, 338]. Therefore, our main objective was to isolate and characterize AAP, a group that has, to date, nearly exclusively been studied in aquatic environments [146].

Our isolation strategy using oligotrophic media and a light regime simulating the increasing day length over transition from Antarctic winter to summer is the first to report on the diversity of culturable aerobic anoxygenic phototrophic bacteria from soils from the SRM (East Antarctica) and gave access to a range of bacteria known to be common inhabitants of soils, including those of Antarctica [23, 46, 216, 302, 365]. Following MALDI-TOF MS and 16S rRNA gene analysis, ~52 % of the 892 identified isolates grouped among known alphaproteobacterial AAP taxa, particularly with *Sphingomonas* (~45 %). The majority of these alphaproteobacterial AAP was retrieved from the liquid enrichments incubated at 4 °C (Table 3). A total of 20 isolates, many obtained from PA medium incubated at 4 °C, grouped with Betaproteobacteria, whereas only a single Firmicutes isolate was picked up. However, none of these grouped among known AAP taxa. No Gammaproteobacteria and Gemmatimonadetes were isolated. This observation is in accordance with previous observations made in these samples with culture-independent approaches. Clone libraries and Illumina MiSeq sequencing of *puf(L)M* and *bchL*/*bchX* genes revealed a dominance of alphaproteobacterial AAP, including many of the groups isolated here [277, 338]. Only 0.65 % of the ~680000 *pufM* Illumina reads grouped with beta- and gammaproteobacterial *pufM*, while 16S rRNA gene sequencing also revealed a high relative abundance of Alphaproteobacteria and in particular AAP taxa. On the other hand, Beta- and

Gammaproteobacteria, Gemmatimonadetes and Firmicutes were (nearly) absent in these samples and the SRM in general [275] [Tahon et al. Submitted (Chapter 3)]. Although similar patterns could be observed in both approaches (*i.e.* dominance of alphaproteobacterial AAP taxa), the comparison of culture-dependent and culture-independent 16S rRNA and *puf(L)M* gene datasets only revealed a limited overlap, indicating that both approaches provide complementary information about a community's diversity which might be missed when using a single approach. The low overlap may be explained by limitations of these methods. Amplicon sequencing is dependent on DNA extraction and the use of primers. These steps do not retrieve all diversity. Indeed, recent metagenome analysis led to the discovery of a novel bacterial phylum that had always remained hidden because of mismatches in the most commonly used primers for 16S rRNA gene sequencing [366]. While cultivation might overcome this problem, many isolates are still resistant to the commonly used cultivation techniques. The low number of types shared will also have been biased by the cultivation setup. Firstly, although specific cultivation conditions for the targeted growth of aerobic phototrophic bacteria were used, in addition to oxygenic and anoxygenic phototrophic bacteria, we also isolated a high number of non-phototrophs lacking *pufLM* and *bchL*/*bchX*. A portion of these isolates may have phototrophic capacities dependent on rhodopsins or BchL, however, these features may have been missed as a result of primer mismatch, or the presence of a different rhodopsin type or a type 1 photochemical reaction center. For example, very little rhodopsin data has been retrieved from currently available terrestrial Antarctic metagenomes on MG-RAST [207] and IMG [208]. Most rhodopsin data, however, originates from aquatic environments [190]. As it is well known that this protein family holds an enormous diversity [179, 183], currently available primers may be unsuitable to capture terrestrial Antarctic rhodopsin variants, whereas annotation pipelines may be unable to detect these variants in terrestrial metagenomes. Future advances in physiological characterization and genome analyses may resolve this question. Secondly, the unexpected and abundant growth of Cyanobacteria in our liquid enrichments may have restricted the growth of other phototrophic bacteria in these setups. Finally, the high number of colonies, their very slow growth and miniscule colony size, in combination with manual picking introduced an additional bias. It was not possible to isolate every single colony. Additionally, the limited biomass production impeded many of the analyses and called for a modification of standard operating procedures as used for high biomass producing fast growing organisms. Development of new innovative and high throughput strategies will be necessary to cultivate and characterize a larger proportion of the Antarctic biodiversity.

Of the 75 isolates showing phototrophy potential (*i.e.* positive PCR for a phototrophy gene), 67 grouped with alphaproteobacterial AAP taxa (Table S2). Although the abundance of alphaproteobacterial AAP is well recognized in various aquatic environments all over our planet [146, 174, 367, 368], to date little is known about their diversity, distribution and role in terrestrial ecosystems. The majority of potential phototrophs, and ~45 % our isolates in general, grouped with *Sphingomonas* (Table 2, Figure S5). To assess whether light can enhance growth, five of these *Sphingomonas* strains (with PCR-detectable *pufLM* or/and *bchL*/*bchX* phototrophy genes) were grown in media containing various amounts of carbon sources, in light and in dark conditions (Table 4). In at least one of the media tested, four strains showed a significantly higher CO_2 production or/and higher growth rate when grown in light, confirming the beneficial effect of light on growth of AAP. The presence of *Sphingomonas* in Antarctica is not unusual, as several members of this group have previously been isolated from a range of cold ecosystems, including Antarctica [46, 369, 370] and are well known for their phototrophic capacities [371, 372]. In recent years, analyses have also revealed the metabolic diversity of several *Sphingomonas* strains as well as their capacity to adapt to extreme cold, high UV-B radiation and arid conditions, indicating they are ideal candidates for survival in extreme oligotrophic Antarctic systems [369, 373-375]. The second most recovered group of alphaproteobacterial phototrophs was highly related to *Methylobacterium*. Some *Methylobacterium* strains are well known for their tolerance to high UV radiation and dehydration, which are common environmental conditions in Antarctic soils [376]. Indeed, Romanovskaya et al. (2009) reported representatives of this genus, and in particular *Methylobacterium extorquens*, a bacterium known to possess a type 2 photochemical reaction center, in terrestrial biotopes on several Antarctic islands [377, 378], whereas PufM, BchL and BchX sequences from members of this taxon have previously been reported from terrestrial locations in the Arctic [205] and Antarctic [338].

Based on 16S rRNA gene sequence data combined with the MALDI-TOF MS dendrogram, 87 of our isolates grouped with the Bacteroidetes, and especially with *Hymenobacter* (Table 3, Figure S4). Members of this genus, of which the type species *Hymenobacter roseosalivarius* was originally isolated from exposed areas from the McMurdo Dry Valleys [379], have been commonly reported from several terrestrial and aquatic Antarctic locations [46, 365, 380]. To our knowledge, no anoxygenic phototrophic members have ever been reported in Bacteroidetes phylum [381]. Remarkably, the *pufLM* genes encoding the conserved proteins

of the type 2 phototrophic reaction center could be amplified from *Hymenobacter* R-68361. This *pufLM* sequence was identical to cloned sequences previously obtained from the same sample [277]. The 16S rRNA gene sequence of strain R-68361 was found to be 93.96 % identical to that of *Hymenobacter terrae* DG7A, a recently described strain isolated from soil samples in Seoul (South Korea) [382]. In addition, a positive result for *bchL*/*bchX* was obtained from *Spirosoma* R-67957 (Table S2). Strains R-68361 and R-67957 were selected for the growth experiments performed here, but due to the very slow growth and low biomass production, insufficient inoculum could be produced. However, during the growth experiments, a second *Spirosoma* isolate (R-68523) showed a significantly higher growth rate in light conditions (Table 4). Strain R-68523 tested negative for phototrophy genes, but was closely related to *Spirosoma* strain R-67957 (Table S2). Thus, to assess whether these isolates are the first phototrophic representatives of the Bacteroidetes, genome and transcriptome analysis will be needed to verify the presence and expression of phototrophic genes.

In addition to taxa known to contain AAP, ~31 % of the 892 isolates were identified as Actinobacteria. These were predominantly isolated from solid media (Table 3). Six of these isolates tested positive in the PCR for *bchL*/*bchX*, although these products co-amplified and therefore not sequenced. None of the Actinobacteria tested positive for rhodopsins or *pufLM* (Table 2). However, three actinobacterial strains (R-68562: *Nocardioides* sp., R-68253: Dermacoccaceae bacterium, R-68535: *Modestobacter* sp.) had a significantly higher growth rate or/and CO_2 production when grown under illumination (Table 4) although we could not detect phototrophy genes in the latter three strains. This might be explained by primer mismatch, as currently available primers may be unable to target all the rhodopsin and photoreaction center diversity. Future physiological characterization and genome analyses may resolve this question. Until then, the phototrophic potential of these Actinobacteria remains unknown. All actinobacterial isolates grouped with the Actinobacteridae, which is similar to observations made in other Antarctic soils [29, 46, 365]. Most of the Actinobaceridae isolates were identified as *Arthrobacter*, *Nocardioides*, *Modestobacter* and *Rhodococcus*. Members of these genera have previously been isolated from several cold regions like Greenland, the Arctic and Antarctica [383-387]. The presence of *Modestobacter*, some of which are known to be cold tolerant and radiation resistant, may be linked to its possible involvement in the weathering of rocks [388, 389]. *Arthrobacter*, on the other hand, is a genus generally associated with the soil compartment and is recognized for its

physiological versatility (*e.g.* altering of its cell wall's fatty acid composition in response to lowered growth temperatures) and its ability to use a wide range of substrates [365, 390].

In addition to preferential growth under continuous light, the growth experiments also revealed that four strains produced significantly more CO_2 in the dark. Remarkably, two of these (R-68473, *Brevundimonas variabilis* and R-68669, *Methylobacterium* sp.) tested positive for phototrophy genes and group with known AAP genera (Table 4). A possible explanation for the absence of light-enhanced growth in these strains in our experimental conditions may be the effect of oxygen. Rathgeber et al. (2012) showed that O_2 levels may greatly influence the efficiency of the phototrophic apparatus of AAP and that, in some cases, this apparatus is not fully functional under normal atmospheric oxygen concentrations, but only under semi-aerobic conditions [391]. Furthermore, growth under continuous light may have resulted in photodynamic damage or a photo-induced growth-inhibitory effect. Indeed, previous research has shown that the puf operon is sensitive to light in some AAP [392], whereas Lami et al. (2009) found that the highest net growth of certain phototrophs appears under dark-light cycles [393]. It should also be noted that both rhodopsin-dependent and Bchl-dependent phototrophy require the transcription of multiple unique genes [137]. Thus, although our approach detected key genes from these pathways, the possible absence of a complete gene cluster may explain these strains' inability to show light-enhanced growth. Furthermore, our observations are based on laboratory testing. It is possible that our bacteria produce BchL or rhodopsins, but only under different conditions.

5.5. Conclusion

This study provides the first data on the culturable aerobic anoxygenic phototrophic bacterial diversity in the Sør Rondane Mountains, East Antarctica. Our isolation strategy resulted in about 52% of isolates belonging to known alphaproteobacterial AAP taxa, while other isolates were distributed over six phyla, including a candidate phylum. In addition, enrichment cultures revealed the presence of Cyanobacteria and even green algae. Growth experiments indicated that while light benefits the growth of several AAP, other AAP isolates grew better in the dark, at least under laboratory conditions. We also demonstrated for the first time that some Bacteroidetes and Actinobacteria isolates may have phototrophic potential. Overall, our results suggest that the ability to adopt a photoheterotrophic lifestyle may provide an advantage in the oligotrophic Antarctic soils surrounding the Princess Elisabeth Station.

5.6. Acknowledgements

This work was supported by the Fund for Scientific Research – Flanders (project G.0146.12). Additional support was obtained from the Belgian Science Policy Office (project CCAMBIO). The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the Flemish Supercomputer Center (VSC) funded by Ghent University, the Hercules Foundation and the Flemish Government – department EWI. We thank Anouk Peeters and Liesbeth Lebbe for excellent technical assistance with growth experiments. This study is a contribution to the State of the Antarctic Ecosystem (AntEco) research program of the Scientific Committee on Antarctic Research (SCAR).

5.7. Supplementary information

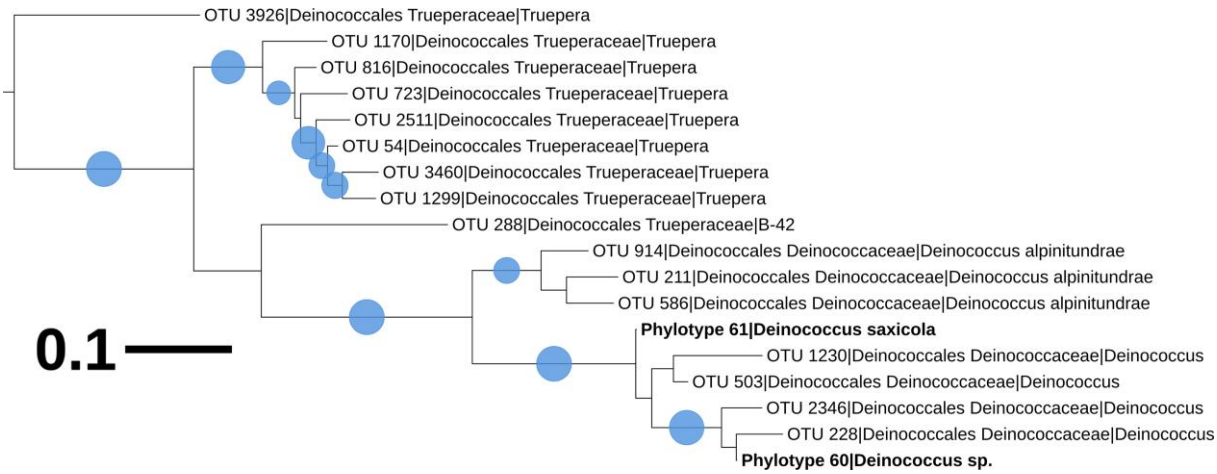

Figure S1 Detailed view of the Deinococcus-Thermus cluster of the maximum likelihood phylogenetic tree (1000 bootstraps) combining partial 16S rRNA gene sequences obtained using a culture-independent and –dependent approach. Representatives of the isolate phylotypes are labelled in bold. For Illumina OTUs [Tahon et al. Submitted (Chapter 3)], identifications at the most detailed taxonomic level are given. Scale bar indicates 0.1 substitutions per position. Bootstrap values are displayed as circles with a diameter reflecting the height of the bootstrap value. Smallest circles represent the lower cutoff of 50 %

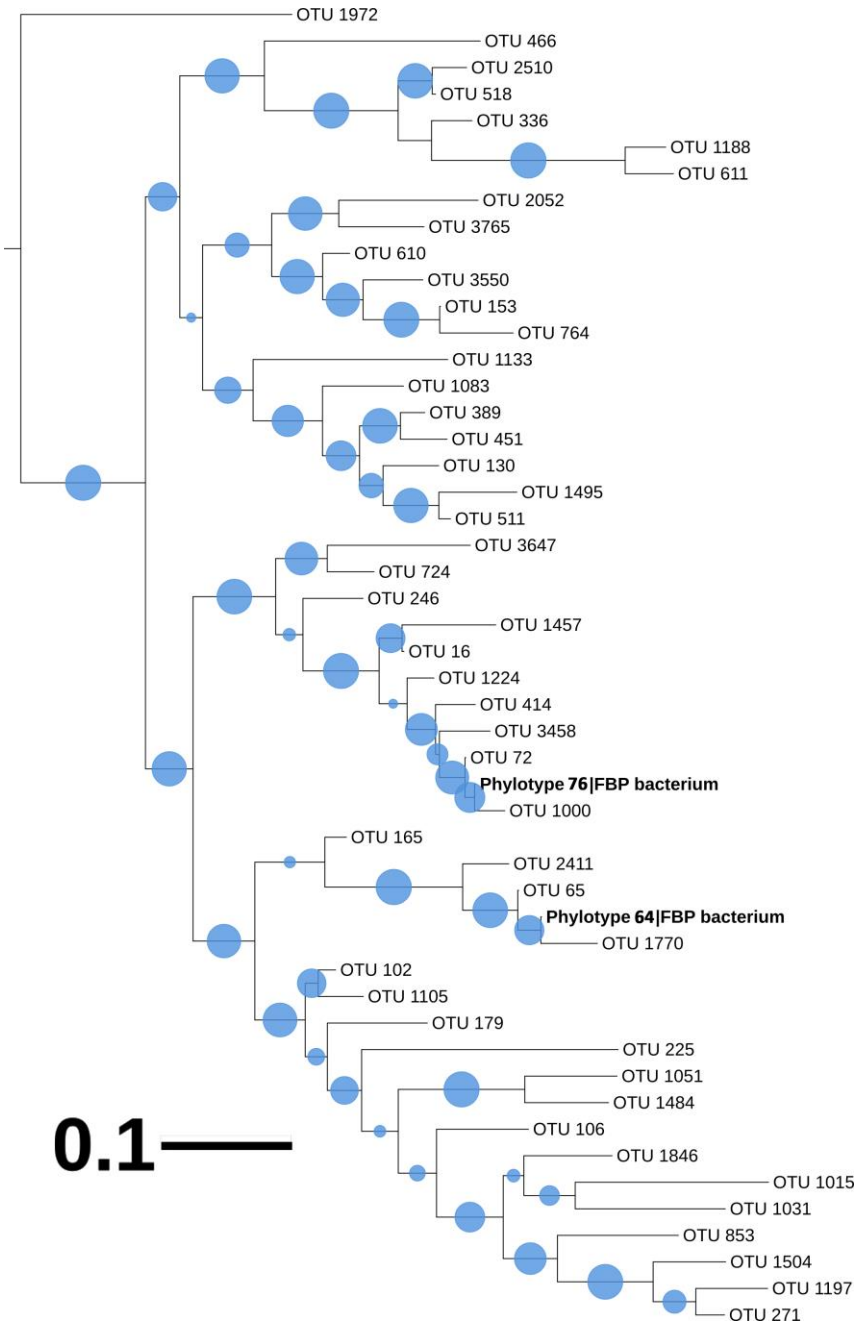

Figure S2 Detailed view of the FBP cluster of the maximum likelihood phylogenetic tree (1000 bootstraps) combining partial 16S rRNA gene sequences obtained using a culture-independent and –dependent approach. Representatives of the isolate phylotypes are labelled in bold. For Illumina OTUs [Tahon et al. Submitted (Chapter 3)], identifications at the most detailed taxonomic level are given. Scale bar indicates 0.1 substitutions per position. Bootstrap values are displayed as circles with a diameter reflecting the height of the bootstrap value. Smallest circles represent the lower cutoff of 50 %

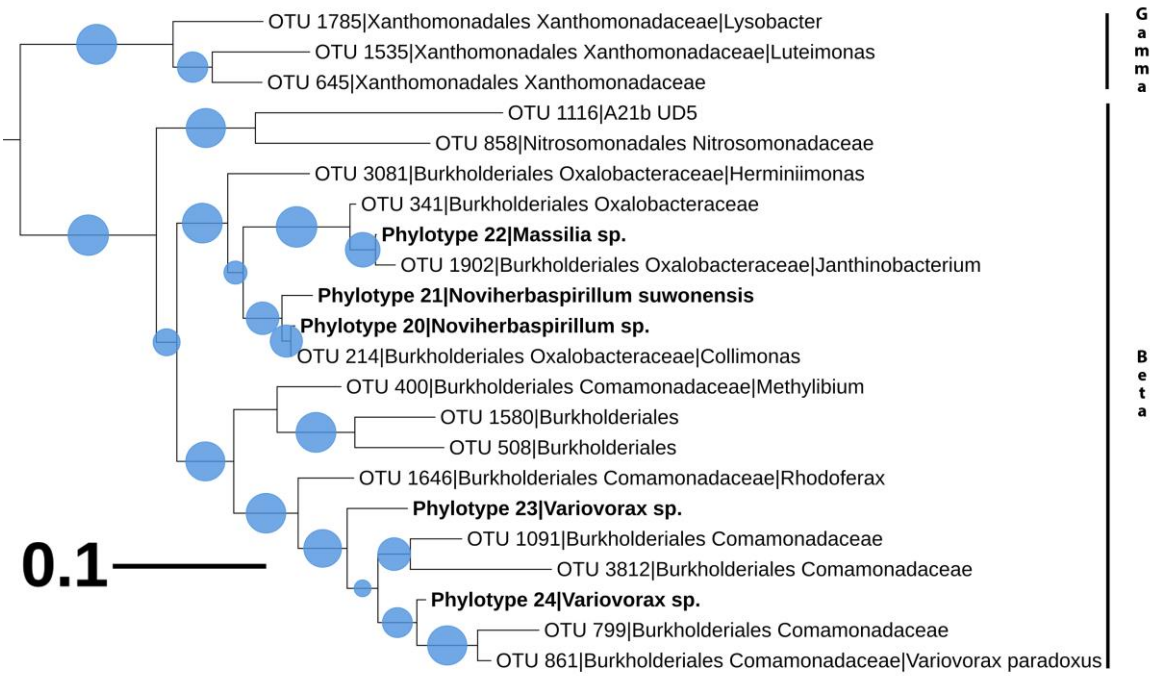

Figure S3 Detailed view of the Beta- and Gammaproteobacteria cluster of the maximum likelihood phylogenetic tree (1000 bootstraps) combining partial 16S rRNA gene sequences obtained using a culture-independent and –dependent approach. Representatives of the isolate phylotypes are labelled in bold. For Illumina OTUs [Tahon et al. Submitted (Chapter 3)], identifications at the most detailed taxonomic level are given. Scale bar indicates 0.1 substitutions per position. Bootstrap values are displayed as circles with a diameter reflecting the height of the bootstrap value. Smallest circles represent the lower cutoff of 50 %

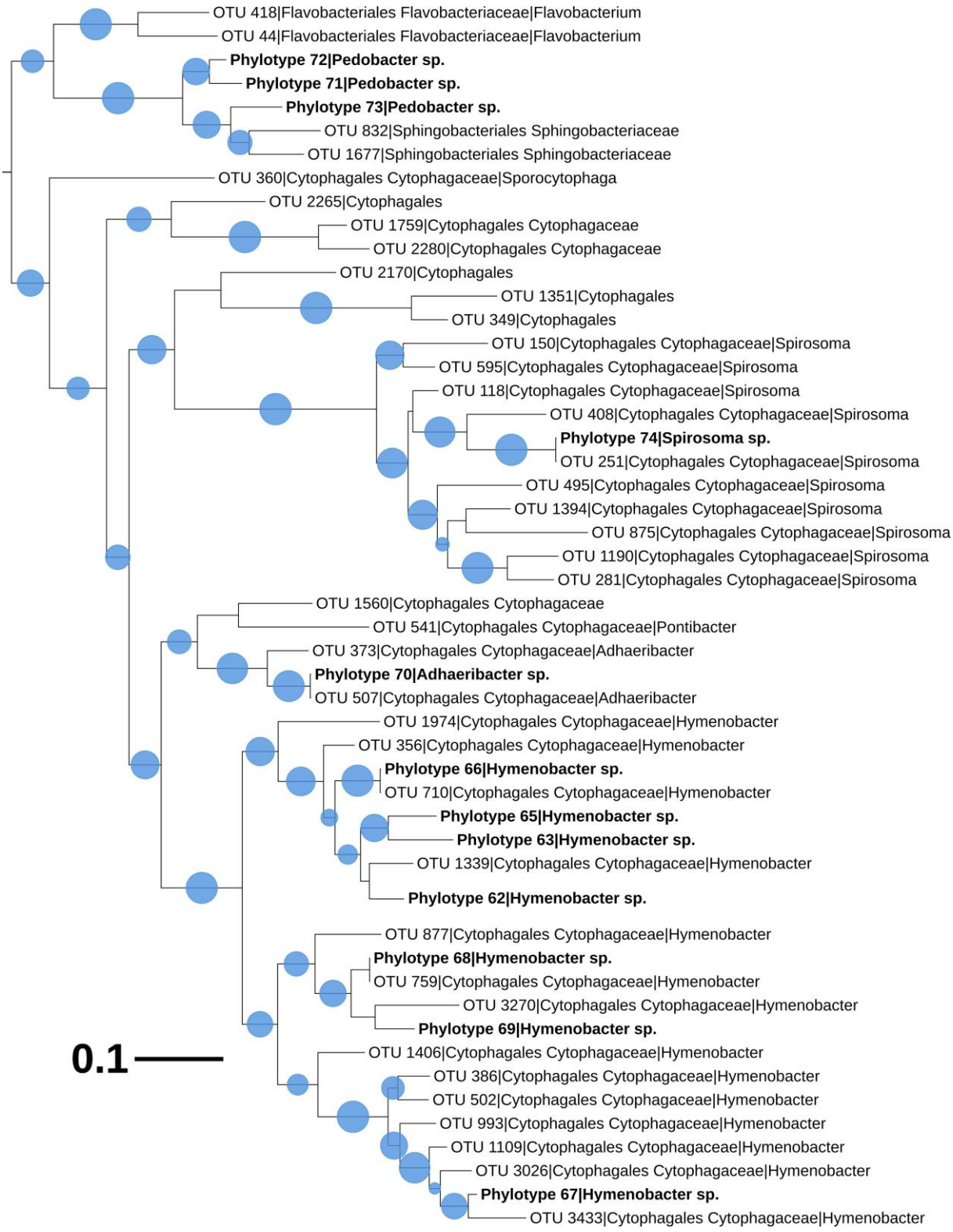

Figure S4 Detailed view of the Bacteroidetes cluster of the maximum likelihood phylogenetic tree (1000 bootstraps) combining partial 16S rRNA gene sequences obtained using a culture-independent and –dependent approach. Illumina sequences grouping with the orders Saprospirales and Rhodothermales are not included in the view. Representatives of the isolate phylotypes are labelled in bold. For Illumina OTUs [Tahon et al. Submitted (Chapter 3)], identifications at the most detailed taxonomic level are given. Scale bar indicates 0.1 substitutions per position. Bootstrap values are displayed as circles with a diameter reflecting the height of the bootstrap value. Smallest circles represent the lower cutoff of 50 %

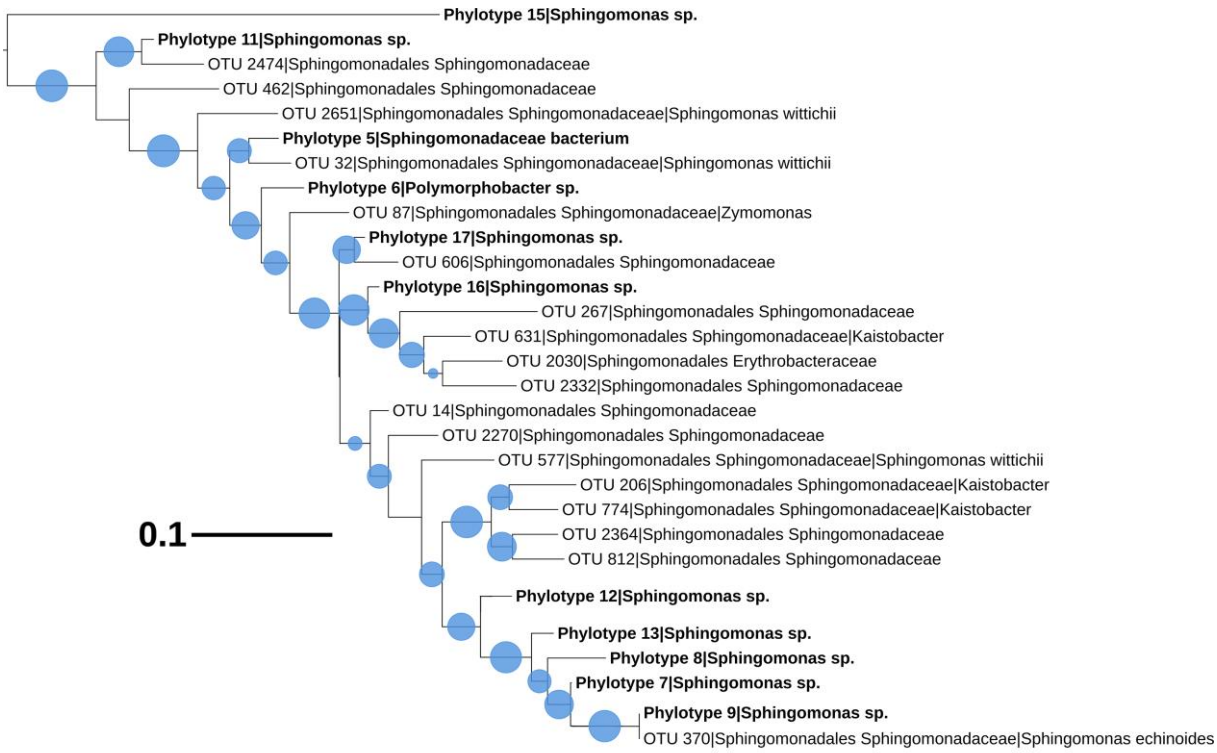

Figure S5 Detailed view of the Sphingomonadales cluster (Alphaproteobacteria) of the maximum likelihood phylogenetic tree (1000 bootstraps) combining partial 16S rRNA gene sequences obtained using a culture-independent and –dependent approach. Representatives of the isolate phylotypes are labelled in bold. For Illumina OTUs [Tahon et al. Submitted (Chapter 3)], identifications at the most detailed taxonomic level are given. Scale bar indicates 0.1 substitutions per position. Bootstrap values are displayed as circles with a diameter reflecting the height of the bootstrap value. Smallest circles represent the lower cutoff of 50 %

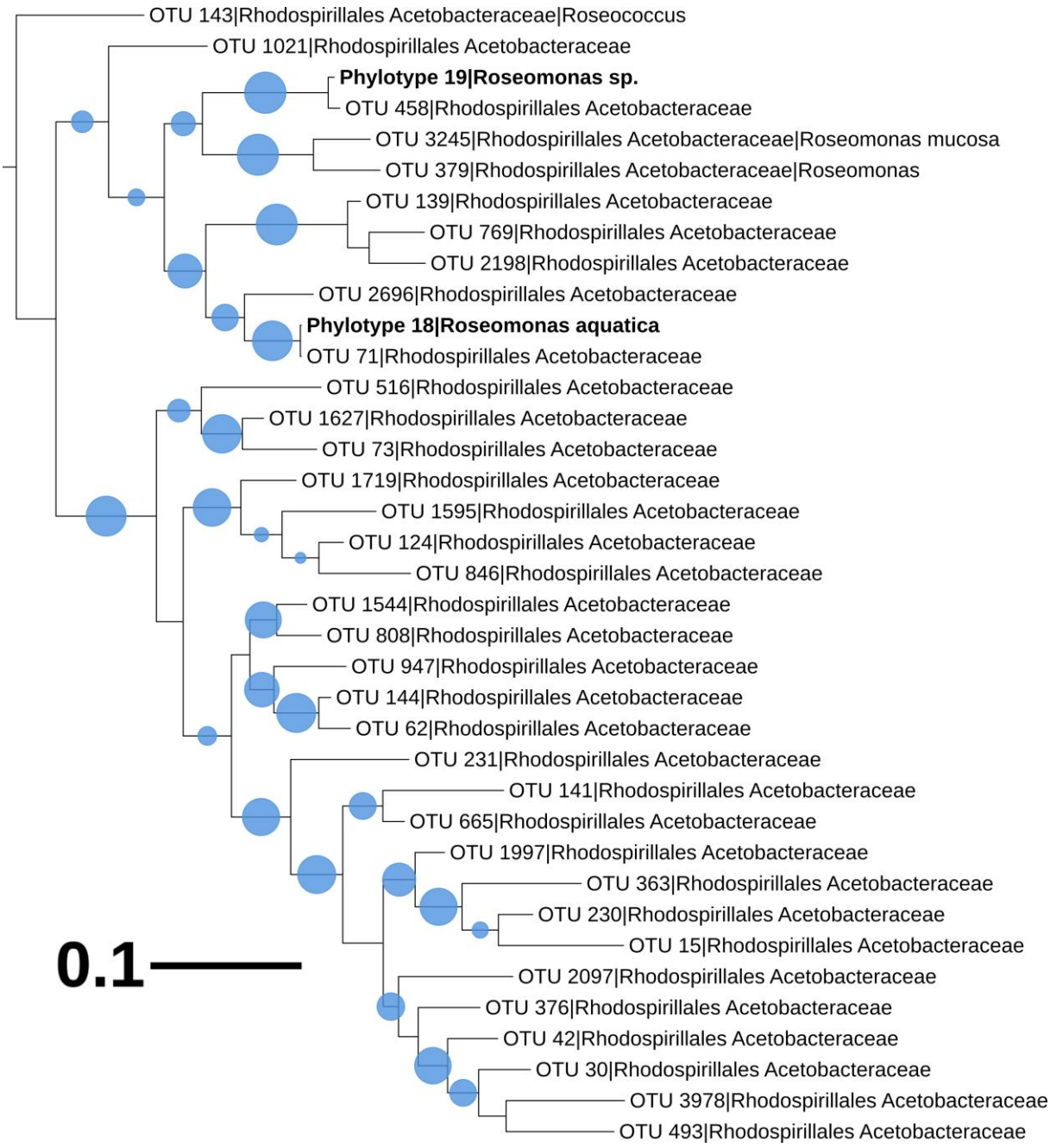

Figure S6 Detailed view of the Rhodospirillales cluster (Alphaproteobacteria) of the maximum likelihood phylogenetic tree (1000 bootstraps) combining partial 16S rRNA gene sequences obtained using a culture-independent and –dependent approach. Representatives of the isolate phylotypes are labelled in bold. For Illumina OTUs [Tahon et al. Submitted (Chapter 3)], identifications at the most detailed taxonomic level are given. Scale bar indicates 0.1 substitutions per position. Bootstrap values are displayed as circles with a diameter reflecting the height of the bootstrap value. Smallest circles represent the lower cutoff of 50 %

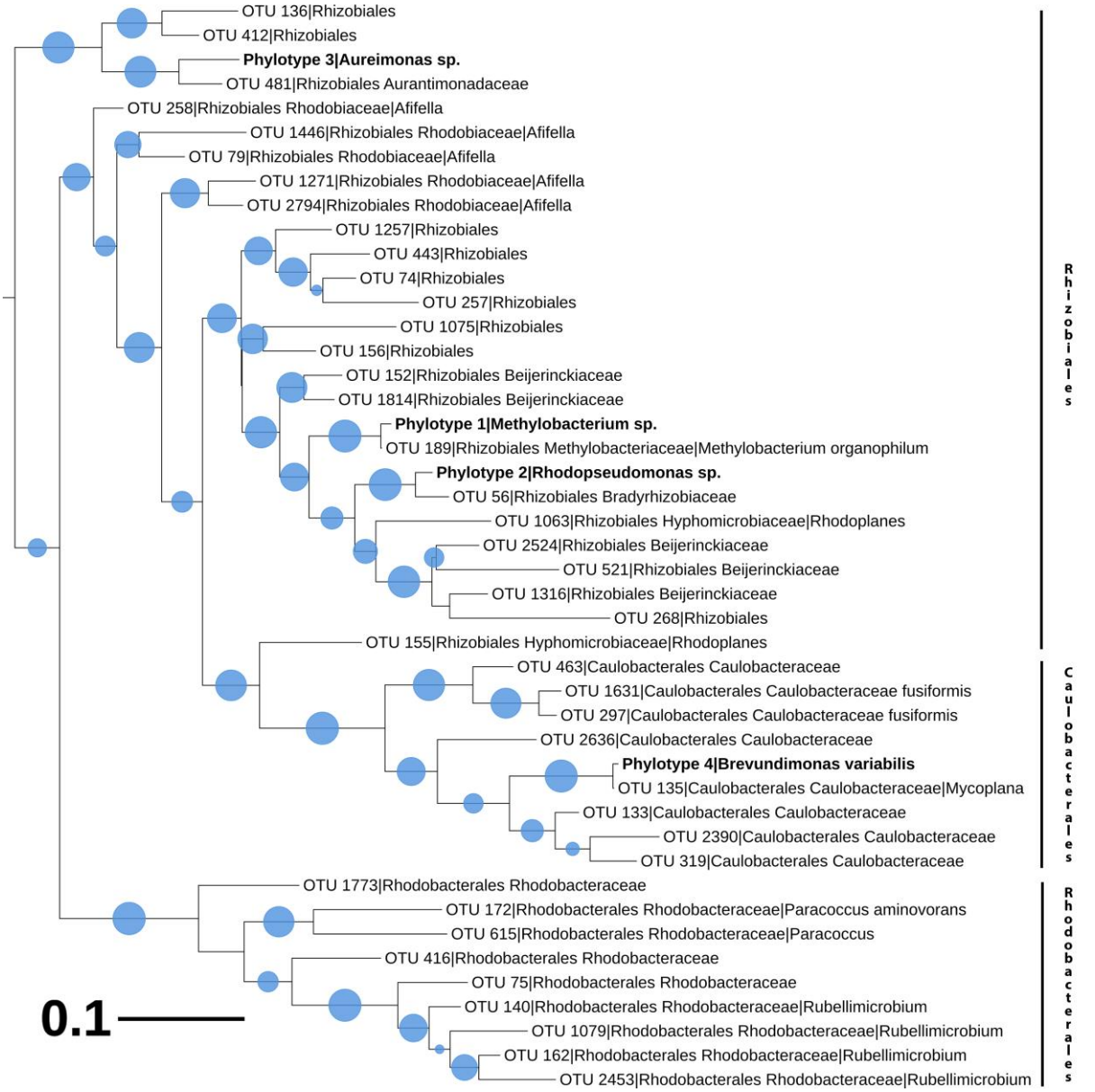

Figure S7 Detailed view of the Caulobacterales, Rhizobiales and Rhodobacterales clusters (Alphaproteobacteria) of the maximum likelihood phylogenetic tree (1000 bootstraps) combining partial 16S rRNA gene sequences obtained using a culture-independent and –dependent approach. Representatives of the isolate phylotypes are labelled in bold. For Illumina OTUs [Tahon et al. Submitted (Chapter 3)], identifications at the most detailed taxonomic level are given. Scale bar indicates 0.1 substitutions per position. Bootstrap values are displayed as circles with a diameter reflecting the height of the bootstrap value. Smallest circles represent the lower cutoff of 50 %

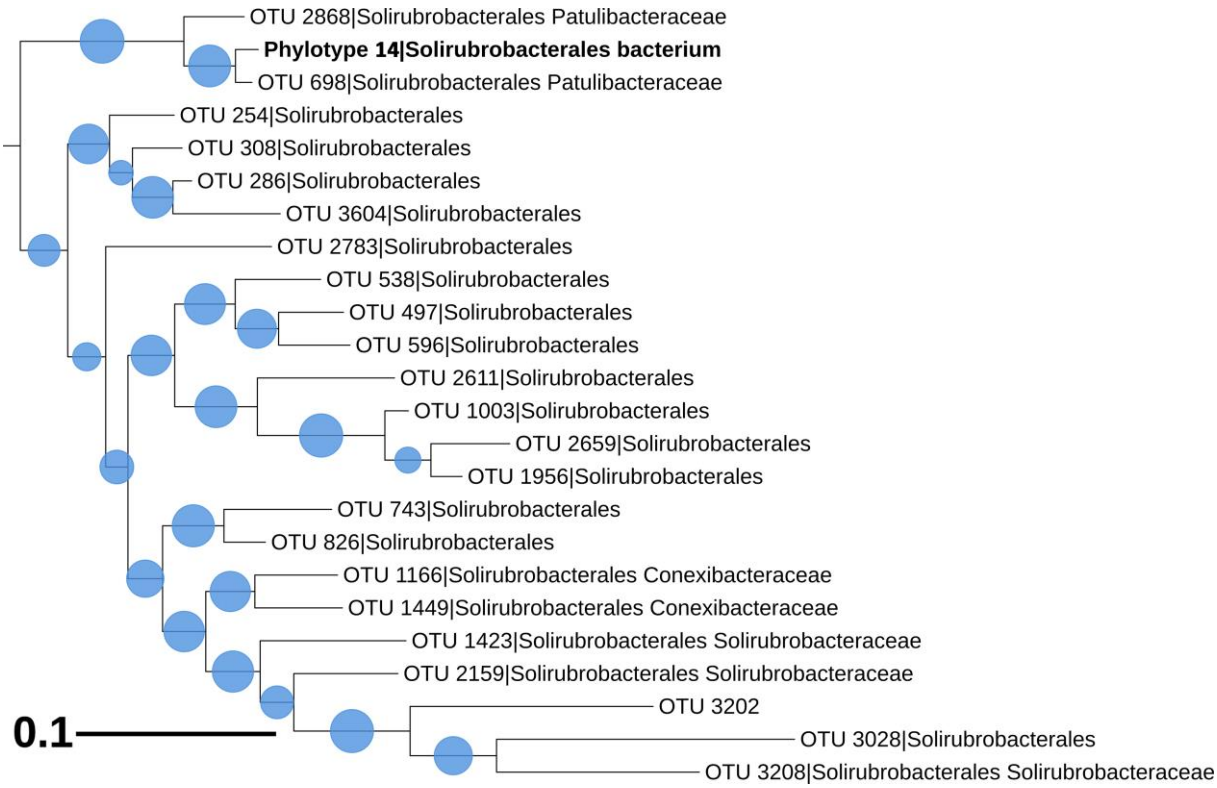

Figure S8 Detailed view of the Solirubrobacterales cluster (Actinobacteria) of the maximum likelihood phylogenetic tree (1000 bootstraps) combining partial 16S rRNA gene sequences obtained using a culture-independent and –dependent approach. Representatives of the isolate phylotypes are labelled in bold. For Illumina OTUs [Tahon et al. Submitted (Chapter 3)], identifications at the most detailed taxonomic level are given. Scale bar indicates 0.1 substitutions per position. Bootstrap values are displayed as circles with a diameter reflecting the height of the bootstrap value. Smallest circles represent the lower cutoff of 50 %

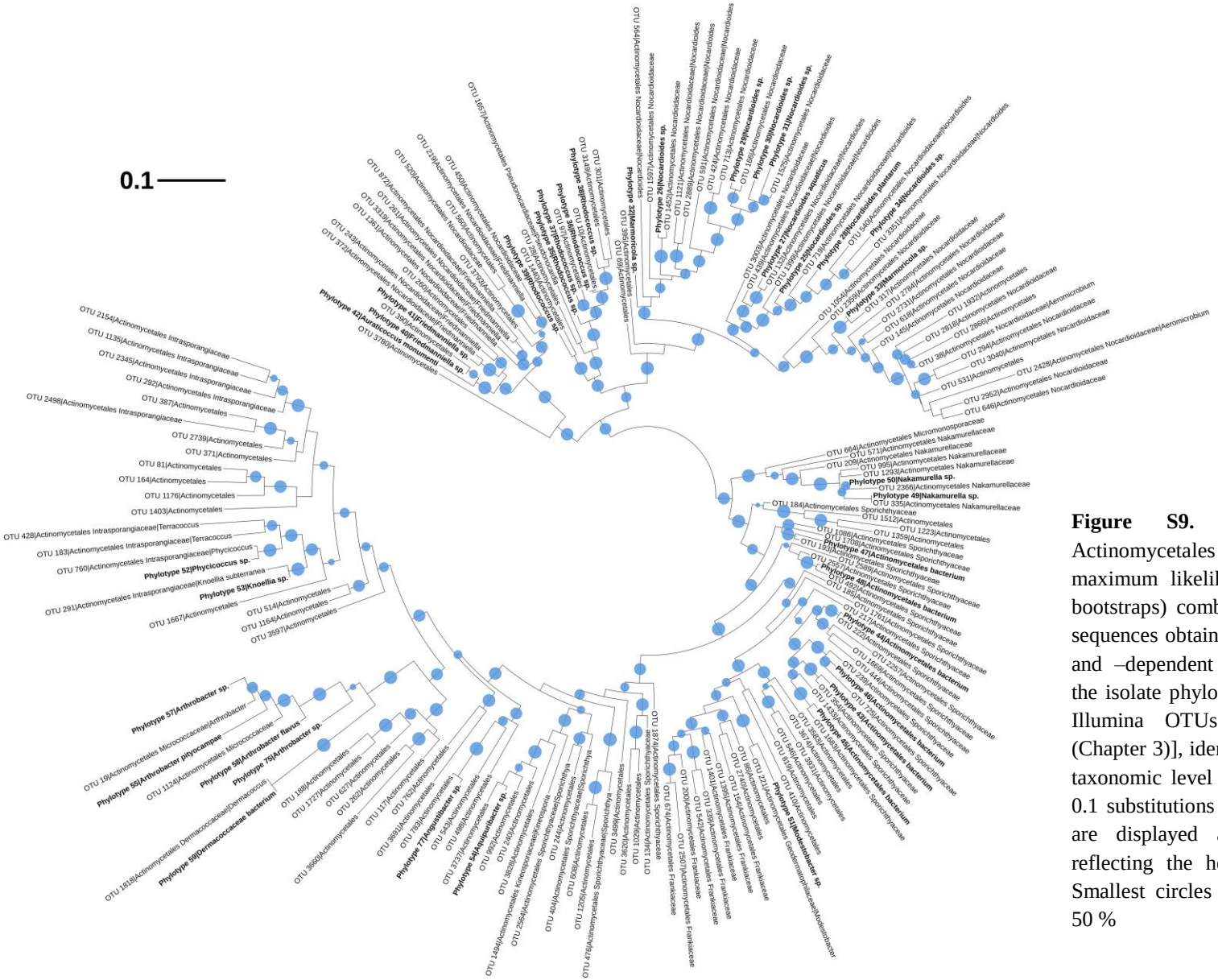

Figure S9. Detailed view of the Actinomycetales cluster (Actinobacteria) of the maximum likelihood phylogenetic tree (1000 bootstraps) combining partial 16S rRNA gene sequences obtained using a culture-independent and –dependent approach. Representatives of the isolate phylotypes are labelled in bold. For Illumina OTUs [Tahon et al. Submitted (Chapter 3)], identifications at the most detailed taxonomic level are given. Scale bar indicates 0.1 substitutions per position. Bootstrap values are displayed as circles with a diameter reflecting the height of the bootstrap value. Smallest circles represent the lower cutoff of 50 %

Table S1 Number of isolates per condition

	4 °C			15 °C		
PH medium	**Solid**	**Liquid shaken**	**Liquid not-shaken**	**Solid**	**Liquid shaken**	**Liquid not-shaken**
KP2	45	30	30	45	20	30
KP15	30	30	30	45	20	20
KP43	45	32	30	45	30	30
KP53	30	32	30	45	30	25
	4 °C			**15 °C**		
PA medium	**Solid**	**Liquid shaken**	**Liquid not-shaken**	**Solid**	**Liquid shaken**	**Liquid not-shaken**
KP2	35	38	38	45	20	20
KP15	41	44	34	45	20	30
KP43	45	30	30	45	20	20
KP53	30	30	28	45	20	20

Table S2 Detailed overview of all 330 represenative isolates enclosed in the 77 phylotypes listed in Table 2. All strains were tested for presence of *nifH*/*bchL*/*bchX* and *pufLM*, however, only positive PCR results are indicated

Phylotype	Strain	Accession no.	Nearest phylogenetic neighbor			*nifH*/*bchL*/*bchX*	*pufLM*
			ID	Sequence similarity (%)	Accession no.		
1	R-68457	KY386309	*Methylobacterium iners* 5317S-33				+
1	KP15.18.4.PA.S	KY386323	*Methylobacterium iners* 5317S-33				+
1	R-68525	KY386327	*Methylobacterium iners* 5317S-33				+
1	KP15.21.4.PA.S	KY386331	*Methylobacterium iners* 5317S-33				+
1	R-68173	KY386380	*Methylobacterium iners* 5317S-33	98.52	EF174497	+	
1	R-68388	KY386396	*Methylobacterium iners* 5317S-33				+
1	R-68391	KY386399	*Methylobacterium iners* 5317S-33				+
1	R-68554	KY386402	*Methylobacterium iners* 5317S-33				+
1	R-68669	KY386414	*Methylobacterium iners* 5317S-33			+	+
1	KP2.9b.4.PA.S	KY386471	*Methylobacterium iners* 5317S-33				+
2	R-67878	KY386506	*Rhodopseudomonas pseudopalustris* DSM123	98.67	AB498818		
3	R-68373	KY386318	*Aureimonas ferruginae* CC-CFT023	96.64	JQ864240	+	
4	R-68476	KY386398	*Brevundimonas variabilis* ATCC15255				+
4	R-68394	KY386404	*Brevundimonas variabilis* ATCC15255				+
4	KP2.19.15.PH.NS	KY386407	*Brevundimonas variabilis* ATCC15255				+
4	R-68473	KY386409	*Brevundimonas variabilis* ATCC15255			+	+
4	R-68297	KY386412	*Brevundimonas variabilis* ATCC15255				+
4	R-68298	KY386416	*Brevundimonas variabilis* ATCC15255			+	+
4	R-68299	KY386422	*Brevundimonas variabilis* ATCC15255			+	+
4	R-68483	KY386436	*Brevundimonas variabilis* ATCC15255			+	+
4	R-67742	KY386439	*Brevundimonas variabilis* ATCC15255			+	
4	R-68295	KY386469	*Brevundimonas variabilis* ATCC15255	99.79	AJ227783	+	+
5	R-67883	KY386496	*Polymorphobacter multimanifer* 262-7	94.94	AB649056	+	
5	KP53.16.15.PA.NS	KY386580	*Polymorphobacter multimanifer* 262-7				+
6	R-68699	KY386562	*Polymorphobacter multimanifer* 262-7	95.99	AB649056		
7	R-68084	KY386306	*Sphingomonas faeni* MA-olki				
7	KP15.16.4.PA.V	KY386315	*Sphingomonas faeni* MA-olki				
7	R-67991	KY386320	*Sphingomonas faeni* MA-olki				+
7	R-68089	KY386319	*Sphingomonas faeni* MA-olki				
7	R-68414	KY386322	*Sphingomonas faeni* MA-olki				
7	R-68418	KY386339	*Sphingomonas faeni* MA-olki				
7	R-68527	KY386348	*Sphingomonas faeni* MA-olki				
7	R-68539	KY386359	*Sphingomonas faeni* MA-olki				
7	R-67789	KY386360	*Sphingomonas faeni* MA-olki				

Phylotype	Strain	Accession no.	Nearest phylogenetic neighbor				
			ID	Sequence similarity (%)	Accession no.	*nifH*/*bchL*/*bchX*	*pufLM*
7	R-68448	KY386370	*Sphingomonas faeni* MA-olki			+	
7	R-68170	KY386374	*Sphingomonas faeni* MA-olki				
7	R-67846	KY386377	*Sphingomonas faeni* MA-olki			+	
7	R-67984	KY386379	*Sphingomonas faeni* MA-olki	99.62	AJ429239	+	
7	KP2.34b.15.PH.V	KY386451	*Sphingomonas faeni* MA-olki				+
7	R-67744	KY386468	*Sphingomonas faeni* MA-olki				+
7	KP43.10b.4.PH.V	KY386477	*Sphingomonas faeni* MA-olki				+
7	R-68208	KY386485	*Sphingomonas faeni* MA-olki			+	
7	R-68340	KY386497	*Sphingomonas faeni* MA-olki				
8	R-68700	KY386453	*Sphingomonas oligophenolica* JCM 12082	98.24	AB018439		
9	R-68087	KY386314	*Sphingomonas aerolata* NW12				
9	R-68461	KY386351	*Sphingomonas aerolata* NW12				
9	KP2.23.15.PH.NS	KY386419	*Sphingomonas aerolata* NW12				
9	R-68261	KY386466	*Sphingomonas aerolata* NW12				
9	R-68499	KY386473	*Sphingomonas aerolata* NW12				
9	R-68344	KY386472	*Sphingomonas aerolata* NW12				
9	R-68493	KY386479	*Sphingomonas aerolata* NW12				
9	R-68425	KY386482	*Sphingomonas aerolata* NW12				
9	R-68427	KY386487	*Sphingomonas aerolata* NW12				
9	R-68341	KY386490	*Sphingomonas aerolata* NW12				
9	R-68103	KY386491	*Sphingomonas aerolata* NW12				
9	R-68279	KY386493	*Sphingomonas aerolata* NW12				
9	R-68104	KY386499	*Sphingomonas aerolata* NW12			+	+
9	R-68435	KY386501	*Sphingomonas aerolata* NW12				+
9	R-68105	KY386502	*Sphingomonas aerolata* NW12				
9	R-68335	KY386511	*Sphingomonas aerolata* NW12			+	
9	R-68336	KY386515	*Sphingomonas aerolata* NW12				
9	KP43.28.4.PH.NS	KY386528	*Sphingomonas aerolata* NW12				
9	R-68566	KY386531	*Sphingomonas aerolata* NW12				
9	R-68274	KY386532	*Sphingomonas aerolata* NW12	99.09	AJ429240		
9	R-68308	KY386534	*Sphingomonas aerolata* NW12			+	
9	R-68023	KY386540	*Sphingomonas aerolata* NW12				
9	R-68025	KY386544	*Sphingomonas aerolata* NW12				
9	R-68275	KY386549	*Sphingomonas aerolata* NW12				
9	R-68232	KY386553	*Sphingomonas aerolata* NW12				
9	KP43.5.15.PH.S	KY386557	*Sphingomonas aerolata* NW12				
9	R-68305	KY386561	*Sphingomonas aerolata* NW12				+

Phylotype	Strain	Accession no.	Nearest phylogenetic neighbor				
			ID	Sequence similarity (%)	Accession no.	*nifH*/*bchL*/*bchX*	*pufLM*
9	R-68495	KY386563	*Sphingomonas aerolata* NW12			+	
9	R-68343	KY386567	*Sphingomonas aerolata* NW12				
9	R-68309	KY386568	*Sphingomonas aerolata* NW12				
9	R-67894	KY386575	*Sphingomonas aerolata* NW12				
9	KP53.14.4.PH.S	KY386576	*Sphingomonas aerolata* NW12				
9	R-68345	KY386579	*Sphingomonas aerolata* NW12				
9	R-68383	KY386582	*Sphingomonas aerolata* NW12			+	
9	R-68291	KY386584	*Sphingomonas aerolata* NW12			+	
9	R-68245	KY386586	*Sphingomonas aerolata* NW12				
9	R-68129	KY386587	*Sphingomonas aerolata* NW12				
9	R-68246	KY386589	*Sphingomonas aerolata* NW12				
9	R-68445	KY386594	*Sphingomonas aerolata* NW12				
9	R-68042	KY386595	*Sphingomonas aerolata* NW12				
9	R-68381	KY386598	*Sphingomonas aerolata* NW12				
9	R-68380	KY386601	*Sphingomonas aerolata* NW12				
9	R-68046	KY386603	*Sphingomonas aerolata* NW12				
9	R-68578	KY386606	*Sphingomonas aerolata* NW12				
9	R-68351	KY386607	*Sphingomonas aerolata* NW12				
9	R-68574	KY386613	*Sphingomonas aerolata* NW12				
9	R-68036	KY386622	*Sphingomonas aerolata* NW12				
9	R-68349	KY386621	*Sphingomonas aerolata* NW12				
9	R-68512	KY386624	*Sphingomonas aerolata* NW12				
9	R-68037	KY386625	*Sphingomonas aerolata* NW12			+	
9	KP53.9.15.PH.S	KY386626	*Sphingomonas aerolata* NW12				
9	R-68348	KY386628	*Sphingomonas aerolata* NW12				
9	KP53.9a.4.PH.NS	KY386629	*Sphingomonas aerolata* NW12				
11	R-68222	KY386535	*Sphingomonas yunnanensis* YIM 003	97.52	AY894691		
12	R-68269	KY386302	*Sphingomonas cynarae* SPC-1				
12	R-68271	KY386310	*Sphingomonas cynarae* SPC-1				+
12	R-68375	KY386329	*Sphingomonas cynarae* SPC-1			+	
12	R-68304	KY386332	*Sphingomonas cynarae* SPC-1	98.24	HQ439186		+
12	R-67996	KY386342	*Sphingomonas cynarae* SPC-1				
12	R-68199	KY386373	*Sphingomonas cynarae* SPC-1			+	
12	R-68171	KY386378	*Sphingomonas cynarae* SPC-1				
12	R-68355	KY386381	*Sphingomonas cynarae* SPC-1				
12	R-68392	KY386400	*Sphingomonas cynarae* SPC-1				+
12	R-68317	KY386417	*Sphingomonas cynarae* SPC-1				+

Phylotype	Strain	Accession no.	Nearest phylogenetic neighbor				
			ID	Sequence similarity (%)	Accession no.	*nifH*/*bchL*/*bchX*	*pufLM*
12	KP2.25b.4.PA.NS	KY386428	*Sphingomonas cynarae* SPC-1			+	+
12	KP2.32b.4.PA.NS	KY386447	*Sphingomonas cynarae* SPC-1				+
13	R-68091	KY386335	*Sphingomonas mucosissima* CP173-2				
13	R-67999	KY386352	*Sphingomonas mucosissima* CP173-2				
13	R-68524	KY386357	*Sphingomonas mucosissima* CP173-2	98.27	AM229669		
15	R-67917	KY386305	*Sphingomonas yantingensis* 1007			+	+
15	R-68270	KY386304	*Sphingomonas yantingensis* 1007	96.79	JX566547	+	+
15	KP15.19.15.PH.S	KY386324	*Sphingomonas yantingensis* 1007			+	+
15	KP15.20.15.PH.V	KY386328	*Sphingomonas yantingensis* 1007				+
15	R-67859	KY386337	*Sphingomonas yantingensis* 1007			+	+
15	R-68007	KY386366	*Sphingomonas yantingensis* 1007			+	+
15	R-68267	KY386368	*Sphingomonas yantingensis* 1007				+
16	R-68405	KY386312	*Sphingomonas hunanensis* JSM 083058			+	+
16	R-67790	KY386362	*Sphingomonas hunanensis* JSM 083058				
16	KP15.4.15.PH.NS	KY386367	*Sphingomonas hunanensis* JSM 083058				+
16	R-68410	KY386375	*Sphingomonas hunanensis* JSM 083058			+	+
16	R-68292	KY386383	*Sphingomonas hunanensis* JSM 083058				+
16	R-68553	KY386406	*Sphingomonas hunanensis* JSM 083058				+
16	R-68552	KY386415	*Sphingomonas hunanensis* JSM 083058				+
16	R-68400	KY386424	*Sphingomonas hunanensis* JSM 083058				+
16	R-68370	KY386440	*Sphingomonas hunanensis* JSM 083058				+
16	R-68161	KY386455	*Sphingomonas hunanensis* JSM 083058			+	+
16	R-68260	KY386460	*Sphingomonas hunanensis* JSM 083058	98.52	FJ527417	+	+
16	R-68242	KY386578	*Sphingomonas hunanensis* JSM 083058				
17	R-67954	KY386585	*Sphingomonas pituitosa* EDIV	95.58	AJ243751		
18	R-68475	KY386403	*Roseomonas aquatica* TR53	99.93	AM231587		+
19	R-68165	KY386462	*Roseomonas tokyonensis* K-20	98.6	AB297501	+	
20	R-67997	KY386345	*Noviherbaspirillum psychrotolerans* PB1	98.62	JN390675		
20	R-68490	KY386488	*Noviherbaspirillum psychrotolerans* PB1				
21	R-68579	KY386599	*Noviherbaspirillum suwonense* 54105-65	98.95	JX275858		
22	R-67978	KY386448	*Massilia eurypsychrophila* B528-3	97.7	KJ361504		
22	R-67982	KY386454	*Massilia eurypsychrophila* B528-3				
23	R-67959	KY386465	*Variovorax boronicumulans* BAM-48				
23	R-67932	KY386486	*Variovorax boronicumulans* BAM-48	97.55	AB300597		
23	KP43.23.4.PA.S	KY386522	*Variovorax boronicumulans* BAM-48				
23	R-68338	KY386543	*Variovorax boronicumulans* BAM-48				
23	R-67927	KY386558	*Variovorax boronicumulans* BAM-48				

Phylotype	Strain	Accession no.	Nearest phylogenetic neighbor				
			ID	Sequence similarity (%)	Accession no.	*nifH*/*bchL*/*bchX*	*pufLM*
24	R-67871	KY386393	*Variovorax ginsengisoli* Gsoil 3165	99.26	AB245358		
24	R-67891	KY386627	*Variovorax ginsengisoli* Gsoil 3165				
25	KP15.17.15.PA.V	KY386317	*Nocardioides ginsengagri* BX5-10	97.38	GQ339904		
25	R-68562	KY386518	*Nocardioides ginsengagri* BX5-10				
26	R-68482	KY386444	*Nocardioides salarius* CL-Z59	97	DQ401092		
27	R-68162	KY386458	*Nocardioides aquaticus* EL-17K	99.38	X94145		
28	KP43.13.15.PH.NS	KY386484	*Nocardioides plantarum* NCIMB 12834	100	AF005008		
28	R-68307	KY386526	*Nocardioides plantarum* NCIMB 12834				
28	KP43.9.15.PH.NS	KY386564	*Nocardioides plantarum* NCIMB 12834				
29	KP2.10.4.PA.S	KY386388	*Nocardioides terrigena* DS-17				
29	R-68053	KY386389	*Nocardioides terrigena* DS-17				
29	R-67906	KY386390	*Nocardioides terrigena* DS-17				
29	KP2.14.4.PH.V	KY386397	*Nocardioides terrigena* DS-17				
29	R-68364	KY386401	*Nocardioides terrigena* DS-17				
29	R-68474	KY386405	*Nocardioides terrigena* DS-17				
29	R-68550	KY386421	*Nocardioides terrigena* DS-17				
29	R-68154	KY386423	*Nocardioides terrigena* DS-17	97.79	EF363712		
29	R-67971	KY386425	*Nocardioides terrigena* DS-17				
29	R-68065	KY386430	*Nocardioides terrigena* DS-17				
29	R-67865	KY386438	*Nocardioides terrigena* DS-17				
29	R-67756	KY386443	*Nocardioides terrigena* DS-17				
29	R-68140	KY386461	*Nocardioides terrigena* DS-17			+	
29	R-68014	KY386492	*Nocardioides terrigena* DS-17				
30	R-67827	KY386592	*Nocardioides terrigena* DS-17	96.06	EF363712		
31	R-68372	KY386385	*Nocardioides terrigena* DS-17				
31	R-68145	KY386386	*Nocardioides terrigena* DS-17	97.52	EF363712		
31	R-67747	KY386391	*Nocardioides terrigena* DS-17			+	
31	KP2.23.15.PH.V	KY386420	*Nocardioides terrigena* DS-17				
31	R-68158	KY386446	*Nocardioides terrigena* DS-17				
31	R-68160	KY386452	*Nocardioides terrigena* DS-17				
31	R-68167	KY386463	*Nocardioides terrigena* DS-17				
32	R-67776	KY386307	*Marmoricola korecus* Sco-A36				
32	R-67781	KY386333	*Marmoricola korecus* Sco-A36	98.67	FN386723		
33	R-68467	KY386387	*Marmoricola aquaticus* CBMAI 1089				
33	R-67804	KY386524	*Marmoricola aquaticus* CBMAI 1089	97.75	JN615437		
34	R-67800	KY386498	*Nocardioides antarcticus* M-SA3-94	97.76	KM347967		
34	R-68564	KY386504	*Nocardioides antarcticus* M-SA3-94				

Phylotype	Strain	Accession no.	Nearest phylogenetic neighbor				
			ID	Sequence similarity (%)	Accession no.	*nifH*/*bchL*/*bchX*	*pufLM*
34	R-68503	KY386512	*Nocardioides antarcticus* M-SA3-94				
34	R-68432	KY386521	*Nocardioides antarcticus* M-SA3-94				
34	R-68485	KY386533	*Nocardioides antarcticus* M-SA3-94				
34	R-68233	KY386554	*Nocardioides antarcticus* M-SA3-94				
34	R-68486	KY386559	*Nocardioides antarcticus* M-SA3-94				
34	R-68497	KY386560	*Nocardioides antarcticus* M-SA3-94				
34	R-68352	KY386605	*Nocardioides antarcticus* M-SA3-94				
35	R-68019	KY386523	*Rhodococcus aerolatus* PAMC 27367	96.34	KM044053		
35	R-68235	KY386556	*Rhodococcus aerolatus* PAMC 27367				
36	R-68187	KY386338	*Rhodococcus aerolatus* PAMC 27367	95.64	KM044053	+	
37	R-68509	KY386574	*Rhodococcus aerolatus* PAMC 27367	95.92	KM044053		
38	R-67872	KY386395	*Rhodococcus aerolatus* PAMC 27367	95.85	KM044053		
38	R-68258	KY386617	*Rhodococcus aerolatus* PAMC 27367				
39	R-68460	KY386303	*Rhodococcus fascians* LMG 3623			+	
39	R-68273	KY386321	*Rhodococcus fascians* LMG 3623	98.77	JMEN01000010		
39	R-68530	KY386341	*Rhodococcus fascians* LMG 3623				
39	R-68463	KY386344	*Rhodococcus fascians* LMG 3623				
39	KP15.30.4.PA.NS	KY386358	*Rhodococcus fascians* LMG 3623			+	
39	R-68196	KY386364	*Rhodococcus fascians* LMG 3623				
39	R-68515	KY386604	*Rhodococcus fascians* LMG 3623				
40	R-68537	KY386361	*Friedmanniella sagamiharensis* FB2				
40	R-67749	KY386408	*Friedmanniella sagamiharensis* FB2	96.09	AB445456		
40	R-67754	KY386434	*Friedmanniella sagamiharensis* FB2				
40	R-67745	KY386470	*Friedmanniella sagamiharensis* FB2				
40	R-68220	KY386525	*Friedmanniella sagamiharensis* FB2				
41	R-68221	KY386527	*Friedmanniella luteola* FA1	98.22	AB445453		
42	R-68201	KY386505	*Auraticoccus monumenti* MON 2.2	99.58	FN552748		
43	R-67786	KY386350	*Jatrophihabitans endophyticus* S9650	93.6	JQ346802		
43	R-68511	KY386570	*Jatrophihabitans endophyticus* S9650				
44	R-68701	KY386494	*Jatrophihabitans soli* KIS75-12	94.37	KP017569		
45	R-68182	KY386325	*Frankia alni* ACN14A				
45	R-68183	KY386330	*Frankia alni* ACN14A	94.11	CT573213		
45	R-67752	KY386426	*Frankia alni* ACN14A				
45	R-67882	KY386481	*Frankia alni* ACN14A				
45	R-68215	KY386514	*Frankia alni* ACN14A				
45	R-67881	KY386566	*Frankia alni* ACN14A				
45	R-67888	KY386611	*Frankia alni* ACN14A				

Phylotype	Strain	Accession no.	Nearest phylogenetic neighbor				
			ID	Sequence similarity (%)	Accession no.	*nifH*/*bchL*/*bchX*	*pufLM*
46	R-67810	KY386542	*Cryptosporangium minutisporangium* IFO15962	94.41	AB037007		
47	R-68148	KY386394	*Modestobacter versicolor* CP153-2				
47	R-67836	KY386609	*Modestobacter versicolor* CP153-2	94.28	AJ871304		
47	R-68237	KY386619	*Modestobacter versicolor* CP153-2				
48	R-68223	KY386537	*Sporichthya brevicatena* IFO 16195	93.78	AB006164		
48	R-68238	KY386569	*Sporichthya brevicatena* IFO 16195				
49	KP15.26.15.PA.V	KY386343	*Nakamurella lactae* DLS-10				
49	R-68216	KY386516	*Nakamurella lactae* DLS-10	98.07	AM778124		
50	R-67826	KY386590	*Nakamurella lactae* DLS-10				
50	R-67832	KY386596	*Nakamurella lactae* DLS-10				
50	R-67838	KY386610	*Nakamurella lactae* DLS-10	98.57	AM778124		
51	R-68190	KY386347	*Modestobacter lapidis* MON 3.1				
51	R-68535	KY386365	*Modestobacter lapidis* MON 3.1				
51	R-68354	KY386382	*Modestobacter lapidis* MON 3.1				
51	R-67763	KY386459	*Modestobacter lapidis* MON 3.1				
51	R-68492	KY386483	*Modestobacter lapidis* MON 3.1				
51	R-68563	KY386513	*Modestobacter lapidis* MON 3.1				
51	R-68230	KY386550	*Modestobacter lapidis* MON 3.1	97.28	LN810544		
52	R-68264	KY386392	*Phycicoccus ochangensis* L1b-b9	98.09	GQ344405		
52	R-68327	KY386427	*Phycicoccus ochangensis* L1b-b9				
52	R-68064	KY386429	*Phycicoccus ochangensis* L1b-b9				
52	R-67977	KY386445	*Phycicoccus ochangensis* L1b-b9				
52	KP2.38.4.PH.V	KY386456	*Phycicoccus ochangensis* L1b-b9				
52	R-67902	KY386464	*Phycicoccus ochangensis* L1b-b9				
52	R-68010	KY386474	*Phycicoccus ochangensis* L1b-b9				
52	R-68028	KY386548	*Phycicoccus ochangensis* L1b-b9				
52	KP43.40.4.PA.V	KY386551	*Phycicoccus ochangensis* L1b-b9				
53	R-68061	KY386418	*Knoellia sinensis* DSM12331	98.7	AVPJ01000034		
54	R-67798	KY386475	*Aquipuribacter hungaricus* IV-75			+	
54	R-67807	KY386530	*Aquipuribacter hungaricus* IV-75				
54	KP43.37.15.PA.V	KY386545	*Aquipuribacter hungaricus* IV-75	97.88	FM179321		
54	R-68204	KY386565	*Aquipuribacter hungaricus* IV-75				
55	R-68518	KY386301	*Arthrobacter pityocampae* Tp2	99.34	EU885749		
55	R-68186	KY386336	*Arthrobacter pityocampae* Tp2				
55	R-67785	KY386346	*Arthrobacter pityocampae* Tp2				
55	R-67998	KY386349	*Arthrobacter pityocampae* Tp2				
55	R-68224	KY386538	*Arthrobacter pityocampae* Tp2				

Phylotype	Strain	Accession no.	Nearest phylogenetic neighbor				
			ID	Sequence similarity (%)	Accession no.	*nifH*/*bchL*/*bchX*	*pufLM*
57	R-68156	KY386431	*Arthrobacter agilis* DSM20550				
57	R-68157	KY386442	*Arthrobacter agilis* DSM20550				
57	R-68477	KY386457	*Arthrobacter agilis* DSM20550				
57	R-68144	KY386467	*Arthrobacter agilis* DSM20550				
57	R-68102	KY386476	*Arthrobacter agilis* DSM20550				
57	KP43.11.4.PH.V	KY386478	*Arthrobacter agilis* DSM20550				
57	R-67801	KY386510	*Arthrobacter agilis* DSM20550				
57	R-68017	KY386519	*Arthrobacter agilis* DSM20550				
57	R-67803	KY386520	*Arthrobacter agilis* DSM20550				
57	R-67808	KY386536	*Arthrobacter agilis* DSM20550				
57	R-68027	KY386546	*Arthrobacter agilis* DSM20550				
57	R-68229	KY386547	*Arthrobacter agilis* DSM20550				
57	R-68231	KY386552	*Arthrobacter agilis* DSM20550				
57	R-68508	KY386577	*Arthrobacter agilis* DSM20550				
57	R-68247	KY386593	*Arthrobacter agilis* DSM20550				
57	R-68043	KY386597	*Arthrobacter agilis* DSM20550				
57	R-68045	KY386600	*Arthrobacter agilis* DSM20550				
57	KP53.28.4.PA.V	KY386602	*Arthrobacter agilis* DSM20550				
57	R-67839	KY386614	*Arthrobacter agilis* DSM20550				
57	R-68257	KY386616	*Arthrobacter agilis* DSM20550				
57	R-67817	KY386620	*Arthrobacter agilis* DSM20550				
57	R-67818	KY386623	*Arthrobacter agilis* DSM20550	99.46	X80748		
58	R-67793	KY386372	*Arthrobacter flavus* TB 23	99.25	ALPM01000083		
58	R-68397	KY386413	*Arthrobacter flavus* TB 23				
59	R-68209	KY386489	*Calidifontibacter indicus* PC IW02				
59	R-68253	KY386608	*Calidifontibacter indicus* PC IW02	94.71	EF187228		
59	KP53.41.15.PH.V	KY386615	*Calidifontibacter indicus* PC IW02				
60	R-68180	KY386316	*Deinococcus marmoris* DSM12784				
60	R-68561	KY386437	*Deinococcus marmoris* DSM12784	100	JNIV01000230		
60	R-68498	KY386507	*Deinococcus marmoris* DSM12784				
61	R-68571	KY386571	*Deinococcus saxicola* AA-1444	99.93	AJ585984		
61	KP53.17.15.PH.V	KY386581	*Deinococcus saxicola* AA-1444				
61	R-68514	KY386612	*Deinococcus saxicola* AA-1444				
62	R-68702	KY386340	*Hymenobacter roseosalivarius* AA718				
62	R-68471	KY386441	*Hymenobacter roseosalivarius* AA718	98.01	Y18833		
62	R-68240	KY386573	*Hymenobacter roseosalivarius* AA718				
62	R-68582	KY386591	*Hymenobacter roseosalivarius* AA718				

Phylotype	Strain	Accession no.	Nearest phylogenetic neighbor				
			ID	Sequence similarity (%)	Accession no.	*nifH*/*bchL*/*bchX*	*pufLM*
63	R-68402	KY386432	*Hymenobacter roseosalivarius* AA718	97.94	Y18833		
65	R-68403	KY386435	*Hymenobacter roseosalivarius* AA718	98.89	Y18833		
66	R-68243	KY386583	*Hymenobacter roseosalivarius* AA718	94.75	Y18833		
66	R-68122	KY386588	*Hymenobacter roseosalivarius* AA718				
67	R-67758	KY386449	*Hymenobacter arcticus* R2-4	98.05	KC213491		
68	R-68178	KY386311	*Hymenobacter terrae* DG7A	93.96	KF862488		
68	R-68353	KY386313	*Hymenobacter terrae* DG7A				
68	R-68361	KY386326	*Hymenobacter terrae* DG7A				+
68	KP15.29.4.PA.NS	KY386354	*Hymenobacter terrae* DG7A				
68	KP15.29b.4.PA.V	KY386355	*Hymenobacter terrae* DG7A				
68	R-68447	KY386356	*Hymenobacter terrae* DG7A				
68	R-68536	KY386363	*Hymenobacter terrae* DG7A				
68	R-68359	KY386369	*Hymenobacter terrae* DG7A				
69	R-68030	KY386555	*Hymenobacter terrae* DG7A	91.29	KF862488		
70	R-68225	KY386541	*Adhaeribacter aquaticus* DSM16391	96.37	AXBK01000007		
71	R-68113	KY386539	*Pedobacter panaciterrae* Gsoil 042				
71	R-68289	KY386572	*Pedobacter panaciterrae* Gsoil 042	97.59	AB245368		
72	R-67967	KY386410	*Pedobacter duraquae* WB2.1-25	98.36	AM491368		
73	R-68191	KY386353	*Pedobacter ruber* W1	95.43	HQ882803		
73	R-68207	KY386480	*Pedobacter ruber* W1				
74	R-68376	KY386334	*Spirosoma rigui* WPCB118				
74	R-68523	KY386371	*Spirosoma rigui* WPCB118				
74	R-68079	KY386376	*Spirosoma rigui* WPCB118	97.94	EF507900		
74	R-67957	KY386411	*Spirosoma rigui* WPCB118			+	
74	R-68424	KY386495	*Spirosoma rigui* WPCB118				
74	R-68502	KY386503	*Spirosoma rigui* WPCB118				
74	R-67885	KY386508	*Spirosoma rigui* WPCB118				
74	R-67936	KY386509	*Spirosoma rigui* WPCB118				
74	R-68108	KY386517	*Spirosoma rigui* WPCB118				
74	R-68505	KY386529	*Spirosoma rigui* WPCB118				
75	R-68384	KY386384	*Arthrobacter oxydans* DSM 20119	99.25	X83408		
77	R-68259	KY386618	*Angustibacter aerolatus* 7402J-48	96.64	JQ639056		
10	R-68085	KY386308	*Geodermatophilus terrae* PB261	99.02	JN033773		
14	R-68159	KY386450	*Conexibacter arvalis* KV-962	94.02	AB597950		
56	R-68670	KY386433	*Paenibacillus frigoriresistens* YIM 016	97.66	JQ314346		
64	R-68168	KY386300	*Nocardioides echinoideorum* CC-CZW004	78.7	KM085325		
76	R-68213	KY386500	*Hippea maritima* DSM 10411	79.07	CP002606		

Chapter 6

Abditibacterium utsteinense sp. nov. the first cultivated representative of the bacterial candidate phylum FBP

Redrafted from:

Tahon G., Tytgat B., Lebbe L., Carlier A. and Willems A. (2017). *Abditibacterium utsteinense* sp. nov. the first cultivated representative of the bacterial candidate phylum FBP. *In preparation*.

Author's contributions:

Conceived and designed the experiments: GT, AW. Performed the experiments: GT, LL. Analyzed the data: GT, BT, LL, AC. Wrote the paper: GT, AW. All authors approved the final manuscript.

Summary

Most bacterial lineages are known only by molecular sequence data from environmental surveys, representing the uncultivated majority. One of these lineages, candidate phylum FBP, is widespread in extreme environments on Earth, ranging from polar and desert ecosystems to wastewater and contaminated mine sites. Here we report on the characterization of strain R-68213^{T}, the first cultivated representative of the FBP lineage. The strain was isolated from a terrestrial surface sample from Utsteinen, Sør Rondane mountains, East Antarctica and is a Gram-negative, obligate aerobic, oligotrophic chemoheterotrophic bacterium. It displayed growth in a very narrow pH range, use of only a limited number of carbon sources, but also a metabolism optimized for survival in low-nutrient habitats. Remarkably, phenotypic and genome analysis indicated an extreme resistance against antibiotics and toxic compounds as well as the absence of two ribosomal proteins (L30 and L36). We propose the names *Abditibacterium utsteinense* for this bacterium and *Abditibacteria* for the former candidate phylum FBP.

6.1. Introduction

In the 1970s, Carl Woese pioneered the use of the 16S ribosomal RNA (rRNA) gene as a phylogenetic marker to study the taxonomy of Prokaryotes [394]. Building on his efforts, 16S rRNA gene sequence surveys steadily expanded the known microbial tree of life. The last decade, however, the use of high-throughput sequencing techniques has rapidly revolutionized our knowledge of the microbiological world. Anno 2017, the microbial tree of life has grown from Woese's initial 12 phyla to 30 cultured and estimates of ~250 candidate lineages (*i.e.* phylogenetic lineages without a cultured representative) in the domain Bacteria [395, 396]. Despite the importance of bacteria in most ecosystems, only ~10^5 of the estimated 10^{11} to 10^{12} species that inhabit our planet are currently represented in annotated genome sequence information [397].

Although the last couple of years cultivation has seen a revival, the issue of the Great Plate Count Anomaly still remains [398] and one of the challenges facing microbial ecologists is isolating representatives of candidate lineages. To date, representatives of only ~10^4 bacterial species have been cultured [397]. Given their potential importance in the environment, obtaining new cultured representatives of uncultivated taxa is of great interest. Isolates not only allow deciphering of their ecological significance and physiological properties, information in their genomes can also guide future isolation efforts, potentially leading to further novel discoveries [399].

The first 16S rRNA gene sequence of candidate division FBP was recovered from a lichen-dominated cryptoendolithic community in the McMurdo Dry Valleys, Antarctica [400]. Cultivation-independent studies have reported abundant occurrence of FBP in terrestrial and aquatic areas in the inland Sør Rondane Mountains, East Antarctica [55, 275]. Additionally, a low number of FBP sequences has been found in terrestrial, predominantly polar, desert and contaminated regions all over our planet, [401-416], Chinese stone cultural heritage monuments [417], magmatic rock cavities [418], wastewater [419, 420], hospitals [421, 422], oak trees [423], iron snow [424], aerosol [425], and human and animal skin [426-428].
We previously reported two representatives of the FBP division (strains R-68168 and R-68213^T) among isolates obtained from exposed sites in the proximity of the Belgian Princess Elisabeth Station, Sør Rondane Mountains, East Antarctica [Tahon et al. In preparation (Chapter 5)]. Unfortunately only the latter strain remained viable. In this study, we present a

detailed description of the phenotypic and phylogenetic features, and the complete genome sequence of strain R-68213^{T}, providing the first insights into the lifestyle and metabolic potential of a member of candidate lineage FBP. We also propose the novel genus and species *Abditibacterium utsteinense* for this strain and describe several novel taxa of higher ranks, including the novel phylum Abditibacteria for sequences previously allocated to the candidate phylum FBP.

6.2. Materials and Methods

Strain and isolation conditions

Strain R-68213^T was isolated from an exposed Antarctic soil sample (KP43, 71° 56' 47.3" S, 23° 20' 44.6" E, elevation 1362 m), taken in the vicinity of the Belgian Princess Elisabeth Station (71° 57' S, 23° 20' E, elevation 1382 m), Sør Rondane Mountains, East Antarctica [Tahon et al. In preparation (Chapter 5)]. The top surface soil was sampled aseptically during the austral summer of 2009. The soil sample had a pH of 6.22 (25 °C), a water content of 0.91 %, a total organic carbon content of 2.57 % and a conductivity of 520 µS cm^{-1}. Additional information and environmental parameters on the sampling location are given by Tahon et al. (2016) and Tytgat et al. (2016) [275, 277]. Isolation and cultivation of strain R-68213^T was performed at 18 °C using a low nutrient medium (PH) containing 3.50 mM $K_2HPO_4.3H_2O$, 1.47 mM KH_2PO_4, 0.81 mM $MgSO_4.7H_2O$, 3.42 mM NaCl, 0.58 mM $CaSO_4.2H_2O$, 25 µM $Fe_2(SO_4)_3$, 69.6 nM $ZnSO_4.7H_2O$, 0.252 µM $MnCl_2.4H_2O$, 25.2 nM $CoCl_2.6H_2O$, 10 nM $CuCl_2.2H_2O$ and 25 nM $NiCl_2.H_2O$, 24.32 µM MoO_3, 1 µM V_2O_5 0.25 mM glucose, 0.25 mM sucrose, 0.25 mM sodium succinate, 0.25 mM sodium pyruvate, 0.25 mM sodium acetate, 0.25 mM malate and 15 g Bacto agar (BD) l^{-1}. The final pH of the medium was set to 7.0. The strain was deposited in the Belgian Co-ordinated Collections of Microorganisms (BCCM/LMG, Ghent, Belgium under the accession number LMG 29911^T.

DNA extraction

Bacterial DNA was extracted using an automated Maxwell® DNA preparation instrument. Briefly, a half blue Ansa/Microloop 10 µl (Biosigma) loop of cells was suspended in 300 µl of UltraPure™ Guanidine Isothiocyanate Solution (4 M) and transferred to well 1 of the 7 well Maxwell® cartridge. Subsequently, DNA was extracted using the automated DNA extraction protocol. The final extract was treated with RNAse (2 mg ml^{-1}, 5 µl per 100 µl extract) and incubated at 37 °C for one hour. DNA integrity was checked using agarose gel electrophoresis. Quantification of DNA was performed using the Qubit® 2.0 fluorometer (LifeTechnologies). Afterwards, DNA was stored at -20 °C prior to further analysis.

16S rRNA gene and genome sequencing

Colony purity was confirmed using PCR amplification and sequencing of the 16S rRNA gene. Amplification was performed in two replicates using a Veriti thermal cycler (Life Technologies). The amplification reactions were carried out in 25 µl reaction mixtures containing 1 µl of template DNA, 1x Qiagen PCR buffer (Qiagen), 0.2 mM of each deoxynucleotide triphosphate, 0.625 units of Qiagen *Taq* polymerase (Qiagen). Primers pA and pH [280] were added in a final concentration of 0.1 µM. Amplification was carried out as follows: 95 °C (3min), 3 cycles of 95 °C (2 min 15 s), 55 °C (2 min 15 s), 72 °C (1 min 15 s) and 30 cycles of 95 °C (35 s), 55 °C (1 min 15 s), 72 °C (1 min 15 s). A final extension at 72 °C (7 min) and subsequent cooling at 4 °C finished the amplification reaction. PCR products were purified using a Nucleofast 96 PCR cleanup membrane system (Macherey-Nagel and Tecan Genesis Workstation 200 (Tecan). Sequencing of the 16S rRNA gene was performed using the amplification primers and primers BKL1, *Gamma, *O, Gamma, 3 and *3 [281, 359]. Sequencing was carried out using a BigDye Xterminator™ purification kit (Applied Biosystems) and an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems). The 16S rRNA gene sequence of strain R-68213^T has previously been submitted to the NCBI database under accession number KY386500 [Tahon et al. In preparation (Chapter 5)].

The genome sequence of strain R-68213^T was determined using the Illumina HiSeq2500 platform with 2x 125 bp cycles (BaseClear). De novo assembly of paired-end reads was performed according to Carlier et al. [429]. Final contigs were submitted for genome annotation using the Integrated Microbial Genomes-Expert Review (IMG-ER) platform [430] and the RAST server [431]. Identification and annotation of prophage sequences within the genome was performed using PHASTER [432]. Presence of clustered regularly interspaced short palindromic repeats (CRISPRs) was analyzed using CRISPRFinder [433]. Pairwise average nucleotide identity (ANI) was determined using the built-in tool in the IMG-ER platform [434]. The annotated complete genome sequence of strain R-68213^T (LMG 29911^T) will be deposited in the Genbank database.

Morphology

Gram staining was performed as previously described by MacFaddin (1980) [435]. The shape and size of cells was observed using a phase-contrast microscope (Olympus BX40) and by transmission electron microscopy. For electron microscopy visualization, bacteria were centrifuged at 5000 g for 5 min. Bacteria were subsequently fixed with 3 times diluted, followed by 2 times diluted Karnovsky fixative (4 % paraformaldehyde and 2,5 % glutaraldehyde in 0,1 M sodium cacodylate buffer pH 7.4) each time 20 min at 4 °C (continuously rotation of the samples). Finally the cells were fixed in pure Karnovsky for 1 hour at room temperature followed by overnight fixation at 4 °C. Between the fixation steps samples were centrifuged 5 min at 5000 g and supernatant was removed. Samples were rinsed 3 times for 20 min in 0,134 M sodium cacodylate buffer and after the last rinsing step bacteria were embedded in 1 % low melting agarose, Ultra-pure TM (Invitrogen). Post fixation took place in reduced Osmium (a mixture of 1 ml OsO_4 (4 %), 3 ml sodium cacodylate (0,134 M) and 66 mg $K_3Fe(CN)_6$), for 1 hour at room temperature. After rinsing 3 times 10 min with double distillated water, the samples were “en bloc” stained with 1 % Uranyl acetate in water for 1 hour at room temperature. After rinsing in double distilled water for 3 times for 10 min, samples were stepwise dehydrated, using ethanol series of increasing concentrations (15-100 % ethanol, 10 min each). The cells were subsequently infiltrated with a low-viscosity embedding medium [436], and polymerization was performed at 65 °C for 10 hours. Ultrathin sections were cut on a Leica UC7 ultramicrotome (Leica, Vienna, Austria), with a diamond knife (Diatome, Ltd., Biel, Switzerland) and mounted on formvar coated single slot copper grids (Agar Scientific, Stansed, UK). The sections were post-stained with Uranyl acetate and lead citrate before studying with a Jeol JEM-1010 transmission electron microscope (Jeol Ltd, Tokyo, Japan) operating at 60 kV. Pictures were digitized using a Ditabis system (Pforzheim, Germany).

Fatty acid analysis

The cellular fatty acid composition of strain R-68213^T was determined from cells grown on Reasoner's 2A (R2A) agar for 3 weeks at 18 °C. After fatty acid methyl-ester extraction and separation by gas-liquid chromatography was performed using the MIDI system (MICROBIAL ID Inc.) as previously described [437], fatty acid methyl esters were identified by comparison to the peak library version 5.0.

Pigment extraction and spectral analysis

Single colonies of strain R-68213^T were grown on R2A agar medium at 18 °C. Pigment extraction was performed according to Mohammedi et al. [438] and Henriques et al. [439]. Colony material was transferred to 1 ml of physiological water. Subsequently, bacteria were harvested by centrifugation at 1800 g for 5 min. The supernatant was removed and cells were resuspended in 2 ml of MilliQ and submitted again to centrifugation. This process was repeated twice. Afterwards, cells were resuspended in 1 ml of methanol (≥99.9 %, Sigma-Aldrich), strongly vortexed for 30 s and heated at 85 °C for 20 min. This mixture was centrifuged at 1800 g for 5 min after which the absorbance spectrum of the supernatant was measured between 190 and 1000 nm using a SPECTRAmax PLUS 384 spectrophotometer (Molecular Devices).

Metabolic profile

Unless stated otherwise, metabolic profiling of strain R-68213^T was conducted at 18 °C, in three replicates, using the original isolation medium, but with 5 mM of each of the six carbon sources (medium PH-5). The salt tolerance was tested by growing the strain on PH-5 medium containing 0, 0.02, 0.1, 0.5 and 1.0 % NaCl. Growth at different temperatures (1, 2, 4, 15, 18, 20, 25, 28, 37, 42, 45 and 52 °C) and pH values (5.5-8.5 at intervals of 0.5 pH units) was assessed after up to 12 weeks of incubation. Catalase activity was determined by bubble production in 3 % (v/v) H_2O_2 and oxidase activity was determined using 1 % (w/v) tetramethyl *p*-phenylenediamine. The composition of the medium used to study single carbon source assimilation was that of the PH medium, however, instead of a mix of six different carbon sources (i.e. glucose, sucrose, sodium succinate, sodium pyruvate, sodium acetate, malate), only one was added in a final concentration of 5 mM. To assess growth on common complex media, strain R-68213^T was grown on pure, one-tenth-strength and one-hundredth-strength marine agar, R2A and tripticase soy agar. To assess growth in liquid media, R-68213^T was grown in liquid PH-5 and R2A broth media. Growth in microaerobic atmosphere (80 % N_2, 15 % CO_2 and 5 % O_2) was tested at 18 °C on PH-5 medium as well as on R2A (three replicates each). Growth in the presence of nitrogen sources was assessed on PH-5 medium with 4 mM KNO_3 or/and 4 mM $(NH_4)_2SO4$. The composition of the medium to assess influence of amino acids on growth was that of the PH-5 medium enriched with 2 % casamino acids (Sigma-Aldrich).

Enzymatic profiles and use of sole carbon sources of R-68213^T were tested using API 20E and 20NE (api®), and the GEN III MicroPlate™ (Biolog). Antibiotic sensitivity was determined by growing R-68213^T on R2A and PH-5 medium (three replicates each) and using antimicrobial susceptibility paper discs. Inhibition zones were read after 8 weeks of incubation at 18 °C. Tested antibiotics (Oxoid) are listed in Table S1.

Phylogenetic analysis

A set of reference sequences covering the bacterial and archaeal domains of life was generated using publicly available data from the NCBI (https://www.ncbi.nlm.nih.gov/) and GreenGenes (http://greengenes.lbl.gov/cgi-bin/nph-index.cgi) databases [285-287]. From these databases, all available (near) complete 16S rRNA gene data (length >1200 bp) classified as the phyla FBP, Armatimonadetes and WS1 was sampled. For the other 81 bacterial and 4 archaeal phyla and candidate phyla that were included, data was sampled so that multiple representatives of each phylum were included. This dataset comprised 2463 sequences to which the 16S rRNA sequences of strains R-68168 and R-68213^T were added. Sequences were aligned using Clustal Omega [248, 249] on the Stevin supercomputer at UGent. A total of 100 iterations was performed to optimize the alignment. After manual curation, the final alignment, comprising 2701 nucleotide positions, was used to construct a phylogenetic maximum likelihood (ML) tree with 1000 bootstrap replicates in the FastTree software environment [250]. Parameters used included the general time-reversible model and the discrete gamma model with 20 rate categories. The tree was visualized using the iTOL software [251, 252].

For a concatenated ribosomal protein alignment, 15 ribosomal proteins (L2, L3,L4, L5, L6, L14, L16, L18, L22, L24, S3, S8, S10, S17 and S19) of strain R-68213^T were concatenated and aligned with the alignment previously constructed by Hug et al. [440] using Clustal Omega. The final alignment comprising 2596 amino acid positions from 3084 genomes was used to construct a ML tree (1000 bootstraps) using FastTree, under the Whelan and Goldman evolutionary model and the discrete gamma model with 20 rate categories.

For phylogenetic genome composition analysis, protein encoding genes (PEGs) with positive BLAST hits in distinct phyla (>30 % sequence similarity) were retrieved from the IMG database. Unassigned PEGs were not included in the analysis. Genomes used (IMG genome number is given between brackets) include strain R-68213^T (2700988623), *Fimbriimonas*

ginsengisoli Gsoil 348 (2585427636), *Chthonomonas calidirosea* T49, DSM 23976 (2524614646), Armatimonadetes bacterium JGI M3C4D3-002-C13 (2703719178), Armatimonadetes bacterium RBG_16_58_9 (2708742785), Armatimonadetes bacterium RBG_16_67_12 (2708742774), Armatimonadetes bacterium RBG_19FT_COMBO_69_19 (2708742821), *Deinococcus hopiensis* KR-140, DSM 18049 (2509276064), *Deinococcus deserti* VCD115 (643692020), *Chloroflexus aggregans* DSM 9485 (643348527), *Thermobaculum terrenum* YNP1, ATCC BAA-798 (646311962) and Chloroflexi bacterium RBG_16_47_49 (2708742765).

6.3. Results and Discussion

Source and isolation of R-68213^T

Strain R-68213^T was obtained from a top surface terrestrial sample taken on the East part of Utsteinen ridge (Sør Rondane Mountains, East Antarctica) during the Antarctic summer of 2009. The microbial community inhabiting the site is characterized by the near absence of Cyanobacteria and dominance of Acidobacteria and micro-algae [Tahon et al. Submitted (Chapter 3)]. The strain was obtained during an isolation campaign targeting aerobic phototrophic bacteria using low-nutrient media. The isolation plates were incubated under an increasing light regime (adding 2 h day^{-1} each week to mimic the onset of Antarctic summer) and after 10 weeks incubation at 18 °C a very small individual colony (~0.05 mm Ø) could be picked up [Tahon et al. In preparation (Chapter 5)]. Analysis of the 16S rRNA gene sequence with the EzTaxon database (http://www.ezbiocloud.net/identify) [60] indicated a low similarity with cultured bacteria, whereas comparison with environmental sequences from the GreenGenes database revealed a high similarity (up to 97.89 %) with sequences of the candidate lineage FBP [Tahon et al. In preparation (Chapter 5)].

General genome characteristics

The genome size of R-68213^T is 3.6 Mb with a G+C content of 54.27 % (Table S2). Based on annotation in IMG-ER, the genome contains 3186 PEGs (2243 with an assigned function). Using RAST, 3730 PEGs were annotated of which 51.39 % were hypothetical. Of the identified PEGs in RAST, many were based on assignments marked as unreliable. Therefore, the RAST annotation results were only used as a double-check of the IMG-ER results.

The genome contains one 5S (locus tag: Ga0139171_1302), 16S (Ga0139171_1451) and 23S (Ga0139171_144) rRNA gene. However, Sanger sequencing of the 16S rRNA gene (accession no. KY386500) revealed two degenerate positions. It should be noted that due to the conserved nature of the 16S rRNA gene, multiple copies are often assembled into one 16S rRNA gene, and these regions are known to have inflated coverage values. Using the formula (16S rRNA gene coverage)/(genome coverage) [395], strain R-68213^T is estimated to contain three copies of the 16S rRNA gene. A total of 46 transfer RNAs, representing 43 anticodons were encoded in the genome (Table S3). The implemented ANI calculation in IMG-ER found no genome with an ANI value ≥90. A total of five putative CRISPR regions were identified

(Table S4). These regions, thought to encode functions for the prevention of infection with alien DNA, could indicate frequent contact of R-68213^T with phages [441]. However, only one putative incomplete prophage region (7.4 kb length) was detected in the genome (Table S5).

A striking finding is the absence of two ribosomal proteins (i.e. L30 and L36). Generally, the genes encoding ribosomal proteins L30 and L36 are located in a specific place in the *spc* ribosomal protein gene operon, next to several other highly conserved ribosomal protein encoding genes [442]. The absence of L30 has previously been observed in several bacterial lineages and L30 therefore appears to be non-essential [443, 444]. Ribosomal protein L36, on the other hand, is highly conserved among Bacteria. Previous research with L36 mutants demonstrated the importance of this small protein, as its absence resulted in severe growth restrictions and a higher susceptibility to nuclease of nearby RNA regions [442, 443]. Therefore, the absence of ribosomal protein L36 may be a contributing factor to the very slow growth of R-68213^T.

Phylogenetic placement

The ML trees in Figures 1 and 2 confirm the placement of strain R-68213^T in the coherent cluster of 16S rRNA gene sequences originally defined as the candidate phylum FBP by Lee et al. (2013) [364]. We propose the new genus *Abditibacterium* (referring to a bacterium hiding in remote places) and species *Abditibacterium utsteinense* (referring to Utsteinen, the place of isolation) for strain R-68213^T. Placement of the FBP cluster next to candidate lineage WS1 and Armatimonadetes, with a high bootstrap support, was consistent with observations previously made by Lee et al. (2013) [364]. Recently, a 75 % 16S rRNA gene sequence similarity was suggested as threshold for delineating bacterial phyla [445]. Using this threshold, FBP clearly resolves as a separate phylum. Furthermore, the phylogenetic depth of the FBP cluster is similar to the divergence observed for other phyla. The separate FBP lineage was also apparent in a phylogenetic analysis based on a data set of 15 ribosomal proteins, known to more strongly resolve radiations compared to using 16S rRNA genes [440]. After inclusion of the R-68213^T ribosomal proteins, this strain clearly formed a separate lineage with a close relationship to Armatimonadetes (data not shown). Absence of enough ribosomal protein data from WS1 bacteria prevented inclusion of this candidate lineage in the analysis. Finally, a phylogenetic composition of the genomes of strains R-68213^T and

members of other closely related phyla (i.e. Armatimonadetes, Chloroflexi, Deinococcus-Thermus) showed that only 10.37 % of PEGs of R-68213^T had a positive BLAST hit (>30 % similarity) in genomes of Armatimonadetes bacteria (Figure 3). For Chloroflexi and Deinococcus-Thermus, this was 5.71 and 2.31 % of the PEGs, respectively. These numbers are in line with those obtained for other distinct phyla. For members grouping in the Armatimonadetes, Chloroflexi and Deinococcus-Thermus the percentage of PEGs with a positive BLAST hit in the same phylum is generally much higher (Figure 3). Similar results were previously obtained for the phylum Verrucomicrobia [446]. Based on all aforementioned results, it seems legitimate to establish a new phylum for the FBP lineage for which we propose the name Abditibacteria.

For Armatimonadetes, the percentage of PEGs with a positive BLAST hit in the same phylum were very different (Figure 3). It should be noted that the existence of 10 class-level divisions in this phylum was recently suggested [364]. This contrasts with the use of a 75 % 16S rRNA gene sequence identity threshold for defining a phylum. Using the latter definition, we propose the Armatimonadetes to be a superphylum. With the use of ~950 near full-length 16S rRNA gene sequences in our phylogenetic analyses, a total of 12 candidate phyla, all having the same phylogenetic depth as other phyla, can clearly be resolved within the Armatimonadetes (Figure 1). The subdivision is also corroborated by the phylogenetic composition of available Armatimonadetes genomes that showed a large variation in the number of PEGs with a positive BLAST hit in Armatimonadetes (Figure 3).

In the ML FBP tree, several distinct clusters can be observed (Figure 2). Many of these appear to group sequences with a similar habitat origin. The 16S rRNA gene sequences of strains R-68168 and R-68213^T display a close grouping with sequences retrieved from skin samples. R-68213^T groups in a small cluster with two other sequences retrieved from polar/alpine terrestrial habitats (Figure 2, Table S6). However, a close affiliation with sequences retrieved from skin samples can be observed. The second isolate, R-68168, which we have so far not been able to revive, groups in between sequences originating from human skin (Figure 2, Table S6). This close affiliation with skin microbiota may not be the complete picture as it should be noted that the phylogenetic tree shown here was constructed using near full-length 16S rRNA gene sequences, limiting the number of sequences that could be included. Indeed, using deep-sequencing, partial 16S rRNA gene sequences grouping with FBP have been retrieved from diverse aquatic and terrestrial samples from the Sør Rondane Mountains [55, 275]. Many of these sequences were more closely related to our isolates than FBP sequences from other environments (data not shown).

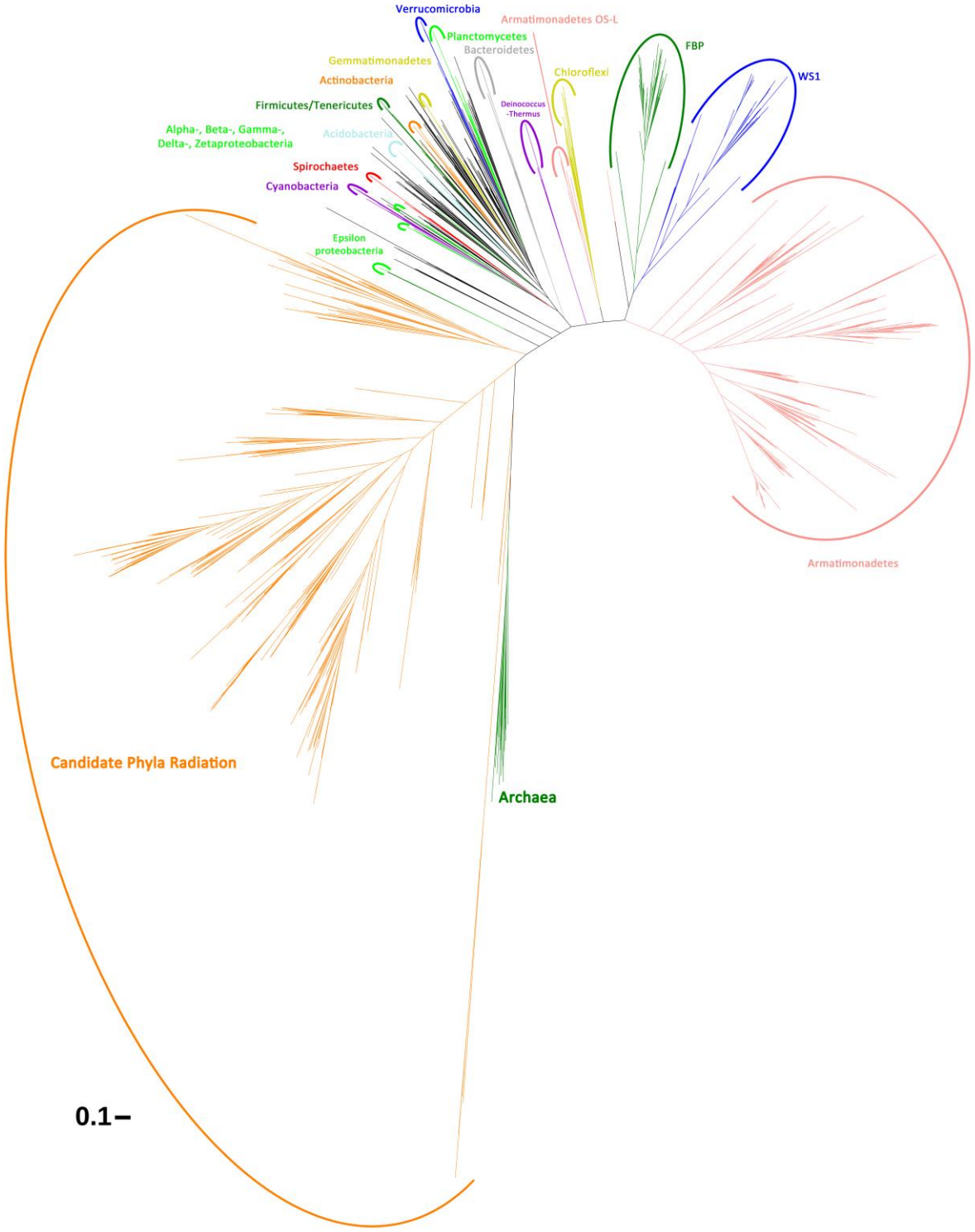

Figure 1 Prokaryotic tree of life (maximum likelihood, 1000 bootstraps), encompassing the total diversity represented by 16S rRNA gene sequences (length >1200 bp). The tree includes 84 bacterial and 4 archaeal phyla (Microgenomates and Parcubacteria are considered as phyla). Major lineages are colored and named. Candidate phyla radiation are given a single color. Scale bar indicates 0.1 substitutions per site. Details on tree construction are given in the Methods section

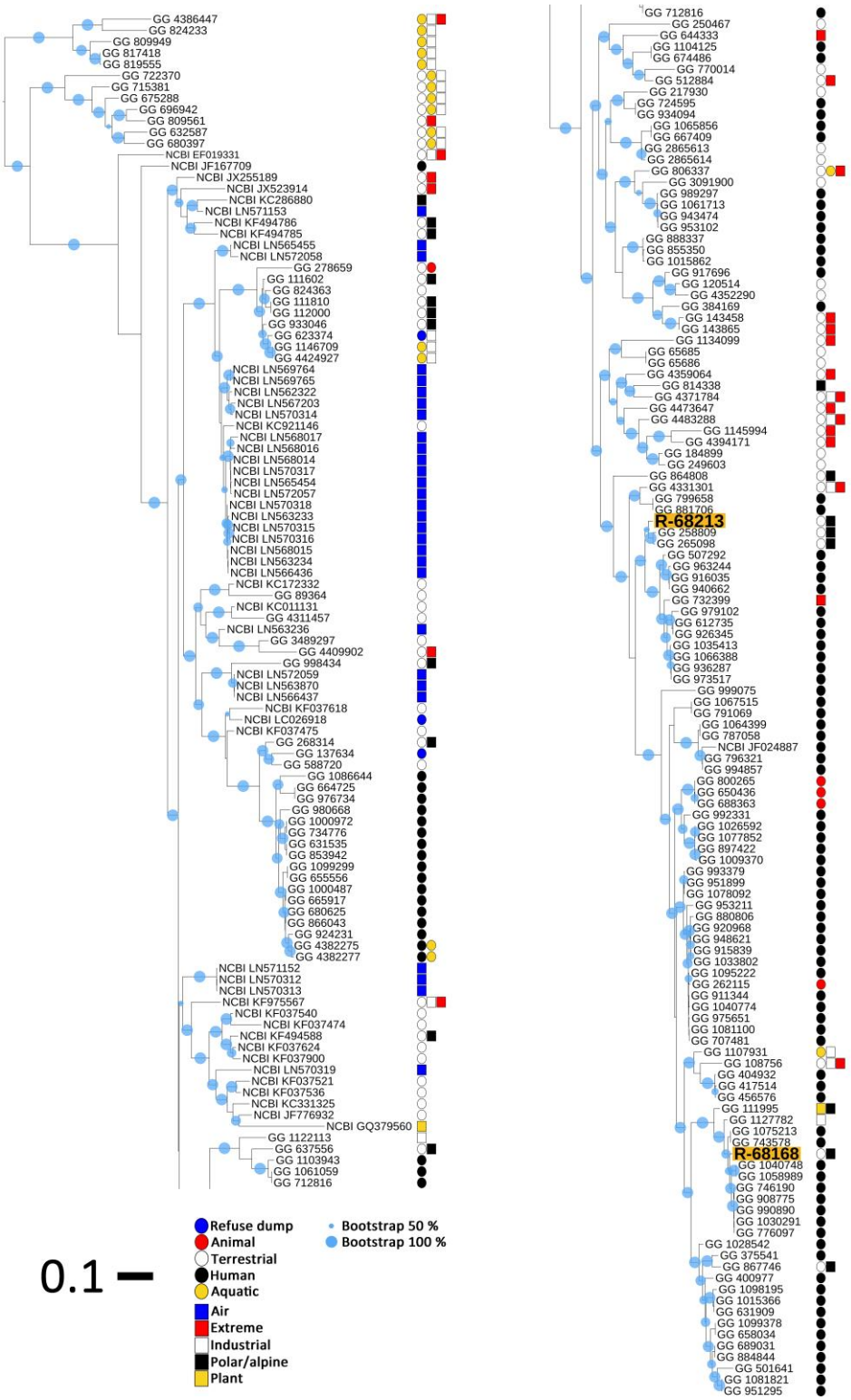

Figure 2 ML phylogenetic tree (1000 bootstrap replicates) of full length FBP 16S rRNA gene sequences. Scale bar indicates 0.1 substitutions per position. Positions of strains R-68168 and R-68213^{T} have been marked with an orange background. For reference data, accession numbers are listed. Accession numbers are preceded by the database from which the sequence was retrieved (NCBI = National Center for Biotechnology Information, GG = GreenGenes). Habitat origin(s) of sequences are represented by colored rectangles or circles Detailed information on sequences and their origin can be found in Table S6. Bootstrap values >50 % are displayed as circles with a diameter reflecting the height of the bootstrap value. Smallest circles represent the lower cutoff of 50 %

Morphology, physiology and chemotaxonomic characteristics

R-68213^T produced flat, smooth, circular colonies (0.25-1.3 mm in diameter) on agar-based media in approximately 6-12 weeks at 18 °C. Colonies were light-pink colored with a light-red center. Only extremely little growth could be observed in liquid media. Genome annotation results revealed several genes (e.g. phytoene synthase (Ga0139171_104206), phytoene desaturase (Ga0139171_104205)) used for production of lycopene and lycopene derivatives. These observations were corroborated by the purified pigment spectrum (Figure S1) which revealed a profile strongly resembling that of 3,4-didehydrolycopene [447]. Small differences between the profiles may be explained by the different solvents used. Strain R-68213^T grew between 1 and 45 °C, with an optimum at 18 °C, under aerobic conditions (Table S2). No growth appeared at 52 °C or under microaerophilic conditions. Growth below 1 °C could not be tested. Given the abundant presence of this phylum in continental Antarctica it can be hypothesized that growth at sub-zero temperatures may be possible [55, 275]. The strict aerobic lifestyle was corroborated by the identification of a complete F-type ATPase in the genome (Ga0139171_104209-104216). Strain R-68213^T tested positive for catalase and negative for oxidase (Table S2). The positive catalase result was corroborated by the presence of catalase (Ga0139171_12428) and manganese catalase (Ga0139171_11523) in the annotated genome. R-68213^T showed growth in a pH range pH 6.5-8.0, with an optimum between pH 7.0 and 7.5 (Table S2). No NaCl was required for growth and up to 0.5 % was tolerated on solid medium (Table S2). R-68213^T did not require $Fe_2(SO_4)_3$, $ZnSO_4.7H_2O$, $MnCl_2.4H_2O$, $CoCl_2.6H_2O$, $CuCl_2.2H_2O$, $NiCl_2.H_2O$ and MoO_3 for growth (Table S2). The main fatty acids observed are 16:1 ω7c/15 iso 2OH (34.3 %) and 18:1 ω 7c (15.6 %). A full overview of all fatty acids of strain R-68213^T, the majority being even-numbered ones, is given in Table S7.

Gram staining resulted in variable results, perhaps because of the age of the very slow growing cultures. However, transmission electron microscopy revealed a Gram-negative cell wall (Figure 4). This was confirmed by the presence of multiple outer membrane proteins in the annotated genome. Cells were approximately 0.6-1.0 µm in width, 0.65-1.3 µm in length and divide by binary fission (Table S2). The latter relies on the cell division protein FtsZ (Ga0139171_105174) [448]. Subsequently, FtsA (Ga0139171_109111) is one of the first proteins involved in the cell division process. Interestingly, R-68213^T also contains cell division protein YlmF (Ga0139171_10323) known to have an overlapping role with FtsA [448]. No spores or flagella were detected in ultrathin sections analyzed with an electron

microscope (Figure 4), and no genes associated to sporulation and flagellar motility were annotated in the genome.

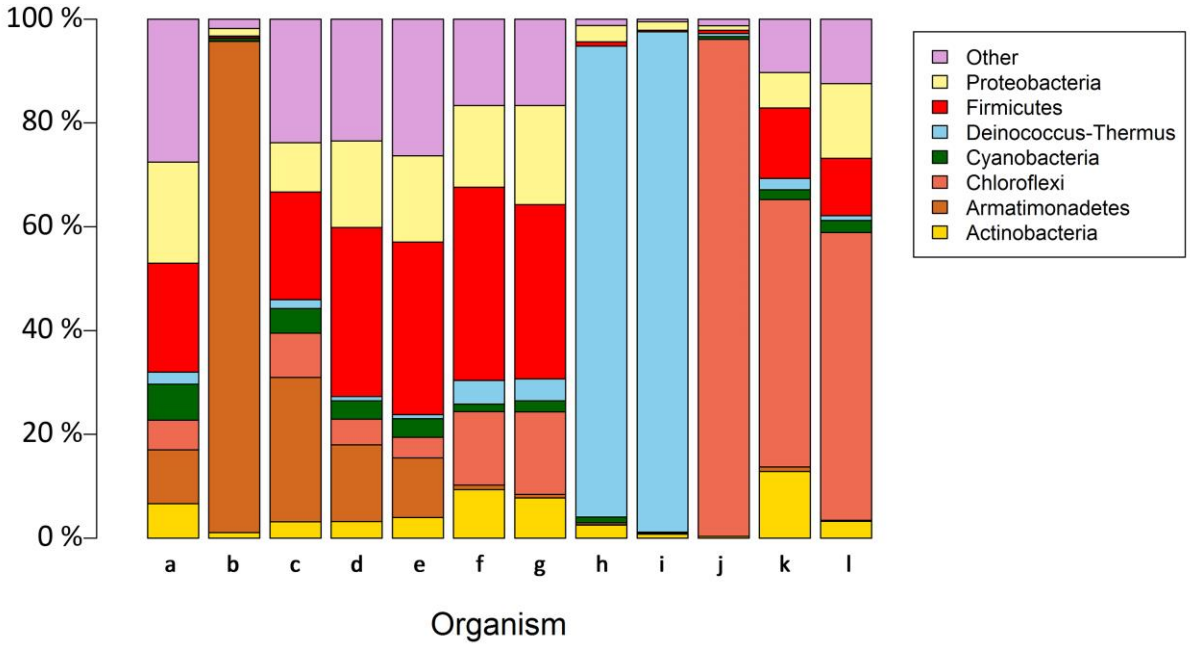

Figure 3 Phylogenetic composition of the genome of strain R-68213 (a) compared with representatives of the phyla Armatimonadetes (b-g), Deinococcus-Thermus (h-i) and Chloroflexi (j-l). Bar charts display the percentage of protein encoding genes with a positive BLAST hit (>30 % sequence identity) in genomes of distinct phyla. Unassigned proteins were excluded from the analysis. Only phyla for which at least one of the genomes had 5 % or more hits are shown. Other phyla are grouped together in the 'Other' group. The following genomes were used: Strain R-68213 (a), *Fimbriimonas ginsengisoli* Gsoil 348 (b), *Chthonomonas calidirosea* T49, DSM 23976 (c), Armatimonadetes bacterium JGI M3C4D3-002-C13 (d), Armatimonadetes bacterium RBG_16_58_9 (e), Armatimonadetes bacterium RBG_19FT_COMBO_69_19 (f), Armatimonadetes bacterium RBG_16_67_12 (g), *Deinococcus hopiensis* KR-140, DSM 18049 (h), *Deinococcus deserti* VCD115 (i), *Chloroflexus aggregans* DSM 9485 (j), *Thermobaculum terrenum* YNP1, ATCC BAA-798 (k) and Chloroflexi bacterium RBG_16_47_49 (l)

Core metabolism

The genome sequence of strain R-68213^T allowed the first reconstruction of the core metabolism of a member of the candidate lineage FBP. The central metabolism proceeds via the Embden-Meyerhof-Parnas pathway (glycolysis), the pentose phosphate pathway and the tricarboxylic acid cycle. Glycolysis is started by the processing of α- and β-D-glucose via the enzymes glucokinase (Ga0139171_10512, Ga0139171_11251) and glucose-6-phosphate isomerase (Ga0139171_10667). For the final step of the pathway, R-68213^T may depend on two enzymes: pyruvate kinase (Ga0139171_103229) and pyruvate phosphate dikinase (Ga0139171_1112). Whereas pyruvate kinase catalyzes the irreversible conversion of ADP and phosphoenolpyruvate to pyruvate and ATP, pyruvate phosphate dikinase catalyzes the

reversible reaction of AMP and phosphoenolpyruvate to ATP and pyruvate. The presence of these two similar enzymes may give R-68213^T an advantage over other organisms in the low-nutrient energy-limited Antarctic environment, as use of pyruvate phosphate dikinase results in an increase of two to five ATP per mol glucose during glycolysis [449]. Upon completion of glycolysis, pyruvate is converted into acetyl CoA using the pyruvate dehydrogenase complex, thereby linking the glycolysis pathway to the tricarboxylic acid cycle where Acetyl CoA is combined with succinic acid into Succinyl CoA via the 2-oxoglutarate dehydrogenase complex (Ga0139171_107103-107104, Ga0139171_107106, Ga0139171_107111) and isocitrate dehydrogenase (Ga0139171_10570). In addition to the Embden-Meyerhof-Parnas pathway, the pentose phosphate pathway can also supply NADPH and building blocks for other biosynthetic pathways. The genome of R-68213^T contained all key enzymes of the pentose phosphate pathway. Key enzymes of the Entner-Doudoroff pathway were absent. No complete photosynthesis and carbon fixation pathways were present in the genome.

In many Antarctic terrestrial sites, nitrogen concentrations are extremely low [220, 352] and nitrogen may thus be a limiting factor for growth. To overcome limited availability of nitrogen, the genome of R-68213^T revealed a versatile nitrogen metabolism potentially relying on the uptake of extracellular ammonium, urea, nitrate and nitrite using an ammonium transporter (Ga0139171_11942), an urea transporter (Ga0139171_102202-102206), a formate/nitrite transporter (Ga0139171_11331) and a nitrate/nitrite transporter (Ga0139171_10210). Subsequently, urea can be converted into ammonia using urease (Ga0139171_10974-10979), whereas nitrate is reduced into nitrite. Furthermore, as revealed by the presence of multiple key enzymes for assimilatory reduction of nitrate to ammonia, intracellular nitrite may be converted into ammonia. In turn, the annotated genome revealed R-68213^T's potential to assimilate ammonia into biomass, for example into L-glutamine and L-glutamate by using glutamine synthetase (Ga0139171_102200, Ga0139171_10671), glutamate synthase (Ga0139171_101177) and glutamate dehydrogenase (Ga0139171_10138).

A sulfate transport system (Ga0139171_10395-10398), sulfate permease (Ga0139171_10231), sulfate adenylyltransferase (Ga0139171_104129) and adenylylsulfate kinase (Ga0139171_1281) encode for uptake and activation of sulfate by assimilatory sulfate reduction. This pathway is, however, incomplete, as genes for the reduction of sulfite to sulfide are missing. No sulfatases were present in the genome. For phosphate, R-68213^T seems to depend on a phosphate transport system (Ga0139171_11289-11294). For uptake and resistance against other components (e.g. iron, molybdate, mercury), the genome of R-68213^T encodes several active transporters.

The substrate utilization spectrum recorded by laboratory testing of strain R-68213^T was quite narrow. In addition to five of the carbon sources present in the original isolation medium, only a few amino acids, monosaccharides, carboxylic alcohols and acids were used (Tables S2 and S7). Using higher concentrations of nutrients, biomass production could not be increased, indicating that R-68213^T has a metabolism adapted to oligotrophic growth conditions (Table S2). The R-68213^T genome contains several transporters for translocation of sugars (*e.g.* xylose, Ga0139171_10243-10245) into the cytoplasm. Furthermore, despite the narrow substrate spectrum, the genome encoded at least 57 glycoside hydrolases (GHs), distributed over 22 families, in particular GH43 and GH2 (Table S10). Using the CAZy database [450], a total of 38 GHs could be assigned a predicted activity. The presence of various β-glucosidases, cellulose esterases, and endo- and exoglucanases may reflect the adaptation of R-68213^T to the Sør Rondane Mountains' soil where the presence of mosses in several sites may be an essential source of cellulose for organisms such as R-68213^T [275, 451]. Furthermore, several green algae and bacteria, the main inhabitants of these areas, are known to produce cellulose [55, 275, 452-455]. Indeed, previous research indicated a positive correlation between the relative abundance of FBP and total organic carbon content in these soils [275]. Of the GHs with an assigned function, only a few were putative endoglucanases, indicating that R-68213^T may be capable of using components such as glucomannan and xylan for growth. The latter is known to be present in the cell wall of green algae [456], which have been retrieved from R-68213^T's isolation site using culture-dependent and -independent approaches [338][Chapters 3 and 5]. The absence of pectin lyases in the annotated genome may be linked to the absence of higher plants in continental Antarctica [24, 213]. Strain R-68213^T contains two α-L-rhamnosidases belonging to the GH78 family (Table S8). L-rhamnose is known to be a component of exopolymers that have an important role for bacterial biofilm establishment and maintenance [457, 458]. Therefore, we hypothesize that R-68213^T may potentially contribute to the degradation of microbial mats present in many terrestrial Antarctic sites [52].

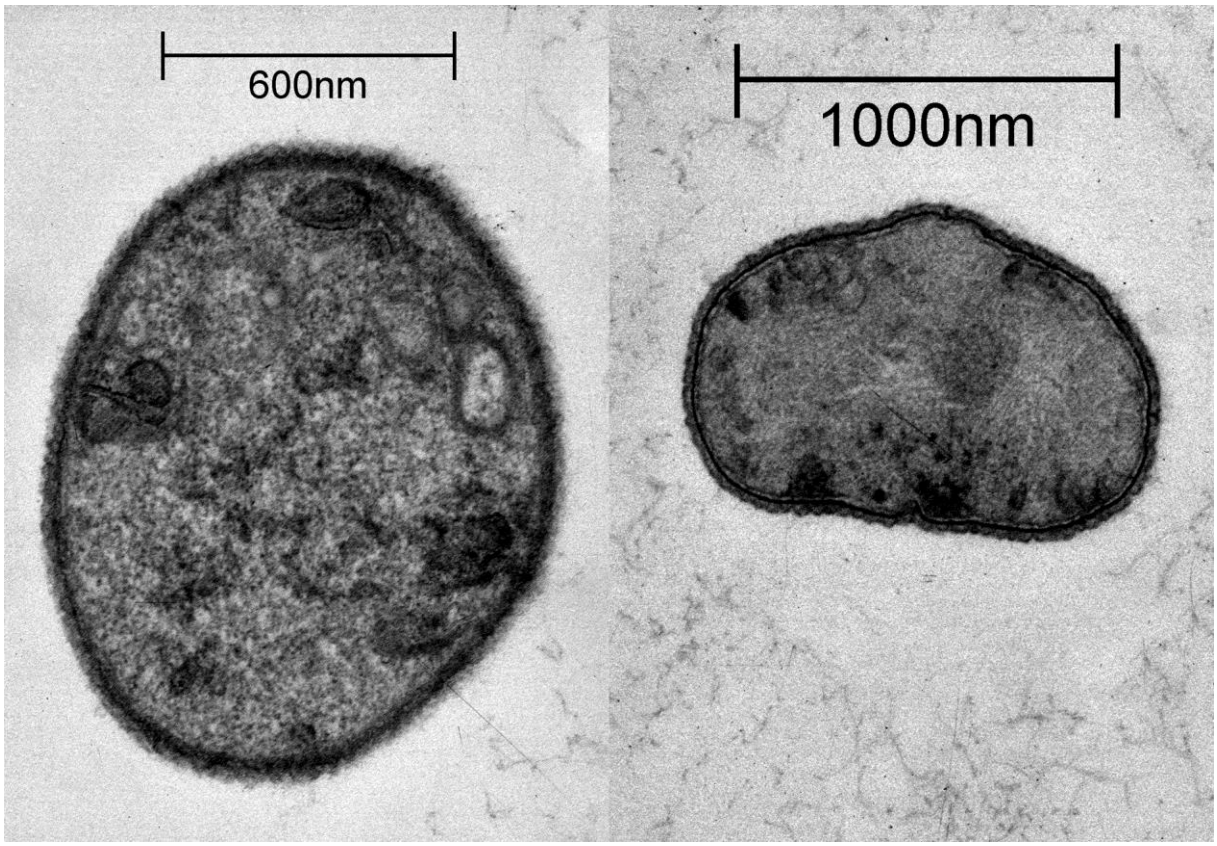

Figure 4 Transmission electron micrographs of strain R-68213^T

Ecological role and concluding remarks

The genome and physiological analysis of R-68213^T in combination with community and habitat analysis have provided insight not only in the ecology and ecological role of this first cultured strain, but also of the phylum Abditibacteria. The analyses all indicate that R-68213^T is an obligate aerobic chemoheterotrophic oligotrophic bacterium adapted to survival in the extreme Antarctic ecosystems. To overcome environmental stress caused by fluctuating temperatures, R-68213^T contains several cold- and heat-shock proteins (e.g. Ga0139171_1261, Ga0139171_101146). R-68213^T produces carotenoids, likely 3,4-didehydrolycopene, which may be regulated through σ-factors in response to oxidative stress caused by the hostile Antarctic environment [16]. Furthermore, the narrow pH range (Table S2) and substrate utilization spectrum of R-68213^T suggest a tight coupling between the microorganism and its environment. Growth in a narrow pH range may indicate limited capacity to buffer the cytoplasmic pH against external variations, making R-68213^T a bacterium adapted to a relatively stable environmental pH. Although many mechanisms for dealing with alkali and acid stress are absent in the genome (e.g. tryptophanase, agmatine deaminase, hydrogenases), the presence of acetyl CoA synthetase (Ga0139171_10840), and K^+/H^+ (Ga0139171_12426),

Ca^{2+}/H^{+} (Ga0139171_11735) and Na^{+}/H^{+} (Ga0139171_10490) antiporters may give R-68213^{T} the ability to deal with a limited amount of acid or alkali stress. Laboratory testing and the annotated genome revealed a high resistance against antibiotics (Tables S1). This may not only explain why this extremely slow growing bacterium is capable of competing with other life forms present in its habitat, but also the presence of this phylum in a variety of, predominantly extreme and/or contaminated, environments on Earth (Figure 2, Table S6). Nevertheless, members of this phylum have, until recently, always evaded isolation. This may be the result of a combination of factors. Firstly, extreme ecosystems have generally been less sampled. Secondly, the extremely slow growth and oligotrophic lifestyle may have prevented these organisms from being isolated, as many isolation campaigns have used of nutrient-rich media and short incubation periods. Thirdly, R-68213^{T} showed growth only under a limited number of conditions. Therefore, commonly used growth media and incubation conditions might not have been suitable for isolating these bacteria. Our study has given a first glimpse in the genomic content of the first cultured representative of the novel phylum Abditibacteria. This information gives the opportunity to set up novel targeted campaigns to isolate and characterize new members, increasing our understanding of this previously uncultivated phylum.

Classification

Description of *Abditibacterium* gen. nov.

Abditibacterium (ab.di.ti.bac.te'ri.um. L. adj. abditus, remote, secluded; -i- connecting vowel; L. neut. n. bacterium, a small rod; N. L. neut. n. *Abditibacterium*, a small rod hiding in remote places).

Strictly aerobic, obligate oligotrophic heterotrophic bacteria. No photosynthetic growth is observed. Cells are free-living, stain Gram-negative and divide by binary fission. Non-motile and non-sporulating. Oxidase negative and catalase positive. Major cellular fatty acids are 16:1 ω7c/15 iso 2OH , 18:1 ω7c and 16:0 3OH . The DNA G+C content is 54-55 %.

Type species is *Abitibacterium utsteinense*

Description of *Abitibacterium utsteinense* sp. nov.

Abditibacterium utsteinense (ut.stein.en'se. N.L. neut. adj. referring to Utsteinen, the location in Antarctica where the first strain was isolated).

The following properties are additional to those given for the genus. Cells are spherical to ovoid, 0.6-1.0 x 0.65-1.3 µm in size. Colonies are circular, smooth and pink with a red center on PH-5 medium. Growth occurs between 1 and 45 °C, with 15-18 °C as optimum. The pH range for growth is 6.5-8.0 with an optimum around 7.0-7.5. NaCl is not required for growth. The following substrates support stable growth on solid PH-5 medium: D-glucose, sucrose, sodium pyruvate, sodium succinate and sodium acetate. Susceptible to the antibiotics quinupristin/dalfopristin, linezolid, netilmicin, gentamicin, kanamycin, streptomycin, amikacin, clindamycin and minocycline (concentrations are given in Table S1). Antibiotics that are tolerated are listed in Table S1. Major cellular fatty acids are 16:1 ω7c/15 iso 2OH , 18:1 ω7c and 16:0 3OH. A full overview of all fatty acids can be found in Table S7. The DNA G+C content of the type strain is 54.27 %.

Type strain, LMG 29911^T (R-68213^T = XYZ xxxxx^T), was isolated from a terrestrial sample taken in the proximity of the Belgian Princess Elisabeth Station, Utsteinen, East Antarctica.

Description of *Abitibacteriaceae* fam. nov.

Abditibacteriaceae (Ab.di.ti.bac.te.ri.a'ce.ae. N. L. neut. n. *Abditibacterium*, type genus of the family; -*aceae*, ending to denote a family; N. L. fem. pl. n. *Abditibacteriaceae*, family of the genus *Abditibacterium*).

The family encompasses bacteria that divide by binary fission. Cells have a cell wall of the Gram-negative type. Affiliation of new species to this family should depend on their phylogenetic position based on the 16S rRNA gene. As recommended by Yarza et al. [445], the sequence similarity threshold for defining families is 86.5 %.
The type genus is *Abditibacterium*.

Description of *Abditibacteriales* ord. nov.

Abditibacteriales (Ab.di.ti.bac.te.ri.al'es. N. L. neut. n. *Abditibacterium,* type genus of the order; *-ales,* ending to denote a family; N. L. fem. pl. n. *Abditibacteriales*, order of the genus *Abditibacterium*).
The description is the same as that given for the family *Abditibacteriaceae*, with the exception that the 16S rRNA gene sequence similarity of newly described strains should be above 82.0 % [445]. The type genus is *Abditibacterium*.

Description of *Abditibacteria* classis. nov.

Abditibacteria (Ab.di.ti.bac.te.ri'a. N. L. fem. pl. n. *Abditibacterium,* type genus of the type order of the class; *-a,* ending to denote a class; N. L. neut. pl. n. *Abditibacteria*, the class of the order *Abditibacteriales*).
The description is the same as that for the family *Abditibacteriaceae*. For assigning a new species to this class, a 16S rRNA gene sequence identity above 78.5 % should be used as a threshold [445]. The type order is *Abditibacteriales*.

Description of *Abditibacteria* phyl. nov.

Abditibacteria (Ab.di.ti.bac.te.ri'a. N. L. fem. pl. n. *Abditibacteriales*, type order of the phylum; N. L. neut. pl. n. *Abditibacteria,* the phylum of the order *Abditibacteriales*).
The phylum *Abditibacteria* is defined based on the phylogenetic analysis of 16S rRNA gene sequences. Members of this phylum for a stable lineage separate from WS1 and Armatimonadetes. Type order is *Abditibacteriales*.

6.4. Acknowledgements

This work was supported by the Fund for Scientific Research – Flanders (project G.0146.12). Additional support was obtained from the Belgian Science Policy Office (project CCAMBIO). The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the Flemish Supercomputer Center (VSC) funded by Ghent University, the Hercules Foundation and the Flemish Government – department EWI. This study is a contribution to the State of the Antarctic Ecosystem (AntEco) research program of the Scientific Committee on Antarctic Research (SCAR). We thank Peter Vandamme for advice on the etymology of the new names proposed. We thank Myriam Claeys and Wim Bert for excellent technical assistance with electron microscopy visualization.

6.5. Supplementary information

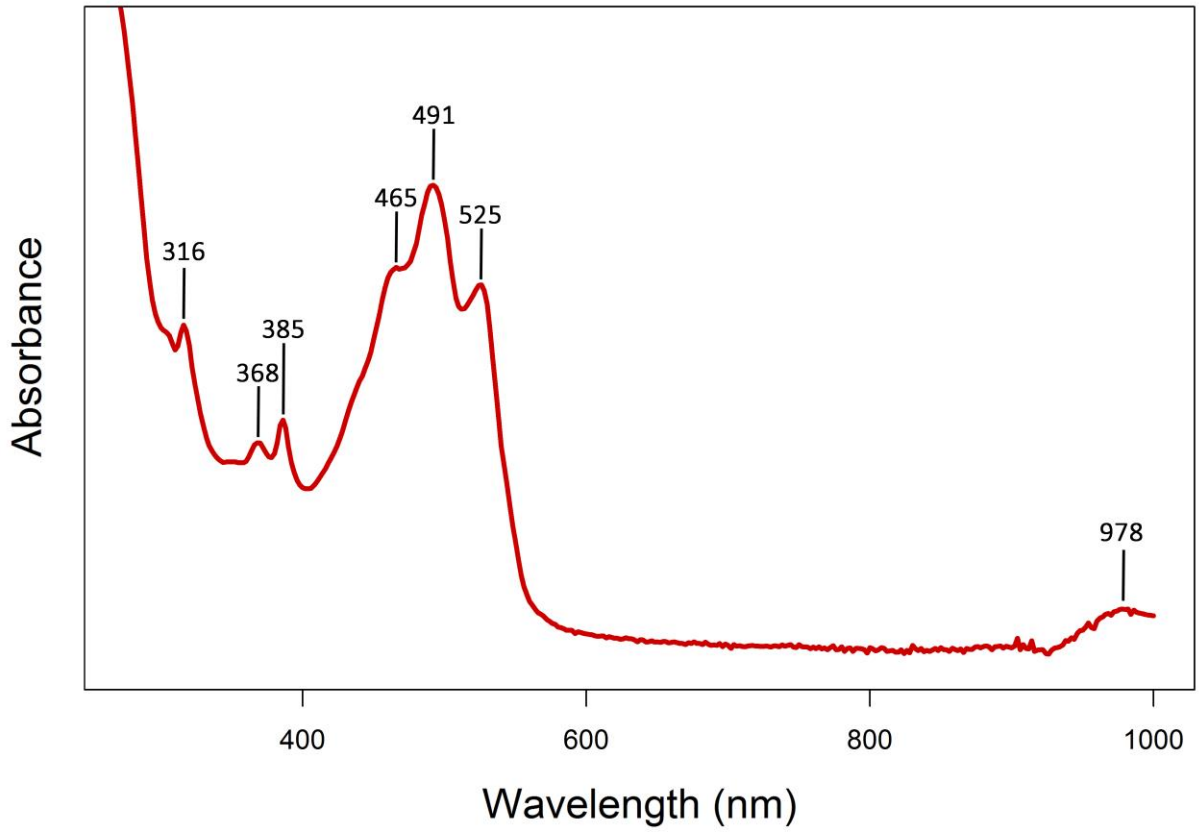

Figure S1 Absorption spectrum of extracted pigment (in methanol) of strain R-68213^T

Supplementary Tables are available at:
http://users.ugent.be/~gktahon/Chapter 6/

Chapter 7

General discussion and future perspectives

Exploring has always been part of man's nature. Since the beginning, mankind sought ways to conquer, investigate and gather knowledge on new places. There is no mountain high enough and no valley low enough to keep us from getting there. But even though our wanderlust takes us to more distant places every year, for example comet 67P/Churyumov-Gerasimenko [459], much is still to be learned about our own planet, especially since our current way of life greatly impacts, often negatively, our unique Earth.

One of the places on our planet that has received little research attention is the Antarctic continent. This is not surprising, given its remoteness, extreme environmental conditions and the costs associated to get there. However, the isolation and environmental conditions of Antarctica together with the relatively low human impact are what make this remote continent interesting. These have led to a unique and relatively untouched environment with an extremely adapted, yet relatively poor biodiversity dominated by microorganisms from all domains of Life [460]. Although the Antarctic ecosystems remain largely understudied, now is the time to investigate them. While most Antarctic habitats escape direct human impact, they are subject to the indirect effects of man's activities [461]. Global warming, as the name says, is omnipresent. Some areas in Antarctica are among the most rapidly warming regions on our planet [462] and research has shown that the delicate ecosystems in these regions appear to respond very quickly to climate changes [463, 464]. Furthermore, the growing human interest in Antarctica may cause great disturbances to the indigenous microscopic communities present. Sampling campaigns go hand in hand with contamination of the environment, not only by fuel spills [465, 466], but also by the unavoidable influx of allochthonous nutrient sources and microorganisms [467, 468].

In addition to documenting the indigenous Antarctic life forms before it is too late, or investigating the impact of humans and climate changes, the study of Antarctic ecosystems and their inhabitants is very interesting for many other reasons. The microbial diversity present is lower compared to more temperate systems [28, 29] and this simplicity could allow scientists, and in particular microbiologists, to reach a thorough understanding of the composition and functioning of these communities. In turn, this knowledge may be used to gain a better understanding of more complex systems. Secondly, Antarctic life has adapted to survive and function under the many extreme environmental conditions that characterize the continent. Yet, little is known about how this life has adapted, how it functions and which mechanisms are used to withstand the physical stressors. Therefore, gaining insight in the functioning of Antarctic microbial life may be interesting from a biotechnological point of

view, as it may lead to an optimization of certain industrial processes. Nevertheless, although Antarctica and its inhabitants potentially harbor many secrets the unraveling of which may prove very useful for several scientific and industrial domains, little is known about it.
Given the vast size and inaccessibility of the continent, studies have been limited to only a few regions, generally in the proximity of the research stations. These are mainly located in the Peninsula or near continental Antarctic coastal areas. As a result, inland regions remain under sampled or even uninvestigated. In addition, only a low number of studies have focused on Antarctic prokaryotes [14] and of these studies not many have investigated the functional or cultivable diversity. It is thus evident that the Antarctic continent harbors a pool of microbiological secrets, waiting to be discovered.

The aim of this PhD research was to assess the presence, diversity and distribution of key protein encoding genes involved in three of the oldest bacterial processes on Earth: autotrophy, diazotrophy and phototrophy, in order to increase our knowledge on functional bacterial diversity in soils of the Sør Rondane Mountains. Because previous analyses had shown that Cyanobacteria are sometimes only present in minute numbers in ice-free Antarctic sites, our focus was on non-cyanobacterial taxa. The approach was four-fold: (i) investigate and compare the diversity of these genes using old (clone library) and new (Illumina MiSeq) culture-independent techniques, (ii) use a high-throughput isolation campaign to grow and characterize auto-, photo- and diazotrophic bacteria from the same samples, (iii) compare the diversity retrieved using culture-dependent and –independent approaches and (iv) assess the potential contribution of the aforementioned mechanisms for growth of selected strains in laboratory experiments.

Our study is unique in that it combined a culture-dependent and culture-independent analysis of the functional bacterial community inhabiting the very scarce ice-free areas of Utsteinen (Antarctic Sør Rondane Mountains). Nevertheless, it is important to recognize that our experimental design did not allow analysis of biological replicates and a high number of environmental variables.
Our project involved the use of four terrestrial samples taken in the proximity of the Princess Elisabeth Station during the Antarctic summer of 2008-2009. Due to logistical restrictions, only a limited number of sites could be sampled and samples consisted of small amounts of material (5-40 g) that had been kept frozen (-20 °C) since collection in February 2009. To obtain a broader view of each site, biological replicates might therefore have been informative.

Additionally, freshly obtained samples would allow determinations on RNA and protein level. In a different study, focusing on the 16S rRNA gene composition of exposed samples from the Sør Rondane Mountains, several environmental parameters were obtained [275]. The small amounts of available sample only allowed determination of a restricted set of environmental variables. Nevertheless, temperature, conductivity and pH were denoted as important environmental drivers [275]. However, no data was available on nitrogen content. This could have been interesting for our study, as it might also have been an important driver impacting bacterial diversity.

Our project involved only a limited number of samples to be tested because initially the focus was mostly on clone libraries and isolations which involve much experimental work. After the first year we had the opportunity to add limited Illumina sequencing and for comparison, we worked with the same four samples. This did not allow determination of correlations between environmental parameters and the diversity results obtained from the different genes. Previously, several authors have reported on such correlations. For example, in different aquatic ecosystems *pufM* gene diversity is negatively correlated with salinity [234, 269]. For future studies it might thus be interesting to include more, and preferably fresh samples, with biological replicates, in order to determine whether correlations are present between functional diversity and environmental variables and how these relate to results from other environments.

7.1. Culture-independent analyses

As good scientific practice requires, we concentrated on a robust experimental design. Regarding microbiological diversity studies, one can argue about the most important step. In my opinion, every single step, no matter how small, from sampling to the final data presentation is equally important and should be performed the best way possible. Therefore, during our work, we spent a lot of time on optimizing our experimental setup.

7.1.1. The influence of DNA extraction methods on bacterial community composition

We first assessed the best DNA extraction method (Chapter 2). Hundreds of methods exist, all with their respective advantages and disadvantages. The use of a certain extraction protocol may greatly influence results (Chapter 2). For comparison with previous work, many authors choose to repeat previously used protocols. The constant evolution and improvement of tools and analyses, however, reduces the advantage of using the same methods as used before. A certain technique might have some time ago given an optimal result, but better approaches may have been developed over time.

Our study was the first to investigate the presence, diversity and distribution of genes involved in auto-, diazo- and phototrophy in the Sør Rondane Mountains. No prior studies had to be compared with. This gave us the liberty to assess the best extraction protocol (*i.e.* providing the most diverse DNA) and hence, optimize the outcome of the study. Other studies in the area, investigating the diversity of eukaryotic and bacteria SSU rRNA genes [54, 55, 275], used a combination of extraction protocols described by Corinaldesi et al. [292] and Zwart et al. [293]. These include removal of extracellular DNA and thus only provide a snapshot of the community of intact cells at the time of extraction. For samples residing in a dark freezer for several years, the living community may have altered from that at the time of sampling. Extracellular DNA may originate from once active, perhaps very abundant members of the *in situ* community. Removal of extracellular DNA might thus exclude these members. Our goal was to provide an overview of the (near) complete diversity of several genes at the sampling sites over a longer period of time. Therefore we chose to use an extraction protocol in which extracellular DNA was not removed.

7.1.2. PCR and primer dependency

The culture-independent part of the project involved amplicon-based sequencing. This has a major limitation: it depends on PCR. PCR involves the use of primers and a polymerase, and results in rather short sequences. Primers, although very often used in surveys, only have a restricted coverage. Even recently, metagenomic analyses revealed a novel lineage of bacteria that had remained hidden because of mismatches with the most commonly used 16S rRNA gene primers [366]. Updating primers may be possible, but the aim of designing universal primers is eutopian [469]. Furthermore, primers target a specific gene region which may greatly influence the amount of phylogenetic diversity obtained and may even result in the amplification of homologous regions from other genes (Chapters 2 and 3) [111, 470]. The primer synthesis also plays an important role. For degenerate primers it is very important to mention that no HPLC purification should be performed during synthesis. This will shift the relative abundances of the different primer versions in the mixture[8] and may bias the final PCR result [471, 472]. When performing amplicon-based research, it is therefore important to select a primer set that targets the broadest diversity, displays no non-specific binding, and amplifies a long enough fragment. If possible, multiple complementary primer sets should be used.

Primer limitations could have been avoided by using a metagenomics approach. Although this was not possible within the project budget, such an approach would also have allowed the analysis of key protein encoding genes involved in other pathways. It is important to recognize that we only addressed the, to our knowledge, most abundant and widespread autotrophic, diazotrophic and phototrophic processes. Metagenome data would have allowed analysis of genes encoding vanadium- and iron-dependent nitrogenases, or key genes in the five other known CO_2 fixation pathways serving autotrophic growth, or type 1 photochemical reaction centers and rhodopsins. During our study, all attempts to amplify proteo- and actinorhodopsin genes failed (Chapters 2, 4 and 5). This may have been the result of primer mismatch or a very low abundance in the community. A metagenome approach might have resolved this problem. However, working with metagenomes also has drawbacks. Firstly, the amount of data generated is enormous, complicating its informatic analysis [473]. Secondly, most communities are highly diverse resulting in an incomplete coverage of all genomes after sequencing. As a result, reads from the same gene may not overlap, making it impossible to

[8] Source: http://www.eurofinsgenomics.com/en/products/dnarna-synthesis/degenerate-bases.aspx

directly compare them through sequence alignment [474, 475]. In case different reads do overlap, it is not evident if they originate from the same or distinct genomes, which can challenge sequence assembly [476, 477]. Secondly, contamination poses a challenge to environmental surveys, because the identification and removal of sequence contaminants is problematic [473, 478, 479]. Thirdly, the gene(s) of interest only make up a very small portion of this data and filtering out the gene(s) of interest may prove (extremely) difficult [480, 481]. Generally, extraction of genes is based on comparing the metagenome reads with known sequence data and only withholding those that share a certain degree of similarity. This poses several limitations: (i) the known reference sequence data should contain as much diversity as possible. Absence of reference diversity may cause reads not being extracted from the metagenome. (ii) The reference data should be well-curated, as otherwise wrong data might be extracted [482]. For example, when extracting *nifH*, highly similar *nifH*-like genes encoding subunits of enzymes present in the (bacterio)chlorophyll synthesis pathway should be excluded. (iii) The use of a non-optimal cutoff (*i.e.* percentage of similarity) may lead to the recovery of false positives or false negatives. Since these cutoffs differ between genes and even between taxa, the optimal values should be determined prior to analysis, as they may greatly influence the diversity retrieved [481, 482]. However, similar to primer-based approaches, the main limitation during extraction of data from a metagenome is our current knowledge on a certain gene. Screening metagenomes for sequences sharing a certain degree of similarity or conserved amino acids with known diversity will fail to extract novel highly dissimilar types from new/unknown microorganisms that, for example, lack amino acid residues that are considered conserved among known diversity. Nevertheless, if those exist and given a high enough sequencing coverage, thorough analysis of a metagenome may reveal new diversity which, in turn, will lead to novel discoveries [483]. Another drawback of metagenome sequencing is the read length and high error rate. Illumina HiSeq platforms offer reads up to 250 bp. These 250 bp reads are, however, of a lower quality and highly subjected to errors in the last 100 bp[9]. As a result, shorter, but higher quality reads of 100 or 150 bp are far more commonly used for metagenome sequencing. Even pairing reads generates only short fragments which strongly limits and thwarts subsequent data analyses. Recreating longer fragments may resolve this problem, although this proves very difficult in practice, especially for complex communities, due to micro heterogeneity [342]. PacBio, Ion Torrent and Roche 454 offer longer read lengths, but are less commonly used due to their higher cost per base

[9] Source: http://www.illumina.com/documents/products/datasheets/datasheet_hiseq2500.pdf

pair and vulnerability to errors [484]. Although anno 2017 these analyses may not be this straightforward, deep sequencing technologies and data analysis tools improve rapidly. At the current rate, it is only a matter of years before metagenome analysis will be the golden standard for culture-independent research.

7.1.3. Advantages of 'outdated' techniques

High-throughput sequencing and meta-omics have revolutionized our knowledge about the bacterial world. The enormous amounts of data generated by these platforms have completely overshadowed the old culture-independent techniques. Despite this, one should not forget the advantages of these approaches. In a first study (Chapter 2), we used clone libraries to obtain a first view on the diversity of several protein encoding genes in our samples. Compared to the high-throughput amplicon sequencing platforms (*i.e.* Illumina MiSeq, Roche 454, Ion Torrent), this technique offers several major advantages. First, clone libraries permit the use of longer fragments. Where the most commonly used Illumina MiSeq sequencing platform is currently restricted to reads of 300 bp[10], and hence an assembly of maximum ~550 bp, clone libraries permit the use of much longer amplicons[11] [485]. Sequencing of clone inserts is generally done using Sanger sequencing. Compared to deep sequencing platforms, Sanger sequencing has a very low error rate, resulting in high-quality data [486, 487]. In turn, longer high-quality amplicons provide more detailed information about the gene of interest, thus improving phylogenetic analysis and leading to a more robust result. For long genes, these (near) complete sequences may also lead to an update or even the designing of internal primers which may be used in short amplicon-based techniques. Nevertheless, clone libraries also pose some issues. Over the last couple of years, the cost of high-throughput sequencing has greatly lowered. In combination with the large data output, nowadays, the price per sequence is a thousand fold lower than for a clone sequence. Additionally, obtaining a similar number of clones as there are sequences generated by deep sequencing would be extremely time-consuming and in fact be virtually impossible. In my opinion, to get a first view on the diversity of a certain gene or to study low-diverse communities, clone libraries are still very attractive. The results obtained from such an approach might help in deciding whether to continue with other analyses (*e.g.* high-throughput sequencing) or not.

[10] Source: http://www.illumina.com/systems/miseq.html
[11] Source: https://www.thermofisher.com/be/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/molecular-cloning/cloning/common-applications-strategies.html

7.1.4. The data analysis pipeline used

An integral aspect of our project was the use of Illumina MiSeq sequencing to investigate the diversity of key genes in carbon and nitrogen fixation, and light-harvesting (Chapters 3 and 4). Although the use of this sequencing platform has become the golden standard for environmental surveys, processing of such data requires numerous decisions. The use of certain programs, parameters and cutoffs will influence the final results. Throughout our analyses, a very stringent pipeline was used to guarantee data of a very high quality. The choice to use USEARCH [282] instead of Mothur [246] for processing reads was based on the previous work by Bjorn Tytgat [471]. His analyses of exposed samples from the same region (*i.e.* Sør Rondane Mountains) included a mock community. Prior to analyzing sequence data from the samples, a thorough investigation of the data of this mock community was performed. While the latter contained DNA from only 22 bacterial organisms, processing of the mock-community's sequence data with Mothur resulted in a tenfold increase of OTUs. The use of USEARCH and UPARSE [284] for sequence filtering and OTU clustering led to the recovery of 20-32 OTUs from this mock community, highlighting the more stringent performance of this program [471]. Our analyses of protein encoding genes did not allow sequencing of mock communities. This was partially compensated by additional quality control steps for protein-encoding genes that cannot be used when working with, for example, 16S rRNA genes. For future research, especially if more samples are sequenced, I would advise inclusion of a mock community for the different protein encoding genes. This will definitely benefit processing, as it may give a clearer view on the use of certain cutoff values for OTU clustering. During our analyses, sequences were clustered at 95 % similarity on protein level. While clustering sequences in OTUs is an integral part of processing amplicon data, it should be noted that there is no single correct or ideal cutoff value. To date, many different values have been used, hence complicating comparison between different studies. Furthermore, different binning tools exist. Subtle differences in the algorithms used will influence the clustering result and thus also the richness retrieved [488, 489].

Compared to pipelines used in other surveys, our data processing pipeline had several advantages. Before merging, we have opted to remove amplification primer sequences from the reads. Many authors do not remove primer sequences or remove them only after merging and a first quality filtering step. As shown by Schirmer et al. [490, 491], the first 10 and last 25 bp of MiSeq reads are highly subjected to sequencing errors. During merging, the presence of these areas may lead to removal of the sequence due to a high number of mismatches in the

overlapping region. Later on, errors may lead to an inflation of the number of OTUs recovered. Removing these areas prior to merging will thus clearly benefit the quality.

A second advantage was the limited amplicon length. Although providing somewhat less phylogenetic resolution, this resulted in a longer and sometimes even complete overlap between forward and reverse reads. A big overlap in combination with a minimum of mismatches allowed, will greatly improve sequence quality. The biggest strongpoints of our pipeline were, however, the additional quality control steps to exclude non-functional genes. For 16S rRNA gene data, it is virtually impossible to detect indels. When working with protein encoding sequences, translation to amino acids and subsequent removal of sequences containing stop codons, a frameshift resulting in loss of function or absence of multiple required highly conserved amino acids will further improve the data quality. Translation to amino acids may also allow removal of contaminants. During sequence analysis, I sometimes came across sequences from other genes in the datasets (*e.g.* presence of 16S rRNA in the *cbbL* dataset). These were probably the result of cross-contamination during the library prep. Although the latter was very carefully performed to avoid cross-contamination, it is impossible to completely rule out transfer of microdroplets. Combining different genes in one sequencing run will allow removal of a certain degree of cross-contamination. In my opinion, when working with different protein encoding genes, the aforementioned steps should always be performed.

Removal of sequences lacking necessary conserved amino acids is a step that is open for improvement and discussion. We chose to remove all sequences that lacked at least one or two conserved positions (depending on the number of conserved sites present in the amplicon). This is a very stringent approach that will probably have led to removal of existing and currently unknown new diversity, as it is well known that such sites are never 100 % conserved across a gene family [162].

7.1.5. Deep sequencing data is not used to its full potential

High-throughput amplicon sequencing data is hardly ever used to its full potential. Although these techniques offer a detailed view of the community structure, they do have their limitations. The main restriction is the (currently) limited read length and relatively high occurrence of sequencing errors. Short sequences do not allow identification at more detailed phylogenetic levels. Therefore, many studies only describe communities on phylum or class level. This, however, hardly gives any information about the potential function of a

community's members. Secondly, the presence of sequencing errors leads, in many cases, to the use of certain analysis tools or processing steps. One such controversial step is the removal of low abundant sequences/OTUs (*i.e.* singletons, doubletons etc.). Many authors choose to remove these data prior to further analysis, as they claim these sequences are merely the result of errors [492-494]. Previous research has already shown that this is not the case. In many cases, provided solid data quality control was performed, these low abundant sequences represent rare and/or unique lineages of dormant or inactive bacteria that may grow when the right conditions are met [54, 315, 316, 495]. Furthermore, their low abundance may be the result of PCR bias or the restricted lysis of these cells during DNA preparation. They may thus even represent important *in situ* taxa [496-498]. In my opinion, low abundant sequences should not be removed, that is, if the quality control pipeline is stringent enough.

Similarly, the use of denoising programs (*e.g.* UNOISE 2 [492], IPED [499]) to reduce the error rates, or chimera removal is cause for discussion. Some of these algorithms require the removal of low-abundant sequences prior to analysis, because they are more prone to contain errors [492]. Others "correct" so-called errors or remove low-abundant sequences in an OTU because they only show minimal differences with the most abundant sequence enclosed in this OTU [499]. Regarding chimera filtering, it is well known that these tools have mediocre performance. Schmidt et al. showed that, even on using a curated set of high-quality full-length 16S rRNA gene sequences, the most commonly used chimera filtering tools always filter out data [500]. Our own analyses also highlighted the poor performance of these tools (data not shown). Upon chimera filtering with the Uchime model [247] in Mothur [246], multiple so-called chimeric sequences were found to be identical to other (non-chimeric) sequences from our datasets. In many cases, the latter were retrieved in high abundances. Therefore, in my opinion, implementing denoising or chimera-removal steps may greatly influence results and their use should depend on the hypothesis of the study and the stringency of the other tools used. For example, if the goal is to investigate all diversity, I would avoid denoising as it might lead to the removal of low abundant, but perhaps very important diversity. When performed, chimera filtering should always be followed by manual curation to rule out false positives.

7.2. Cultivation-dependent analyses

7.2.1. The importance of cultivation

The last decade, deep sequencing meta-omics have revolutionized our knowledge about the bacterial world. We now better than ever before understand the functioning and composition of bacterial communities in a variety of environments all over our planet [55, 191, 275, 501, 502]. Although culture-independent approaches moved cultivation strategies to the background, the last few years, microbiologists have become more and more aware of the limitations of an exclusive focus on sequencing [342]. Amplicon sequencing, for example, only gives us an insight in who is present (16S rRNA genes) or in the diversity of certain protein encoding genes. For the latter, linking data to organisms is hindered by possible gene duplication and horizontal gene transfer events [263, 339, 340]. Metagenomic approaches, on the other hand, may reveal functional potential. However, retrieving information from such data sets may be hard since a considerable portion of genes cannot be assigned a function during annotation. For simple communities, it may be possible to recreate genomes or genome fragments from metagenome data. Nevertheless, this approach is biased by genomic micro-heterogeneity [341, 342]. Sequencing may provide us with an overload of data on bacterial community structure and potential function, however, certain functions and behavior (*e.g.* pH optimum or growth kinetics) can only be retrieved from cultured bacteria. Further downstream, these isolates might even become important in certain biotechnological approaches.

The last couple of years, culturing has started to live a second youth [397, 503]. However, the issue of the Great Plate Count Anomaly still remains. The laborious nature and complexity of bringing bacteria in to culture has been known for decades. Current estimates say that approximately 10^{11} to 10^{12} microbial species inhabit our planet. Of these, only ~10^4 have actually been cultured [397, 503]. This number is, however, rapidly growing as a result of culturomics studies (*i.e.* culturing using hundreds of culture conditions) [503, 504]. The unquestionably powerful metagenomic tools have provided sequence information of ~10^5 species [397]. Given the advantages of both approaches, characterizing microbial life on our planet will require a combination of classic microbiological techniques and culture-independent sequencing. This joint approach might even allow microbiologists to say farewell to the long-standing Great Plate Count anomaly, as it might turn uncultivable bacteria into cultivable ones ready to be studied in the lab [397].

Motivated by the large diversity of key protein encoding genes retrieved from our samples, an isolation campaign was set up to examine the culturable diversity of aerobic anoxygenic phototrophic bacteria dependent on the widespread type 2 reaction center (Chapter 5). The aim of this culture-dependent part was three-fold: (i) isolate aerobic phototrophic bacteria from our samples, (ii) test the influence of light on growth in laboratory experiments and (iii) compare this data with identifications obtained using culture-independent techniques.
It is important to note that we also tested the non-cyanobacterial representative isolates for presence of *cbbL*, *cbbM* (Calvin cycle) and *nifH*. The absence of these genes does not necessarily mean that our isolates are incapable of fixing carbon dioxide and/or nitrogen. Primer mismatch or the presence of a different autotrophic or diazotrophic pathway may have caused the negative results.

7.2.2. Technical aspects

The isolation approach used in our study was, however, limited. Only a small number of isolates (20 to 45) was randomly picked up per sample and condition. The main reason for this is the laborious culturing work inherent to an isolation campaign. The use of automated systems for the isolation of bacteria may eliminate this bottleneck, although many improvements will have to be made to the currently available systems. A colony picker robot present at the LM-UGent lab was initially to be used for isolating the bacteria. This very soon turned out to be a dead end due to technical shortcomings. Many of our isolates represented minute nearly invisible colonies on plate and picking these up by using the automated system was impossible. Similar to other available colony picker robots, the picking method (K2 automated colony picker system, KBiosystems) depends on locating colonies using a camera. This posed two major drawbacks: (i) the camera failed to localize micro-colonies and (ii) picking of micro-colonies was not possible.

Manual picking over a one-year incubation period eventually resulted in ~1600 isolates and a high number of cyanobacterial liquid mixtures. High-throughput MALDI-TOF MS dereplication of our isolates immediately posed a second bottleneck, as the little available biomass in combination with slow growth did not allow the use of standard protocols. To overcome this problem, we devised a simple miniaturized protocol that worked for many strains (67 %). Future advances will have to be made, however, to allow working with slow-growing low biomass producing strains.

The identity and phototrophic potential of the reduced set of 330 representatives was confirmed through amplification of 16S rRNA, and rhodopsin and *puf(L)M* genes, respectively. Of these representatives, 40 % grouped with known (potential) AAP genera. In combination with the MALDI-TOF MS dataset, this clearly showed that the majority of our isolates, grouped with known AAP taxa. Most of these genera had previously been retrieved from the PufM Illumina MiSeq dataset.

Only 38 % of the potential AAP isolates tested were positive for key genes in a phototrophic pathway. However, only a limited number of all isolates was tested. Therefore, this number may change when investigating other potential AAP. Furthermore, primer mismatch or the occurrence of different pathways/systems may have left the phototrophic potential of some strains undetected. In this respect, several isolates were also subjected to laboratory testing in which the influence of light on growth was determined by following the increase in CO_2 produced. Technical restrictions and the slow growth and low biomass production of our isolates again hampered the experiment. Nevertheless, several of the tested strains indeed revealed preferential growth under continuous illumination or some in darkness. Among the strains displaying better growth in light, one even lacked detectable phototrophic genes, hence clearly highlighting the potential primer mismatch or presence of other gene types. Interestingly, several strains testing positive for key phototrophy genes grew better in complete darkness. It seems logic that this should be the other way around. Perhaps, similar to methicillin-resistant *Staphylococcus aureus* that produces an enzyme that degrades skin secretions into compounds that are toxic to itself [505], our strains produce a toxin in light which inhibits further growth in our setup. *In situ*, these strains may rely on a cooperation with other bacteria that remove excess of the toxin, hence allowing the phototrophs to fully establish an advantageous growth in light. Additionally, the light intensity or the light regime used may eplain the absence of better growth in light. Previously it has been shown that bacteria display optimal growth under different light intensities and that incubation under a day/night cycle sometimes also benefits growth and survival [151, 506]. Despite the positive outcome of the growth experiment, it would be interesting to follow the growth over even longer periods (*i.e.* >10 weeks), because many of our strains did not yet reach stationary phase at the end of the experiment. In addition, following the number of cells using flowcytometry, and the expression of genes would also have benefited this experiment.

I do consider the isolation campaign a success. Our strategy of using oligotrophic media, prolonged incubation periods and a light regime mimicking the *in situ* onset of Antarctic summer led to the isolation of a large number of potential AAP isolates, some representing novel species and genera. Our approach mainly detected potential phototrophs grouping among known AAP taxa. However, a functional *pufLM* was also found in one of the *Hymenobacter* strains (Bacteroidetes). This phylum is not known to contain phototrophic members [147] and additional analyses of this strain may broaden our knowledge about the diversity and evolution of phototrophs. The liquid enrichment and isolation strategy resulted in a large number of cyanobacterial/green algal mixtures. Although we did not investigate these in our lab, they are being examined by our colleagues at the University of Liège and have already led to the addition of multiple new strains to the BCCM/ ULC polar cyanobacterial culture collection.

Next to AAP, a large number of other taxa were also obtained. The use of low nutrient concentrations has become a widely adopted strategy to increase cultivability of rare microbiota [507-510]. In our isolation campaign, this led to the cultivation of bacteria representing novel species, genera and families (Chapter 5). Additionally, we succeeded in isolating the first cultured representatives of a phylum only known to exist from environmental sequences (Chapter 6). The genomic landscape, physiology and chemotaxonomic characteritics of one of these strains (R-68213^{T}, *Abditibacterium utsteinense*) provided more insight in its ecological role, indicating this microorganism is an obligate oligotrophic chemoheterotrophic bacterium adapted to survival in a range of hostile environments. Taken together, the detailed analysis of this strain not only explains why it had always evaded isolation, but also provides important baseline information for future targeted isolation campaigns.

A more high-tech approach would have benefited our isolation campaign. Technical restrictions allowed the testing of a limited number of isolation conditions. Plating on solid media containing different solidifying agents might successfully target other groups. Indeed, it has been shown that the use of gellan gum better supports growth of certain taxa, compared to agar [508]. Aerobic anoxygenic phototrophic bacteria perform phototrophy under aerobic conditions. Our strategy included growth under normal atmospheric oxygen levels. For some AAP, these concentrations result in a lowered functionality of the phototrophic apparatus [391]. Therefore, incubation under different oxygen concentrations might lead to the growth of different phototrophs. Furthermore, phototrophs are known to be able to use an extremely

broad range of light wavelengths [159, 160, 166-168, 179]. The targeted use of light sources exhibiting only certain wavelengths may further improve isolation of phototrophs. Therefore, revisiting these cultivation experiments with completely new terrestrial samples should result in even more pronounced positive results.

7.3. Reconstruction of the *in situ* food web

The results obtained provide new insights in which bacteria might be present and which bacteria could be able to perform a certain function. This new knowledge allows a partial reconstruction of the *in situ* food web. However, because of limitations specific to our approach and results, including i) possible influence of horizontal gene transfer, ii) the use of only four samples, iii) absence of *in situ* activity values, iv) absence of data on other pathways and v) no availability of physicochemical data or detailed data from viruses, Archaea and Eukaryotes, the results only allow a tentative reconstruction and speculation about the potential role of some of the (bacterial) groups found in the investigated sites.

Similar to many other environments on our planet, there would be no life in Antarctica without energy from the sun. Given the absence of plants and high amounts of nutrients in the investigated sites, producers relying on phototrophy, and in particular on photosynthesis, will form the first trophic level and hence act as primary producers. In this respect, our results suggest that Cyanobacteria and green algae are the most important, depending on the location. For green algae, two of our datasets suggested presence and high importance of members of the Trebouxiophyceae (Chapters 3 and 4). Although no 18S rRNA gene sequence data is available for the sites investigated here, their presence is corroborated by sequencing results of nearby exposed sites. The work of Obbels et al. (2016) showed that the Eukaryotic community at several sites on Utsteinen ridge, with visible presence of crusts, mosses and algae, were dominated by members of Chlorophyta and in particular Trebouxiophyceae [55]. On Utsteinen nunatak, however, similar soil samples revealed a much lower relative abundance of these groups. Interestingly, Obbels et al. (2016) showed that many sites in the Sør Rondane Mountains with lower numbers of Chlorophyta displayed higher relative abundance of Cyanobacteria [55]. Using results of different datasets obtained in our work, a similar conclusion could be made about the main primary producers inhabiting the investigated areas. Nevertheless, our analyses also revealed the presence and diversity of many other primary producer bacteria relying on the Calvin-Benson-Bassham cycle (Chapters 2 and 3). Although mostly found in lower numbers than oxygenic phototrophs, they should not be overlooked, as it has previously been shown that low abundant organisms can be important contributors [496]. Of special interest could be members of *Bradyrhizobium*. As observed from our results, 16S rRNA, *cbbL*, *pufM*, *bchL* and *bchX* gene sequences highly similar to those of *Bradyrhizobium* were recovered in higher relative abundance from samples

with high numbers of Cyanobacteria and low numbers of green algae (Chapters 2-4). Many of our sequences displayed high similarities to those of *Bradyrhizobium* sp. 23321. This bacterium was demonstrated to be a non-nodulating strict oligotroph adapted to survival in a wide range of environments [511]. Given the low levels of nutrients and absence of vascular plants, the presence of such an organism in Utsteinen is not unlikely and therefore it could play an important role in primary production in sites with low numbers of green algae. Additionally, members of the Actinobacteria could act as important carbon dioxide fixers in our samples as our results showed a dominance and broad diversity of actinobacterial type IC (or IE) RuBisCO. Actinobacteria are an ecologically significant group which play a vital role in several biological processes. In many known autotrophic Actinobacteria, CO_2 fixation is linked to other important processes. Under environmental stress some link autotrophy to ammonia oxidation, whereas others rely on H_2 or sulfur as an energy source for carbon fixation [79]. Other Actinobacteria (*e.g.* some *Streptomyces* spp.) are known to produce the carbon monoxide dehydrogenase enzyme. In these bacteria, this enzyme facilitates microbial growth in nutrient deprived conditions by oxidizing carbon monoxide into CO_2, which is further fixed into microbial biomass through the Calvin-Benson-Bassham cycle [512]. Therefore, the versatility and high relative abundance (16S rRNA and *cbbL* genes) of Actinobacteria could make them the most important primary producers in exposed areas in Utsteinen (Chapters 2 and 3).

Anno 2017, five other pathways that serve autotrophic growth have been elucidated [93]. Some of these exist only in anaerobic or microaerophilic environments (Chapter 1). Although we worked with top surface samples, the presence of microenvironments in the investigated sites could allow microorganisms to use these processes. Several of the alternative pathways have been found in Archaea. Although we used primers specific for bacterial 16S rRNA genes and did not retrieve archaeal 16S rRNA gene sequences, it is not excluded that the primers also pick up some archaeal 16S rRNA genes. However, previous studies have shown that Archaea are not very abundant in other Antarctic areas [17, 102, 513-515]. Therefore it is likely that they are not very important primary producers. Nevertheless, in addition to primary producers using the Calvin-Benson-Bassham cycle, the possible importance of organisms using one of the five other described carbon dioxide fixation pathways should not be overlooked as it has already been shown that low abundant organisms may very well be the most active ones *in situ* [496].

Although we were unable to determine the nitrogen content, composition and fixation rates in the four sampling locations, it can be expected that nitrogen fixation will also be an important primary process in this oligotrophic environment. As indicated by our results, Cyanobacteria appear to be the main nitrogen-fixing microorganisms. Although many Cyanobacteria are known to be involved in nitrogen fixation, it seems like *Nostoc* spp. are the main contributors. This hypothesis is corroborated by available data from other ice-free areas in Antarctica [125, 126, 128, 130-133]. Thus, although sometimes only present in very low numbers, Cyanobacteria – and in particular *Nostoc* – act a primary nitrogen fixers for the communities inhabiting exposed areas of the Sør Rondane Mountains.

In addition to producers, consumers also play important roles. The different levels of consumers present in ice-free Antarctic areas not only feed on waste products of primary producers, but also on nutrients provided by non-living sources. We discuss here a few examples of taxa we detected in several of our datasets. *Modestobacter* might be involved in granite weathering, perhaps making certain granite-related molecules available for other inhabitants and thus also acting as some kind of 'producer' [388, 389]. Additionally, weathering of the parent material could be important for the community because of the creation of new niches. Armatimonadetes may also perform a special role in the food web: they were shown to consume certain monosaccharides (*e.g.* components of hemicelluloses, which many cellulolytic species cannot utilize) from the environment, thereby facilitating degradation of complex polysaccharides (*e.g.* cellulose) by other organisms known for their ability to use a wide range of substrates (*e.g. Sphingomonas*, *Arthrobacter*) [365, 369, 373-375, 390, 516]. Phototrophic methylotrophs could also play important roles. Several methylotrophic bacteria use methanesulfonate as carbon, sulfur and energy source to support growth [517, 518]. Methanesulfonate, formed by chemical oxidation of dimethylsulfide in the atmosphere, is constantly deposited in Antarctica in snow and by dry deposition, and the stable concentration in ice cores indicates constant methanesulfonate degradation [517, 518]. Since methanesulfonate is a major intermediate in the sulfur cycle, phototrophic methylotrophs could be important contributors to successful biogeochemical cycling.

The aforementioned partial reconstruction of the *in situ* food web is, however, only speculation. To unravel the full extent of the functioning of the communities inhabiting the investigated sites in the Sør Rondane Mountains, new analyses including more samples are needed.

7.4. Future perspectives

Our world is changing at an enormous pace and our hunger for knowledge seems bigger than ever. In the field of microbiology, new studies are published daily, however, only a very small percentage of those focus on Antarctic microbiology. Nevertheless, of those, most focus on coastal areas, which is where most research stations are located, leaving our knowledge on microbial systems and functions in the Antarctic fragmented [14]. Lately, the rise of deep sequencing technologies gave us more insight in the bacterial communities inhabiting different Antarctic ecosystems. The total bacterial community structure is, however, only one small piece of a complex puzzle: it only tells us who is present. To further understand life in this extreme environment, other questions have to be answered: (i) what is the total community structure, (ii) what is the functional diversity, (iii) who is active *in situ*, (iv) what is the relation between (functional) community structure and the (changing) environment and (v) what is the impact of mankind, both directly or indirectly.

Our study was carried out in a microbiological lab with a focus on bacteria. To fully understand who is present and which interactions are at play, all players should be investigated (*i.e.* viruses, Archaea, Bacteria, Eukaryotes). Without taking into account the impact of all Life in these environments, it will remain difficult to extract meaningful (micro)biological information. Therefore, I believe that in order to evaluate players involved in certain processes, joint efforts should be made between different labs/researchers specialized in investigating different organisms. This should also benefit the experimental design. The combined knowledge and experience of a variety of people may lead to novel approaches and robust analyses of the entire community inhabiting a certain area.

Currently available samples do not allow a detailed investigation of the once active *in situ* (functional) community. Residing in a dark temperature controlled room certainly influenced the bacterial communities inhabiting the samples. To increase our knowledge on Life in exposed Antarctic soils, new sampling campaigns should be organized. Replicates should be taken at each selected sampling location at given time intervals. This would not only incorporate microscale heterogeneity at the sites, but also provide information on the evolution of the community. Upon sampling, environmental variables should be determined as soon as possible. The presence of basic laboratory material at the Belgian research station will enable this. Determination of some environmental parameters currently requires quite

some sample material. Given the difficulties to sample in continental Antarctica, efforts should be made to optimize these protocols for small sample volumes. On site extraction of DNA and especially RNA and proteins will provide much more detailed information compared to extraction after transportation to Belgium. DNA would allow new amplicon-based approaches and/or determination of a metagenome sequence. This is a good starting point, but although certain genes or pathways may be very diverse in these data, they may not be important or dominant in the soils near the Princess Elisabeth Station. RNA and proteins would allow metatranscriptomics and metaproteomics approaches which are far more informative [342]. Furthermore, each site should be constantly monitored to track changes in for example temperature and precipitation. These may later help explain changes in the (functional) community composition at a given sampling location. Constant monitoring would also provide us with a better understanding of the effects of ongoing climate changes or human impact in Antarctica.

Although culture-independent approaches will shine a light on the total community and its functioning, DNA sequences that cannot be attributed to known organisms remain a problem [398]. Furthermore, upon DNA extraction, all spatial information at micron-to-millimeter-scales as well as dynamics of individual cells in the sample, are lost. In addition to improving our knowledge about genes and their expression, it is important to connect this information to the molecular physiology of individual taxa and assess how bacteria interact in a community [397]. Cultivation of organisms will thus remain a much needed step, because isolates provide an understanding of the actual bacteria and their physiology. Since many species are resistant to classic cultivation techniques, an *in situ* cultivation, or culturomics approach might overcome this problem [398], hence filling in gaps in our knowledge of the Antarctic microscopic life. For *in situ* cultivation, the use of micro- or nanochambers (*e.g.* Ichip) could lead to the cultivation of bacteria that have restricted growth under laboratory conditions [519, 520]. The culturomics approach in the laboratory should include a very large number of isolation conditions, *i.e.* different media, solidifying agents, headspace concentrations, temperatures, humidity, light intensities, etc. In this respect, information obtained from (meta)genomes may be helpful [521]. Furthermore, gathering enough environmental data would allow a more targeted cultivation approach. Culturomics has proven extremely useful for other environments, for example the human gut [503]. Such a setup should also be employed for Antarctic samples. I believe that future isolation campaigns should definitely include incubation at sub-zero temperatures. Given the *in situ* conditions, bacteria in these

regions are used to functioning at extremely low temperatures. Indeed, previous research has demonstrated bacterial genome replication at subzero temperatures (up to -20 °C) in Arctic permafrost [522], whereas respiration and metabolic activity has been observed at temperatures of -39 °C [523, 524]. While incubation between 0 and 20 °C would lead to the recovery of psychrotrophic species, performing the experiments at negative temperatures between 0 and 50 °C could lead to the (slow) growth of an unseen diversity of psychrophilic bacteria that are currently being missed as they may be overgrown or fail to grow at warmer temperatures. In this respect, the large diversity of taxa only known to exist from environmental sequences is of particular interest. Amplicon-based sequencing of 16S rRNA genes revealed that the exposed areas in the proximity of the Princess Elisabeth Station are inhabited by many bacterial phyla of which very little is known (*i.e.* only few cultured representatives, *e.g.* Acidobacteria) or which are known only to exist from environmental sequencing data (*e.g.* candidate divisions TM7, OD1, WPS-2) [275]. Our approach (Chapter 5) led to the isolation of the first representative of one of these candidate phyla (*i.e.* FBP). This phylum has originally been detected [400] and now also isolated from Antarctic samples. In Chapter 6, the numerous laboratory tests and genome analysis of one of these strains highlighted, among others, its extremely slow growth, high resistance to toxic compounds and narrow substrate utilization spectrum. Given the broad diversity of this group in exposed areas near the Belgian research station, a culturomics approach with media and settings based on information obtained in Chapter 6 may lead to isolation of many more members. In turn, the newly obtained knowledge on these organisms could then be used to further optimize cultivation experiments or biotechnological approaches.

Finally, it is very important to test the importance of newly obtained strains and certain metabolic processes. Despite the fact that we observed quite some diversity of key genes involved in carbon and nitrogen fixation, and phototrophy, little data is available on their contribution to the actual processes. It might be interesting to set up experiments to determine the actual *in situ* fixation rates, not only of the entire community, but also of certain strains. Following actual *in situ* rates is also interesting from an ecological point of view. It is well known that some Antarctic regions are among the most rapidly warming ones on our entire planet [462]. This affects not only the omnipresent microscopic, but also the little macroscopic life on the continent. In recent years, increasing temperatures have led to a southward expansion of the habitats of the two native Antarctic vascular plants (*i.e.* *Colobanthus quitensis* and *Deschampsia antarctica*) [10, 26]. In many, and maybe even all

places on Antartica, a changing environment may have a serious impact on the community composition and functioning, whether or not by the introduction of allochthonous life forms. For example, an increase in temperature or a change in light conditions could greatly influence the rate by which a certain process is carried out [69, 525, 526], hence affecting the entire community. A potentially interesting way to test *in situ* rates is through the evaluation of isotopic pairing. For testing of certain strains that gave promising results under laboratory conditions, *in situ* testing might be done using micro- or nanochambers. This would not only allow to investigate growth and activity of these organisms in their natural habitat, but also under changing environmental conditions. All together, these above experiments would allow more insight in the composition and functioning of Antarctic communities. Afterwards, this knowledge could be applied in biotechnological applications, as reference for future analyses, or for modelling and extrapolation to more complex systems.

Summary

Antarctica is considered one of the last pristine environments on Earth. Its geographic location has resulted in a thick ice-layer covering approximately 99.7 % of the continent's surface area. The little exposed sites are predominantly found along coastal areas or in the Antarctic Peninsula, although some are located in more inland continental Antarctic regions. Although Life was long thought to be absent from these sites, they are inhabited by a wealth of organisms that are mainly microscopic due to Antarctica's longtime isolation, extreme environmental conditions, historical climate change and geological events. These microbial communities are characterized by relatively simple food webs in which Cyanobacteria are often thought to be responsible for primary production, providing organic matter to other organisms through their auto- and diazotrophy (*i.e.* carbon dioxide and nitrogen fixation). While the importance of Cyanobacteria has been demonstrated in many of the Earth's environments, members of this phylum have been found in very low numbers, or were even undetectable, in multiple terrestrial Antarctic sites. For such sites, little is known about who else is present that may contribute to vital primary production processes.

The absence of higher life forms in combination with extreme environmental conditions have led to nutrient-limited Antarctic soils. To overcome this, the bacterial inhabitants may be expected to use alternative energy sources. The abundant sunlight during the austral summer gives phototrophs, *i.e.* those capable of harvesting sunlight to generate chemical energy, a unique advantage. Nevertheless, only very little is known about the diversity of phototrophic bacteria in terrestrial areas on our planet.

Given the scarce knowledge about Bacteria in most terrestrial Antarctic sites and the growing impact of climate change, tourists and scientists, it is of vital importance to understand who is there and who does what before it is too late. This study focused on the presence, diversity and activity of autotrophic, diazotrophic and phototrophic bacteria inhabiting exposed sites of the Sør Rondane Mountains, Dronning Maud Land, East Antarctica using a dual approach combining traditional and new DNA sequencing techniques, and a targeted isolation campaign followed by growth experiments. Because other studies revealed that Cyanobacteria are sometimes only present in very low numbers in ice-free Antarctic areas, including those of the Sør Rondane Mountains, we focused on non-cyanobacterial taxa.

In a first study (Chapter 2) we investigated the presence and diversity of protein encoding genes playing a key role in the most important and widespread autotrophic, diazotrophic and phototrophic pathways, using clone libraries. Notwithstanding the fact that this technique only provides a glimpse of the complete diversity, a large variety of non-cyanobacterial RuBisCO

genes were retrieved, indicating that multiple bacteria may have an important role in carbon fixation. Unexpectedly, nitrogen fixation seemed restricted to Cyanobacteria. At the same time, the two main bacterial phototrophy mechanisms were investigated for the first time in terrestrial Antarctica. Only one mechanism, dependent on (bacterio)chlorophyll, was found present in exposed samples from the Sør Rondane Mountains. Amplification and sequencing of *pufLM* genes indicated that aerobic bacteria using the widespread type 2 photochemical reaction center are present throughout the investigated sites. Key genes involved in rhodopsin-dependent phototrophy, on the other hand, could not be amplified from the DNA.

Next, using these insights, the diversity of these genes was further investigated using the at present most commonly used deep sequencing platform, Illumina's MiSeq (Chapters 3 and 4). The greater sequencing depth verified and extended results obtained using clone libraries. A large diversity of non-cyanobacterial typetype IC RuBisCO genes could be retrieved from the samples. The non-cyanobacterial type IA was found only in minute numbers. Unfortunately, deep sequencing did not extend our view on nitrogen fixing bacteria, as again only cyanobacterial genes were retrieved. Illumina sequencing greatly improved our insights on (bacterio)chlorophyll-dependent phototrophic microorganisms, especially due to the addition of sequencing data for other genes, co-amplified with *nifH*, playing key roles in this pathway.

Based on the insights obtained from the culture-independent analyses, a large scale isolation campaign was performed to recover aerobic phototrophic bacteria from the samples (Chapter 5). The targeted use of oligotrophic media, different incubation conditions and use of a light regimen to mimic the day/night cycle during the onset of austral summer resulted not only in a high number of bacteria belonging to phototrophic taxa, but also other bacteria making up new species, genera, families and even a new phylum. Furthermore, lab testing revealed the potential of several strains, some without known phototrophic capacities, to grow better in light than dark conditions. Unfortunately, due to the slow growth and low biomass production we were not able to include all strains in these analyses.

Lastly, whole genome sequencing and laboratory testing of a strain representing the first cultured representative of the phylum Abditibacteria (previously known as candidate lineage FBP) led to the first insights in the lifestyle and ecological role of these organisms (Chapter 6). The microorganism, for which we propose the name *Abditibacterium utsteinense*, exhibited a strict aerobic oligotrophic chemoheterotrophic lifestyle and a very high resistance to

antibiotics and toxic compounds. Interestingly, genome analysis revealed the absence of two ribosomal proteins, which may explain its slow growth and low biomass production.

In conclusion, the terrestrial sites near the Belgian Princess Elisabeth Station in the Sør Rondane Mountains harbor an enormous bacterial diversity. This thesis supports the hypothesis that, in addition to Cyanobacteria, many other microorganisms present in these environments may be capable of fixing carbon dioxide and harvesting sunlight as an energy source. Although only a limited number of samples could be investigated, it is expected that the study of other sites may lead to similar results. However, new sampling campaigns followed by culture-independent analysis and especially targeted isolation campaigns will be needed to unravel all the secrets that are held by Antarctic bacteria.

Samenvatting

Antarctica wordt beschouwd als een van de laatste ongerepte gebieden op Aarde. Als gevolg van haar geografische locatie wordt ongeveer 99.7 % van Antarctica's oppervlak permanent bedekt door een dikke ijslaag. De weinige ijsvrije gebieden bevinden zich voornamelijk in kustgebieden en op het Antarctische schiereiland. Slechts een beperkt aantal ijsvrije regio's bevinden zich in het binnenland van Continentaal Antarctica. Alhoewel lang gedacht werd dat er geen leven aanwezig was in deze gebieden, worden ze toch bevolkt door een grote diversiteit aan organismen die voornamelijk microscopisch klein zijn ten gevolge van de extincties volgend op historische klimaatsveranderingen en geologische gebeurtenissen resulterend in extreme omgevingscondities en een zeer beperkte herkolonisatie door de isolatie van Antarctica. De aanwezige microbiële gemeenschappen worden gekenmerkt door relatief eenvoudige voedselwebben waarvan vaak aangenomen wordt dat voornamelijk cyanobacteriën er de rol op zich nemen van primaire producent, en op die manier organisch materiaal ter beschikking stellen van andere organismen door hun autotrofe en diazotrofe (stikstoffixerende) levensstijl. Het belang van cyanobacteriën werd reeds veelvuldig aangetoond in vele ecosystemen op aarde. Toch blijken cyanobacteriën regelmatig in heel kleine aantallen aanwezig te zijn in terrestrisch Antarctische gebieden. In dergelijke gevallen is tot op heden heel weinig gekend over welke andere organismen aanwezig zijn die zouden kunnen bijdragen aan primaire productie.

De afwezigheid van hogere levensvormen en de extreme omgevingscondities hebben geleid tot lage hoeveelheden aan nutriënten in Antarctische gronden. Daarom kan verwacht worden dat de aanwezige bacteriën ook andere energiebronnen kunnen benutten. Het gebruik van zonlicht als energiebron kan fototrofen een uniek voordeel geven tijdens de Antarctische zomer. Licht kan immers gebruikt worden om chemische energie aan te maken. Toch is tot op heden heel weinig gekend over de diversiteit aan fototrofe bacteriën die aanwezig is in terrestrische gebieden op Aarde.

Door de toenemende impact van klimaatsverandering, het stijgend aantal toeristen en wetenschappers, en onze geringe kennis over Antarctische bacteriën, is het van vitaal belang om die kennis uit te breiden voor het te laat is. Deze doctoraatsthesis heeft gefocust op de aanwezigheid, diversiteit en activiteit van autotrofe, diazotrofe en fototrofe bacteriën in terrestrische stalen afkomstig uit het Sør Rondanegebergte, Dronning Maud Land, Oost Antarctica. Hiervoor werd gebruik gemaakt van oude en nieuwe DNA-sequentietechnieken en een gerichte isolatiecampagne gevolgd door groei-experimenten. Omdat andere studies hadden aangetoond dat cyanobacteriën soms in hele lage aantallen aanwezig zijn in ijsvrije

Antarctische regio's, inclusief deze van het Sør Rondanegebergte, focuseten we op niet-cyanobacteriële taxa.

In een eerste studie (Chapter 2) werd, met behulp van clone libraries, de aanwezigheid en diversiteit van eiwitcoderende genen nagegaan die een centrale rol spelen in de meest wijdverspreide autotrofie-, diazotrofie- en fototrofiemechanismen. Niettegenstaande deze techniek een beperkte resolutie heeft, toch leverde ze een grote diversiteit aan niet-cyanobacteriële RuBisCO-genen op, wat erop wijst dat tal van andere bacteriën een belangrijke rol zouden kunnen spelen in CO_2-fixatie. Voor stikstoffixatie bleken uitsluitend cyanobacteriën verantwoordelijk te zijn. Tegelijkertijd werd de aanwezigheid en diversiteit van de twee belangrijkste bacteriële fototrofiemechanismen voor de allereerste keer aangetoond in terrestrisch Antarctica. Enkel fototrofie gebruik makend van (bacterio)chlorofyl werd teruggevonden in grondstalen afkomstig uit het Sør Rondanegebergte. Amplificatie en sequeneren van *pufLM* genen toonde aan dat bacteriën die gebruik maken van het wijdverspreide type-2 fotochemisch reactiecentrum aanwezig zijn in alle onderzochte sites. Rhodopsinegenen konden niet geamplifeceerd worden.

Op basis van de bekomen inzichten werd de diversiteit van deze genen verder onderzocht door gebruik van de Illumina MiSeq, het meest gebruikte sequentieplatform op dit ogenblik (Chapters 3 en 4). De grotere sequentiediepte bevestigde niet alleen de resultaten bekomen met de clone libraries, maar breidde deze ook uit. Bij onderzochte autotrofiegenen werd wederom een zeer grote diversiteit aan het niet-cyanobacteriële type IC RuBisCO teruggevonden, terwijl het niet-cyanobacteriële type IA amper gedetecteerd werd. Net zoals bij de clone libraries werden voor stikstoffixatie enkel cyanobacteriële genen gedetecteerd. Het aanwenden van Illumina sequencing zorgde voor een overvloed aan informatie over fototrofe bacteriën die gebruik maken van (bacterio)chlorofyl. Dit kwam hoofdzakelijk door extra data afkomstig van andere sleutelgenen voor dit proces die samen met *nifH* geamplificeerd werden.

De resultaten van de cultuuronafhankelijke studies werden gebruikt om een grootschalige isolatiecampagne op te zetten om aerobe fototrofen op te groeien uit de terrestrische stalen (Chapter 5). Het gebruik van oligotrofe media, verschillende incubatiecondities en een lichtregime om de overgang van zuidpoolwinter naar -zomer na te bootsen leidde niet alleen tot isolatie van een groot aantal bacteriën toebehorend aan fototrofe taxa, maar ook andere

bacteriën die nieuwe species, genera, families en zelfs een nieuw fylum vormen. Daarnaast bevestigden testen in het labo het potentieel van sommige stammen om beter te groeien in het licht dan in het donker. Opmerkelijk genoeg konden bij enkele van deze stammen geen fototrofiegenen gedetecteerd worden. Toch konden niet alle stammen getest worden al gevolg van hun zeer trage groei en lage biomassaproductie.

In een laatste studie werden het genoom en gedrag onder labocondities van een stam die de eerste gecultiveerde representatieve is van het fylum Abditibacteria (vroeger gekend als kandidaatfylum FBP) bestudeerd (Chapter 6). De bekomen resultaten gaven nieuwe inzichten in de levensstijl en ecologische rol van deze organismen. De onderzochte stam, waarvoor we de naam *Abditibacterium utsteinense* voorstellen, beschikt over een strikt aerobe oligotrofe chemoheterotrofe levenswijze en een zeer hoge resistentie tegen antibiotica en toxische stoffen. Opmerkelijk genoeg bleken twee ribosomale eiwitten niet aanwezig in het genoom, wat een mogelijke verklaring zou kunnen zijn voor de trage groei en geringe biomassaproductie.

Samengevat blijkt dat de terrestrische regio's in de nabijheid van het Belgische Prinses Elisabethstation in het Sør Rondanegebergte een enorme weelde aan bacteriën bevatten. Deze doctoraatsthesis ondersteunt de hypothese dat, naast cyanobacteriën, vele andere aanwezige micro-organismen in deze omgevingen zouden kunnen bijdragen aan CO_2-fixatie en gebruik van licht als energiebron. Alhoewel maar een gering aantal stalen kon onderzocht worden, wordt verwacht dat het bestuderen van andere sites gelijkaardige resultaten zal opleveren. Toch zullen in de toekomst nieuwe staalnamecampagnes gevolgd door cultuuronafhankelijke analyses en vooral gerichte isolatiecampagnes nodig zijn om alle geheimen te onsluieren die verborgen liggen in Antarctische bacteriële gemeenschappen.

References

1. Kennedy, A.D., *Antarctic Terrestrial Ecosystem Response to Global Environmental Change.* Annual Review of Ecology and Systematics, 1995. **26**(1): p. 683-704.
2. Chown, S.L. and P. Convey, *Spatial and temporal variability across life's hierarchies in the terrestrial Antarctic.* Philosophical Transactions of the Royal Society of London B: Biological Sciences, 2007. **362**(1488): p. 2307-2331.
3. Elsworth, G., et al., *Enhanced weathering and CO2 drawdown caused by latest Eocene strengthening of the Atlantic meridional overturning circulation.* Nature Geosci, 2017. **10**(3): p. 213-216.
4. Convey, P., et al., *Antarctic terrestrial life – challenging the history of the frozen continent?* Biological Reviews, 2008. **83**(2): p. 103-117.
5. Baker, G.C., L. Ah Tow, and D.A. Cowan, *PCR-based detection of non-indigenous microorganisms in 'pristine' environments.* Journal of Microbiological Methods, 2003. **53**(2): p. 157-164.
6. Bargagli, R., *Antarctica: Geomorphology and Climate Trends*, in *Antarctic Ecosystems: Environmental Contamination, Climate Change, and Human Impact*. 2005, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 1-41.
7. Peck, L.S., et al., *Genomics: applications to Antarctic ecosystems.* Polar Biology, 2005. **28**(5): p. 351-365.
8. Ugolini, F.C. and J.G. Bockheim, *Antarctic soils and soil formation in a changing environment: A review.* Geoderma, 2008. **144**(1–2): p. 1-8.
9. Lythe, M.B. and D.G. Vaughan, *BEDMAP: A new ice thickness and subglacial topographic model of Antarctica.* Journal of Geophysical Research: Solid Earth, 2001. **106**(B6): p. 11335-11351.
10. Convey, P., et al., *Antarctic climate change and the environment.* Antarctic Science, 2009. **21**(06): p. 541-563.
11. Turner, J., et al., *Antarctic Climate Change and the Environment*. 2009, Cambridge: SCAR & Scott Polar Research Institute. 526.
12. Bargagli, R., *Environmental contamination in Antarctic ecosystems.* Science of The Total Environment, 2008. **400**(1–3): p. 212-226.
13. Davies, B.J., et al., *Variable glacier response to atmospheric warming, northern Antarctic Peninsula, 1988–2009.* The Cryosphere, 2012. **6**(5): p. 1031-1048.
14. Chong, C.-W., D.A. Pearce, and P. Convey, *Emerging spatial patterns in Antarctic prokaryotes.* Frontiers in Microbiology, 2015. **6**: p. 1058.
15. Centre, A.A.D. *Ice free areas of Antarctica: buffered by 5 km*. 2010 [cited 2016 May 18th]; Map 13766: Ice free areas of Antarctica: buffered by 5 km]. Available from: https://data.aad.gov.au/aadc/mapcat/display_map.cfm?map_id=13766.
16. Wynn-Williams, D.D., *Microbial Colonization Processes In Antarctic Fellfield Soils : An Experimental Overview (Eleventh Symposium on Polar Biology).* Proceedings of the NIPR Symposium on Polar Biology, 1990. **3**: p. 164-178.
17. Cary, S.C., et al., *On the rocks: the microbiology of Antarctic Dry Valley soils.* Nature Reviews Microbiology, 2010. **8**(2): p. 129-138.
18. Marshall, G.J., *On the annual and semi-annual cycles of precipitation across Antarctica.* International Journal of Climatology, 2009. **29**(15): p. 2298-2308.
19. Palerme, C., et al., *How much snow falls on the Antarctic ice sheet?* The Cryosphere, 2014. **8**(4): p. 1577-1587.
20. Bockheim, J.G., *Soil-Forming Factors in Antarctica*, in *The Soils of Antarctica*, G.J. Bockheim, Editor. 2015, Springer International Publishing: Cham. p. 5-20.
21. Nylen, T.H., A.G. Fountain, and P.T. Doran, *Climatology of katabatic winds in the McMurdo dry valleys, southern Victoria Land, Antarctica.* Journal of Geophysical Research: Atmospheres, 2004. **109**(D3): p. n/a-n/a.
22. Convey, P., *Terrestrial biodiversity in Antarctica – Recent advances and future challenges.* Polar Science, 2010. **4**(2): p. 135-147.

23. Teixeira, L.C.R.S., et al., *Bacterial diversity in rhizosphere soil from Antarctic vascular plants of Admiralty Bay, maritime Antarctica.* ISME J, 2010. **4**(8): p. 989-1001.
24. Convey, P. and M.I. Stevens, *Antarctic Biodiversity.* Science, 2007. **317**(5846): p. 1877-1878.
25. Ozheredova, I.P., et al., *Mechanisms of antarctic vascular plant adaptation to abiotic environmental factors.* Cytology and Genetics, 2015. **49**(2): p. 139-145.
26. Parnikoza, I.Y., I.A. Kozeretska, and V.A. Kunakh, *Vascular Plants of the Maritime Antarctic: Origin and Adaptation.* American Journal of Plant Sciences, 2011. **2**: p. 381-395.
27. Peat, H.J., A. Clarke, and P. Convey, *Diversity and biogeography of the Antarctic flora.* Journal of Biogeography, 2007. **34**(1): p. 132-146.
28. Chown, S.L., et al., *The changing form of Antarctic biodiversity.* Nature, 2015. **522**(7557): p. 431-438.
29. Aislabie, J.M., et al., *Dominant bacteria in soils of Marble Point and Wright Valley, Victoria Land, Antarctica.* Soil Biology and Biochemistry, 2006. **38**(10): p. 3041-3056.
30. Fierer, N., et al., *Cross-biome metagenomic analyses of soil microbial communities and their functional attributes.* Proceedings of the National Academy of Sciences, 2012. **109**(52): p. 21390-21395.
31. Lee, C.K., et al., *The Inter-Valley Soil Comparative Survey: the ecology of Dry Valley edaphic microbial communities.* ISME J, 2012. **6**(5): p. 1046-1057.
32. Chan, Y., et al., *Functional ecology of an Antarctic Dry Valley.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(22): p. 8990-8995.
33. Jensen, H.I., *Report on Antarctic soils. Repts. Sci. Invest. Brit. Antarct. Exped. 1907-1909. Part IV.* Geology, 1916. **2**: p. 89-92.
34. Cowan, D.A., et al., *Microbial ecology and biogeochemistry of continental Antarctic soils.* Frontiers in Microbiology, 2014. **5**.
35. *Glossary of Soil Science Terms 2008*, in *Glossary of Soil Science Terms 2008*. 2008, Soil Science Society of America: Madison, WI. p. 1-82.
36. Bockheim, J.G., *The Soils of Antarctica*. 1 ed. World Soils Book Series, ed. J.G. Bockheim. 2015, Basel: Springer International Publishing Switzerland.
37. Aislabie, J., et al., *Bacterial diversity associated with ornithogenic soil of the Ross Sea region, AntarcticaThis article is one of a selection of papers in the Special Issue on Polar and Alpine Microbiology.* Canadian Journal of Microbiology, 2009. **55**(1): p. 21-36.
38. Zazovskaya, E., et al., *Soils of Queen Maud Land*, in *The Soils of Antarctica*, J.G. Bockheim, Editor. 2015, Springer International Publishing: Basel. p. 21-44.
39. Mergelov, N.S., et al., *Soils of MacRobertson Land*, in *The Soils of Antarctica*, G.J. Bockheim, Editor. 2015, Springer International Publishing: Cham. p. 65-86.
40. Kennedy, A.D., *Water as a limiting factor in the Antarctic terrestrial environment - a biogeographical synthesis.* Arctic and Alpine Research, 1993. **25**(4): p. 308-315.
41. Cowan, D.A. and L.A. Tow, *Endangered Antarctic Environments.* Annual Review of Microbiology, 2004. **58**(1): p. 649-690.
42. McKnight, D.M., et al., *Inorganic N and P dynamics of Antarctic glacial meltwater streams as controlled by hyporheic exchange and benthic autotrophic communities.* Journal of the North American Benthological Society, 2004. **23**(2): p. 171-188.
43. Tiao, G., et al., *Rapid microbial response to the presence of an ancient relic in the Antarctic Dry Valleys.* Nat Commun, 2012. **3**: p. 660.
44. Nelson, A.E., et al., *Age, geographical distribution and taphonomy of an unusual occurrence of mummified crabeater seals on James Ross Island, Antarctic Peninsula.* Antarctic Science, 2008. **20**(5): p. 485-493.
45. Osanai, Y., et al., *Geologic evolution of the Sør Rondane Mountains, East Antarctica: Collision tectonics proposed based on metamorphic processes and magnetic anomalies.* Precambrian Research, 2013. **234**: p. 8-29.

46. Peeters, K., D. Ertz, and A. Willems, *Culturable bacterial diversity at the Princess Elisabeth Station (Utsteinen, Sør Rondane Mountains, East Antarctica) harbours many new taxa.* Systematic and Applied Microbiology, 2011. **34**(5): p. 360-367.
47. Pattyn, F., K. Matsuoka, and J. Berte, *Glacio-meteorological conditions in the vicinity of the Belgian Princess Elisabeth Station, Antarctica.* Antarctic Science, 2009. **22**(1): p. 79-85.
48. Kumar, A., *Katabatic Wind: In Relation With Snow and Glaciers*, in *Encyclopedia of Snow, Ice and Glaciers*, V.P. Singh, P. Singh, and U.K. Haritashya, Editors. 2011, Springer Netherlands: Dordrecht. p. 671-672.
49. Hiruta, S.-i. and Y. Ohyama, *A preliminary report on terrestrial invertebrates in the Asuka station area, Antarctica (16th Symposium on Polar Biology).* Proceedings of the NIPR Symposium on Polar Biology, 1995. **8**: p. 188-193.
50. Ertz, D., et al., *Lichens from the Utsteinen Nunatak (Sør Rondane Mountains, Antarctica), with the description of one new species and the establishment of permanent plots.* 2014, 2014. **191**(1): p. 16.
51. Stevens, M.I. and C.A. D'Haese, *Islands in ice: isolated populations of Cryptopygus sverdrupi (Collembola) among nunataks in the Sør Rondane Mountains, Dronning Maud Land, Antarctica.* Biodiversity, 2014. **15**(2-3): p. 169-177.
52. Tsujimoto, M., et al., *Preliminary description of tardigrade species diversity and distribution pattern around coastal Syowa Station and inland Sør Rondane Mountains, Dronning Maud Land, East Antarctica.* Polar Biology, 2014. **37**(9): p. 1361-1367.
53. Namsaraev, Z., et al., *Biogeography of terrestrial cyanobacteria from Antarctic ice-free areas.* Annals of Glaciology, 2010. **51**(56): p. 171-177.
54. Tytgat, B., et al., *Bacterial diversity assessment in Antarctic terrestrial and aquatic microbial mats: a comparison between bidirectional pyrosequencing and cultivation.* PLoS One, 2014. **9**(6): p. e97564.
55. Obbels, D., et al., *Bacterial and eukaryotic biodiversity patterns in terrestrial and aquatic habitats in the Sør Rondane Mountains, Dronning Maud Land, East Antarctica.* FEMS Microbiology Ecology, 2016. **92**(6): p. fiw041.
56. Suganuma, Y., et al., *East Antarctic deglaciation and the link to global cooling during the Quaternary: evidence from glacial geomorphology and 10Be surface exposure dating of the Sør Rondane Mountains, Dronning Maud Land.* Quaternary Science Reviews, 2014. **97**: p. 102-120.
57. Madigan, M.T., et al., *Brock biology of microorganisms*. 13 ed. 2012.
58. Selesi, D., et al., *Quantification of bacterial RubisCO genes in soils by cbbL targeted real-time PCR.* Journal of Microbiological Methods, 2007. **69**(3): p. 497-503.
59. Kato, S., et al., *Distribution and phylogenetic diversity of cbbM genes encoding RubisCO form II in a deep-sea hydrothermal field revealed by newly designed PCR primers.* Extremophiles, 2012. **16**(2): p. 277-283.
60. Woese, C., *The universal ancestor.* Proceedings of the National Academy of Sciences, 1998. **95**(12): p. 6854-6859.
61. Poole, A., D. Jeffares, and D. Penny, *Early evolution: prokaryotes, the new kids on the block.* BioEssays, 1999. **21**(10): p. 880-889.
62. Brown, J.R., *Ancient horizontal gene transfer.* Nat Rev Genet, 2003. **4**(2): p. 121-132.
63. Zimmer, C., *On the Origin of Eukaryotes.* Science, 2009. **325**(5941): p. 666-668.
64. Martin, W. and M.J. Russell, *On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells.* Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 2003. **358**(1429): p. 59-83.
65. Tolli, J. and G.M. King, *Diversity and structure of bacterial chemolithotrophic communities in pine forest and agroecosystem soils.* Applied and Environmental Microbiology, 2005. **71**(12): p. 8411-8418.

66. Schopf, J.W., *The paleobiological record of photosynthesis.* Photosynthesis Research, 2011. **107**(1): p. 87-101.
67. Tabita, F.R., *The hydroxypropionate pathway of CO2 fixation: Fait accompli.* Proceedings of the National Academy of Sciences, 2009. **106**(50): p. 21015-21016.
68. Zscheischler, J., et al., *A few extreme events dominate global interannual variability in gross primary production*. 2014.
69. Ducat, D.C. and P.A. Silver, *Improving Carbon Fixation Pathways.* Current opinion in chemical biology, 2012. **16**(3-4): p. 337-344.
70. Selesi, D., M. Schmid, and A. Hartmann, *Diversity of Green-Like and Red-Like Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Large-Subunit Genes (cbbL) in Differently Managed Agricultural Soils.* Applied and Environmental Microbiology, 2005. **71**(1): p. 175-184.
71. Bassham, J.A., et al., *The Path of Carbon in Photosynthesis. XXI. The Cyclic Regeneration of Carbon Dioxide Acceptor1.* Journal of the American Chemical Society, 1954. **76**(7): p. 1760-1770.
72. Badger, M.R. and E.J. Bek, *Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle.* Journal of Experimental Botany, 2008. **59**(7): p. 1525-1541.
73. Alfreider, A., et al., *Distribution and diversity of autotrophic bacteria in groundwater systems based on the analysis of RubisCO genotypes.* Systematic and Applied Microbiology, 2009. **32**(2): p. 140-150.
74. Ellis, R.J., *Most abundant protein in the world.* Trends in Biochemical Sciences, 1979. **4**(11): p. 241-244.
75. Kellermann, C., *PhD Dissertation: Autotrophy in Groundwater Ecosystems*, in *Faculty of Biology*. 2009, LMU München: München.
76. Tourova, T.P. and E.M. Spiridonova, *Phylogeny and Evolution of the Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase Genes in Prokaryotes.* Molecular Biology, 2009. **43**(5): p. 713-728.
77. Buchanan, B.B., *Biochemistry & molecular biology of plants*. 2nd ed. ed. Biochemistry and molecular biology of plants. 2015: Chichester : Wiley-Blackwell. 1264 S.
78. Tabita, F.R., *Molecular and cellular regulation of autotrophic carbon dioxide fixation in microorganisms.* Microbiological Reviews, 1988. **52**(2): p. 155-189.
79. Grostern, A. and L. Alvarez-Cohen, *RubisCO-based CO2 fixation and C1 metabolism in the actinobacterium Pseudonocardia dioxanivorans CB1190.* Environmental Microbiology, 2013. **15**(11): p. 3040-3053.
80. Park, S.W., et al., *Presence of duplicate genes encoding a phylogenetically new subgroup of form I ribulose 1,5-bisphosphate carboxylase/oxygenase in Mycobacterium sp. strain JC1 DSM 3803.* Research in Microbiology, 2009. **160**(2): p. 159-165.
81. Murugapiran, S.K., et al., *Thermus oshimai JL-2 and T. thermophilus JL-18 genome analysis illuminates pathways for carbon, nitrogen, and sulfur cycling.* Standards in Genomic Sciences, 2013. **7**(3): p. 449-468.
82. Khadem, A.F., et al., *Autotrophic Methanotrophy in Verrucomicrobia: Methylacidiphilum fumariolicum SolV Uses the Calvin-Benson-Bassham Cycle for Carbon Dioxide Fixation.* Journal of Bacteriology, 2011. **193**(17): p. 4438-4446.
83. Caldwell, P.E., M.R. MacLean, and P.R. Norris, *Ribulose bisphosphate carboxylase activity and a Calvin cycle gene cluster in Sulfobacillus species.* Microbiology, 2007. **153**(7): p. 2231-2240.
84. Van Der Wielen, P.W.J.J., *Diversity of ribulose-1,5-bisphosphate carboxylase/oxygenase large-subunit genes in the MgCl2-dominated deep hypersaline anoxic basin discovery.* FEMS Microbiology Letters, 2006. **259**(2): p. 326-331.
85. Tabita, F.R., et al., *Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2008. **363**(1504): p. 2629-2640.

86. Elsaied, H. and T. Naganuma, *Phylogenetic Diversity of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Large-Subunit Genes from Deep-Sea Microorganisms.* Applied and Environmental Microbiology, 2001. **67**(4): p. 1751-1765.
87. Nakai, R., et al., *Diversity of RuBisCO gene responsible for CO2 fixation in an Antarctic moss pillar.* Polar Biology, 2012. **35**(11): p. 1641-1650.
88. Hanson, T.E. and F.R. Tabita, *A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from Chlorobium tepidum that is involved with sulfur metabolism and the response to oxidative stress.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(8): p. 4397-4402.
89. Furbank, R.T. and W.C. Taylor, *Regulation of Photosynthesis in C3 and C4 Plants: A Molecular Approach.* The Plant Cell, 1995. **7**(7): p. 797-807.
90. Cleland, W.W., et al., *Mechanism of Rubisco: The carbamate as general base.* Chemical Reviews, 1998. **98**(2): p. 549-561.
91. Sidhu, G.K., R. Mehrotra, and S. Mehrotra, *Carbon concentrating mechanisms: in rescue of Rubisco inefficiency.* Acta Physiologiae Plantarum, 2014. **36**(12): p. 3101-3114.
92. Tabita, F.R., et al., *Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs.* Microbiology and Molecular Biology Reviews, 2007. **71**(4): p. 576-+.
93. Hügler, M. and S.M. Sievert, *Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean.* Annual Review of Marine Science, 2011. **3**(1): p. 261-289.
94. Kandler, O. and K.O. Stetter, *Evidence for autotrophic CO2 assimilation in Sulfolobus brierleyi via a reductive carboxylic acid pathway.* Zentralblatt Fur Bakteriologie Mikrobiologie Und Hygiene I Abteilung Originale C-Allgemeine Angewandte Und Okologische Mikrobiologie, 1981. **2**(2): p. 111-121.
95. Berg, I.A., et al., *A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in archaea.* Science, 2007. **318**(5857): p. 1782-1786.
96. Huber, H., et al., *A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum Ignicoccus hospitalis.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(22): p. 7851-7856.
97. Wood, H.G., S.W. Ragsdale, and E. Pezacka, *The acetyl-CoA pathway: a newly discovered pathway of autotrophic growth.* Trends in Biochemical Sciences, 1986. **11**(1): p. 14-18.
98. Bar-Even, A., E. Noor, and R. Milo, *A survey of carbon fixation pathways through a quantitative lens.* Journal of Experimental Botany, 2012. **63**(6): p. 2325-2342.
99. Kong, W., et al., *Diversity and expression of RubisCO genes in a perennially ice-covered Antarctic lake during the polar night transition.* Applied and Environmental Microbiology, 2012. **78**(12): p. 4358-66.
100. Kong, W., et al., *Evidence of form II RubisCO (cbbM) in a perennially ice-covered Antarctic lake.* Fems Microbiology Ecology, 2012. **82**(2): p. 491-500.
101. Dolhi, J.M., et al., *Diversity and spatial distribution of autotrophic communities within and between ice-covered Antarctic lakes (McMurdo Dry Valleys).* Limnology and Oceanography, 2015. **60**(3): p. 977-991.
102. Tebo, B.M., et al., *Microbial communities in dark oligotrophic volcanic ice cave ecosystems of Mt. Erebus, Antarctica.* Frontiers in Microbiology, 2015. **6**: p. 179.
103. Niederberger, T.D., et al., *Carbon-Fixation Rates and Associated Microbial Communities Residing in Arid and Ephemerally Wet Antarctic Dry Valley Soils.* Frontiers in Microbiology, 2015. **6**: p. 1347.
104. Stueken, E.E., et al., *Isotopic evidence for biological nitrogen fixation by molybdenum-nitrogenase from 3.2 Gyr.* Nature, 2015. **520**(7549): p. 666-669.
105. Boyd, E.S. and J.W. Peters, *New insights into the evolutionary history of biological nitrogen fixation.* Frontiers in Microbiology, 2013. **4**: p. 201.
106. Díez, B., et al., *High cyanobacterial nifH gene diversity in Arctic seawater and sea ice brine.* Environmental Microbiology Reports, 2012. **4**(3): p. 360-366.

107. Hamilton, T., E. Boyd, and J. Peters, *Environmental Constraints Underpin the Distribution and Phylogenetic Diversity of nifH in the Yellowstone Geothermal Complex.* Microbial Ecology, 2011. **61**(4): p. 860-870.
108. Raymond, J., et al., *The Natural History of Nitrogen Fixation.* Molecular Biology and Evolution, 2004. **21**(3): p. 541-554.
109. Zhao, Y., et al., *Diversity of Nitrogenase Systems in Diazotrophs.* Journal of Integrative Plant Biology, 2006. **48**(7): p. 745-755.
110. Rees, D.C. and J.B. Howard, *Nitrogenase: standing at the crossroads.* Current Opinion in Chemical Biology, 2000. **4**(5): p. 559-566.
111. Gaby, J.C. and D.H. Buckley, *A comprehensive evaluation of PCR primers to amplify the nifH gene of nitrogenase.* PLoS One, 2012. **7**(7): p. e42149.
112. Hu, Y., C.C. Lee, and M.W. Ribbe, *Vanadium Nitrogenase: A Two-Hit Wonder?* Dalton transactions (Cambridge, England : 2003), 2012. **41**(4): p. 10.1039/c1dt11535a.
113. Hales, B.J., et al., *Isolation of a new vanadium-containing nitrogenase from Azotobacter vinelandii.* Biochemistry, 1986. **25**(23): p. 7253-7255.
114. Robson, R.L., et al., *The alternative nitrogenase of Azotobacter chroococcum is a vanadium enzyme.* Nature, 1986. **322**(6077): p. 388-390.
115. Rubio, L.M. and P.W. Ludden, *Maturation of nitrogenase: a biochemical puzzle.* Journal of Bacteriology, 2005. **187**(2): p. 405-414.
116. Miller, R.W. and R.R. Eady, *Molybdenum and vanadium nitrogenases of Azotobacter chroococcum. Low temperature favours N2 reduction by vanadium nitrogenase.* Biochemical Journal, 1988. **256**(2): p. 429-432.
117. Bishop, P.E., D.M.L. Jarlenski, and D.R. Hetherington, *Evidence for an alternative nitrogen-fixation system in Azotobacter vinelandii.* Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1980. **77**(12): p. 7342-7346.
118. Alloway, B.J., *Molybdenum*, in *Heavy Metals in Soils: Trace Metals and Metalloids in Soils and their Bioavailability*, J.B. Alloway, Editor. 2013, Springer Netherlands: Dordrecht. p. 527-534.
119. Madejón, P., *Vanadium*, in *Heavy Metals in Soils: Trace Metals and Metalloids in Soils and their Bioavailability*, J.B. Alloway, Editor. 2013, Springer Netherlands: Dordrecht. p. 579-587.
120. Betancourt, D.A., et al., *Characterization of Diazotrophs Containing Mo-Independent Nitrogenases, Isolated from Diverse Natural Environments.* Applied and Environmental Microbiology, 2008. **74**(11): p. 3471-3480.
121. Ribbe, M., D. Gadkari, and O. Meyer, *N2 Fixation by Streptomyces thermoautotrophicus Involves a Molybdenum-Dinitrogenase and a Manganese-Superoxide Oxidoreductase That Couple N2 Reduction to the Oxidation of Superoxide Produced from O2 by a Molybdenum-CO Dehydrogenase.* Journal of Biological Chemistry, 1997. **272**(42): p. 26627-26633.
122. MacKellar, D., et al., *Streptomyces thermoautotrophicus does not fix nitrogen.* Scientific Reports, 2016. **6**: p. 20086.
123. Gaby, J.C. and D.H. Buckley, *A global census of nitrogenase diversity.* Environmental Microbiology, 2011. **13**(7): p. 1790-1799.
124. Gaby, J.C. and D.H. Buckley, *A comprehensive aligned nifH gene database: a multipurpose tool for studies of nitrogen-fixing bacteria.* Database: The Journal of Biological Databases and Curation, 2014. **2014**: p. bau001.
125. Olson, J.B., et al., *N2-Fixing Microbial Consortia Associated with the Ice Cover of Lake Bonney, Antarctica.* Microbial Ecology, 1998. **36**(3): p. 231-238.
126. Jungblut, A.D. and B.A. Neilan, *NifH gene diversity and expression in a microbial mat community on the McMurdo Ice Shelf, Antarctica.* Antarctic Science, 2010. **22**(2): p. 117-122.
127. Niederberger, T.D., et al., *Diverse and highly active diazotrophic assemblages inhabit ephemerally wetted soils of the Antarctic Dry Valleys.* Fems Microbiology Ecology, 2012. **82**(2): p. 376-390.
128. Cowan, D.A., et al., *Hypolithic communities: important nitrogen sources in Antarctic desert soils.* Environmental Microbiology Reports, 2011. **3**(5): p. 581-586.

129. Fernández-Valiente, E., et al., *N2-Fixation in Cyanobacterial Mats from Ponds on the McMurdo Ice Shelf, Antarctica.* Microbial Ecology, 2001. **42**(3): p. 338-349.
130. Jung, J., et al., *Change in gene abundance in the nitrogen biogeochemical cycle with temperature and nitrogen addition in Antarctic soils.* Research in Microbiology, 2011. **162**(10): p. 1018-1026.
131. Genuário, D.B., et al., *Characterization of freshwater benthic biofilm-forming Hydrocoryne (Cyanobacteria) isolates from Antarctica.* Journal of Phycology, 2013. **49**(6): p. 1142-1153.
132. Yergeau, E., et al., *Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect.* ISME J, 2007. **1**(2): p. 163-179.
133. Fogg, G.E. and W.D.P. Stewart, *In Situ Determination of Biological Nitrogen Fixation in Antarctica.* British Antarctic Survey Bulletin, 1968. **15**: p. 36-46.
134. Wilde, S.A., et al., *Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago.* Nature, 2001. **409**(6817): p. 175-178.
135. Barber, J., *Photosynthetic energy conversion: natural and artificial.* Chemical Society Reviews, 2009. **38**(1): p. 185-196.
136. Najafpour, M. and B. Pashaei, *Photosynthesis: How and Why?*, in *Advances in Photosynthesis - Fundamental Aspects*, M. Najafpour, Editor. 2012. p. 588.
137. Bryant, D.A. and N.-U. Frigaard, *Prokaryotic photosynthesis and phototrophy illuminated.* Trends in Microbiology, 2006. **14**(11): p. 488-496.
138. Zahnle, K., L. Schaefer, and B. Fegley, *Earth's earliest atmospheres.* Cold Spring Harbor Perspectives in Biology, 2010. **2**(10): p. a004895.
139. Goldblatt, C., et al., *The Eons of Chaos and Hades.* Solid Earth, 2010. **1**(1): p. 1-3.
140. Sagan, C. and G. Mullen, *Earth and Mars: Evolution of Atmospheres and Surface Temperatures.* Science, 1972. **177**(4043): p. 52-56.
141. Zahnle, K.J. and J.C.G. Walker, *The evolution of solar ultraviolet luminosity.* Reviews of Geophysics, 1982. **20**(2): p. 280-292.
142. Olson, J.M., *Photosynthesis in the Archean Era.* Photosynthesis Research, 2006. **88**(2): p. 109-117.
143. Hohmann-Marriott, M.F. and R.E. Blankenship, *Evolution of Photosynthesis.* Annual Review of Plant Biology, 2011. **62**(1): p. 515-548.
144. Allen, J.F., *Origin of Oxygenic Photosynthesis from Anoxygenic Type I and Type II Reaction Centers*, in *The Biophysics of Photosynthesis*, J. Golbeck and A. Est, Editors. 2014, Springer New York: New York, NY. p. 433-450.
145. Buick, R., *When did oxygenic photosynthesis evolve?* Philosophical Transactions of the Royal Society of London B: Biological Sciences, 2008. **363**(1504): p. 2731-2743.
146. Koblížek, M., *Ecology of aerobic anoxygenic phototrophs in aquatic environments.* FEMS Microbiology Reviews, 2015. **39**(6): p. 854-870.
147. Zeng, Y., et al., *Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes.* Proceedings of the National Academy of Sciences, 2014. **111**(21): p. 7795-7800.
148. Bryant, D.A., et al., *Candidatus Chloracidobacterium thermophilum: An Aerobic Phototrophic Acidobacterium.* Science, 2007. **317**(5837): p. 523-526.
149. Feng, Y., et al., *Diversity of aerobic anoxygenic phototrophic bacteria in paddy soil and their response to elevated atmospheric CO2.* Microbial Biotechnology, 2011a. **4**(1): p. 74-81.
150. Harashima, K., et al., *Occurrence of bacteriochlorophyll a in a strain of an aerobic heterotrophic bacterium.* Agricultural and Biological Chemistry, 1978. **42**(8): p. 1627-1628.
151. Hauruseu, D. and M. Koblížek, *Influence of Light on Carbon Utilization in Aerobic Anoxygenic Phototrophs.* Applied and Environmental Microbiology, 2012. **78**(20): p. 7414-7419.
152. Yurkov, V. and J.T. Csotonyi, *New Light on Aerobic Anoxygenic Phototrophs*, in *The Purple Phototrophic Bacteria*, C.N. Hunter, et al., Editors. 2009, Springer Netherlands: Dordrecht. p. 31-55.

153. Atamna-Ismaeel, N., et al., *Bacterial anoxygenic photosynthesis on plant leaf surfaces.* Environmental Microbiology Reports, 2012. **4**(2): p. 209-216.
154. Sato, K., *Bacteriochlorophyll formation by facultative methylotrophs, Protaminobacter ruber and Pseudomonas AM-1.* Febs Letters, 1978. **85**(2): p. 207-210.
155. Fleischman, D. and D. Kramer, *Photosynthetic rhizobia.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1998. **1364**(1): p. 17-36.
156. Harashima, K., T. Shiba, and N. Murata, *Aerobic Photosynthetic Bacteria*. 1989, Tokyo: Japan Scientific Societies Press.
157. Kramer, D.M., A. Kanazawa, and D. Fleischman, *Oxygen dependence of photosynthetic electron transport in a bacteriochlorophyll-containing rhizobium.* FEBS Letters, 1997. **417**(3): p. 275-278.
158. Tanaka, A. and R. Tanaka, *Chlorophyll metabolism.* Current Opinion in Plant Biology, 2006. **9**(3): p. 248-255.
159. Chew, A.G.M. and D.A. Bryant, *Chlorophyll Biosynthesis in Bacteria: The Origins of Structural and Functional Diversity.* Annual Review of Microbiology, 2007. **61**(1): p. 113-129.
160. Masuda, T. and Y. Fujita, *Regulation and evolution of chlorophyll metabolism.* Photochemical & Photobiological Sciences, 2008. **7**(10): p. 1131-1149.
161. Silva, P.J., *With or without light: comparing the reaction mechanism of dark-operative protochlorophyllide oxidoreductase with the energetic requirements of the light-dependent protochlorophyllide oxidoreductase.* PeerJ, 2014. **2**: p. e551.
162. Fujita, Y. and C.E. Bauer, *The Light-Independent Protochlorophyllide Reductase: A Nitrogenase-Like Enzyme Catalyzing a Key Reaction for Greening in the Dark*, in *The Porphyrin Handbook*, R. Guilard, K. Kadish, and K.M. Smith, Editors. 2003, Academic Press: Amsterdam. p. 109-156.
163. Reinbothe, C., et al., *Chlorophyll biosynthesis: spotlight on protochlorophyllide reduction.* Trends in Plant Science, 2010. **15**(11): p. 614-624.
164. Gupta, R.S., *Origin and Spread of Photosynthesis Based upon Conserved Sequence Features in Key Bacteriochlorophyll Biosynthesis Proteins.* Molecular Biology and Evolution, 2012. **29**(11): p. 3397-3412.
165. Croce, R. and H. van Amerongen, *Natural strategies for photosynthetic light harvesting.* Nat Chem Biol, 2014. **10**(7): p. 492-501.
166. Scheer, H., *An Overview of Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications*, in *Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications*, B. Grimm, et al., Editors. 2006, Springer Netherlands: Dordrecht. p. 1-26.
167. Trampe, E. and M. Kühl, *Chlorophyll f distribution and dynamics in cyanobacterial beachrock biofilms.* Journal of Phycology, 2016: p. n/a-n/a.
168. Hamada, F., A. Murakami, and S. Akimoto, *Comparative Analysis of Ultrafast Excitation Energy-Transfer Pathways in Three Strains of Divinyl Chlorophyll a/b-Containing Cyanobacterium, Prochlorococcus marinus.* The Journal of Physical Chemistry B, 2015. **119**(51): p. 15593-15600.
169. Harada, J., et al., *A seventh bacterial chlorophyll driving a large light-harvesting antenna.* Scientific Reports, 2012. **2**: p. 671.
170. Melkozernov, A.N. and R.E. Blankenship, *Photosynthetic Functions of Chlorophylls*, in *Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications*, B. Grimm, et al., Editors. 2006, Springer Netherlands: Dordrecht. p. 397-412.
171. Blankenship, R.E., *Early Evolution of Photosynthesis.* Plant Physiology, 2010. **154**(2): p. 434-438.
172. Hauska, G., et al., *The reaction center of green sulfur bacteria1.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 2001. **1507**(1–3): p. 260-277.
173. Oh-oka, H., *Type 1 Reaction Center of Photosynthetic Heliobacteria†.* Photochemistry and Photobiology, 2007. **83**(1): p. 177-186.

174. Ritchie, A.E. and Z.I. Johnson, *Abundance and Genetic Diversity of Aerobic Anoxygenic Phototrophic Bacteria of Coastal Regions of the Pacific Ocean.* Applied and Environmental Microbiology, 2012. **78**(8): p. 2858-2866.
175. Koh, E.Y., W. Phua, and K.G. Ryan, *Aerobic anoxygenic phototrophic bacteria in Antarctic sea ice and seawater.* Environmental Microbiology Reports, 2011. **3**(6): p. 710-716.
176. Jagannathan, B. and J.H. Golbeck, *Photosynthesis: Microbial A2 - Schaechter, Moselio*, in *Encyclopedia of Microbiology (Third Edition)*. 2009, Academic Press: Oxford. p. 325-341.
177. Perreault, N.N., et al., *Heterotrophic and Autotrophic Microbial Populations in Cold Perennial Springs of the High Arctic.* Applied and Environmental Microbiology, 2008. **74**(22): p. 6898-6907.
178. Koblížek, M., et al., *Genome Sequence of the Marine Photoheterotrophic Bacterium Erythrobacter sp. Strain NAP1.* Journal of Bacteriology, 2011. **193**(20): p. 5881-5882.
179. Ernst, O.P., et al., *Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms.* Chemical Reviews, 2014. **114**(1): p. 126-163.
180. Yutin, N. and E.V. Koonin, *Proteorhodopsin genes in giant viruses.* Biology Direct, 2012. **7**: p. 34.
181. Béjà, O. and J.K. Lanyi, *Nature's toolkit for microbial rhodopsin ion pumps.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(18): p. 6538-6539.
182. Skulachev, V.P., A.V. Bogachev, and F.O. Kasparinsky, *Bacteriorhodopsin*, in *Principles of Bioenergetics*. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 139-156.
183. Bamann, C., et al., *Proteorhodopsin.* Biochimica Et Biophysica Acta-Bioenergetics, 2014. **1837**(5): p. 614-625.
184. Tsukamoto, T., M. Demura, and Y. Sudo, *Irreversible Trimer to Monomer Transition of Thermophilic Rhodopsin upon Thermal Stimulation.* Journal of Physical Chemistry B, 2014. **118**(43): p. 12383-12394.
185. Pushkarev, A. and O. Beja, *Functional metagenomic screen reveals new and diverse microbial rhodopsins.* ISME J, 2016. **10**(9): p. 2331-2335.
186. Béjà, O., et al., *Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea.* Science, 2000. **289**(5486): p. 1902-1906.
187. Giovannoni, S.J., et al., *Proteorhodopsin in the ubiquitous marine bacterium SAR11.* Nature, 2005. **438**(7064): p. 82-85.
188. Gomez-Consarnau, L., et al., *Light stimulates growth of proteorhodopsin-containing marine Flavobacteria.* Nature, 2007. **445**(7124): p. 210-213.
189. Campbell, B.J., et al., *Abundant proteorhodopsin genes in the North Atlantic Ocean.* Environmental Microbiology, 2008. **10**(1): p. 99-109.
190. Boeuf, D., et al., *MicRhoDE: a curated database for the analysis of microbial rhodopsin diversity and evolution.* Database: The Journal of Biological Databases and Curation, 2015. **2015**: p. bav080.
191. Atamna-Ismaeel, N., et al., *Widespread distribution of proteorhodopsins in freshwater and brackish ecosystems.* ISME J, 2008. **2**(6): p. 656-662.
192. Martinez-Garcia, M., et al., *High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton.* ISME J, 2012. **6**(1): p. 113-123.
193. Koh, E.Y., et al., *Proteorhodopsin-Bearing Bacteria in Antarctic Sea Ice.* Applied and Environmental Microbiology, 2010. **76**(17): p. 5918-5925.
194. Petrovskaya, L.E., et al., *Predicted bacteriorhodopsin from Exiguobacterium sibiricum is a functional proton pump.* Febs Letters, 2010. **584**(19): p. 4193-4196.
195. Bohorquez, L.C., C.A. Ruiz-Pérez, and M.M. Zambrano, *Proteorhodopsin-Like Genes Present in Thermoacidophilic High-Mountain Microbial Communities.* Applied and Environmental Microbiology, 2012. **78**(21): p. 7813-7817.
196. Atamna-Ismaeel, N., et al., *Microbial rhodopsins on leaf surfaces of terrestrial plants.* Environmental Microbiology, 2012. **14**(1): p. 140-146.

197. Rusch, D.B., et al., *The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific.* Plos Biology, 2007. **5**(3): p. 398-431.
198. Yooseph, S., et al., *The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families.* PLOS Biology, 2007. **5**(3): p. e16.
199. Sharma, A.K., et al., *Actinorhodopsins: proteorhodopsin-like gene sequences found predominantly in non-marine environments.* Environmental Microbiology, 2008. **10**(4): p. 1039-1056.
200. Warnecke, F., R. Amann, and J. Pernthaler, *Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages.* Environmental Microbiology, 2004. **6**(3): p. 242-253.
201. Sharma, A.K., et al., *Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater Actinobacteria.* ISME J, 2009. **3**(6): p. 726-737.
202. Salka, I., et al., *Distribution of acI-Actinorhodopsin genes in Baltic Sea salinity gradients indicates adaptation of facultative freshwater photoheterotrophs to brackish waters.* Environmental Microbiology, 2014. **16**(2): p. 586-597.
203. Sabehi, G., et al., *New Insights into Metabolic Properties of Marine Bacteria Encoding Proteorhodopsins.* PLoS Biol, 2005. **3**(8): p. e273.
204. Blankenship, R.E., *Molecular Mechanisms of Photosynthesis*. Molecular Mechanisms of Photosynthesis. 2002: Blackwell Science Ltd.
205. Feng, Y., et al., *pH is a good predictor of the distribution of anoxygenic purple phototrophic bacteria in Arctic soils.* Soil Biology and Biochemistry, 2014. **74**(0): p. 193-200.
206. Madigan, M.T., et al., *Rhodoferax antarcticus sp. nov., a moderately psychrophilic purple nonsulfur bacterium isolated from an Antarctic microbial mat.* Archives of Microbiology, 2000. **173**(4): p. 269-277.
207. Wilke, A., et al., *The MG-RAST metagenomics database and portal in 2015.* Nucleic Acids Res, 2016. **44**(D1): p. D590-4.
208. Markowitz, V.M., et al., *IMG/M 4 version of the integrated metagenome comparative analysis system.* Nucleic Acids Research, 2014. **42**(D1): p. D568-D573.
209. Post, E., et al., *Ecological Dynamics Across the Arctic Associated with Recent Climate Change.* Science, 2009. **325**(5946): p. 1355-1358.
210. Smith, R.I.L., *Vascular plants as bioindicators of regional warming in Antarctica.* Oecologia, 1994. **99**(3): p. 322-328.
211. Bockheim, J.G., *Antarctic Soils and Climate Change*, in *The Soils of Antarctica*, G.J. Bockheim, Editor. 2015, Springer International Publishing: Cham. p. 305-314.
212. Matsuoka, N., et al., *Quaternary bedrock erosion and landscape evolution in the Sør Rondane Mountains, East Antarctica: Reevaluating rates and processes.* Geomorphology, 2006. **81**(3-4): p. 408-420.
213. Greene, D.M. and A. Holtom, *Studies in Colobanthus quitensis (Kunth) Bartl. & Deschampsia antarctica Desv. III. Distribution, habitats and perfomance in the Antarctic botanical zone.* British Antarctic Survey Bulletin, 1971. **26**: p. 1-29.
214. Michaud, A.B., M. Sabacka, and J.C. Priscu, *Cyanobacterial diversity across landscape units in a polar desert: Taylor Valley, Antarctica.* FEMS Microbiol Ecol, 2012. **82**(2): p. 268-78.
215. Moodley, K., *M.SC. Thesis: Microbial Diversity of Antarctic Dry Valley Mineral Soil*. 2004: University of the Western Cape.
216. Yergeau, E., et al., *Patterns of bacterial diversity across a range of Antarctic terrestrial habitats.* Environmental Microbiology, 2007. **9**(11): p. 2670-2682.
217. Yergeau, E., et al., *Size and structure of bacterial, fungal and nematode communities along an Antarctic environmental gradient.* Fems Microbiology Ecology, 2007. **59**(2): p. 436-451.
218. Tabita, F.R., *Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective.* Photosynthesis Research, 1999. **60**(1): p. 1-28.
219. Howard, K.S., et al., *Klebsiella pneumoniae nifM Gene Product Is Required For Stabilization And Activation Of Nitrogenase Iron Protein In Escherichia coli.* Journal of Biological Chemistry, 1986. **261**(2): p. 772-778.

220. Fritsen, C.H., A.M. Grue, and J.C. Priscu, *Distribution of organic carbon and nitrogen in surface soils in the McMurdo Dry Valleys, Antarctica.* Polar Biology, 2000. **23**(2): p. 121-128.
221. Zehr, J.P., et al., *Nitrogenase gene diversity and microbial community structure: a cross-system comparison.* Environmental Microbiology, 2003. **5**(7): p. 539-554.
222. Yeager, C.M., et al., *Diazotrophic community structure and function in two successional stages of biological soil crusts from the Colorado plateau and Chihuahuan desert.* Applied and Environmental Microbiology, 2004. **70**(2): p. 973-983.
223. Palovaara, J., et al., *Stimulation of growth by proteorhodopsin phototrophy involves regulation of central metabolic pathways in marine planktonic bacteria.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(35): p. E3650-E3658.
224. Koblížek, M., et al., *Isolation and characterization of Erythrobacter sp. strains from the upper ocean.* Archives of Microbiology, 2003. **180**(5): p. 327-338.
225. Alberti, M., D. Burke, and J. Hearst, *Structure and Sequence of the Photosynthesis Gene Cluster*, in *Anoxygenic Photosynthetic Bacteria*, R. Blankenship, M. Madigan, and C. Bauer, Editors. 1995, Springer Netherlands. p. 1083-1106.
226. Jiang, H., et al., *Response of Aerobic Anoxygenic Phototrophic Bacterial Diversity to Environment Conditions in Saline Lakes and Daotang River on the Tibetan Plateau, NW China.* Geomicrobiology Journal, 2010. **27**(5): p. 400-408.
227. Salka, I., et al., *Rhodoferax-related pufM gene cluster dominates the aerobic anoxygenic phototrophic communities in German freshwater lakes.* Environmental Microbiology, 2011. **13**(11): p. 2865-2875.
228. Achenbach, L.A., J. Carey, and M.T. Madigan, *Photosynthetic and Phylogenetic Primers for Detection of Anoxygenic Phototrophs in Natural Environments.* Applied and Environmental Microbiology, 2001. **67**(7): p. 2922-2926.
229. Cottrell, M.T. and D.L. Kirchman, *Photoheterotrophic Microbes in the Arctic Ocean in Summer and Winter.* Applied and Environmental Microbiology, 2009. **75**(15): p. 4958-4966.
230. Jeanthon, C., et al., *Diversity of cultivated and metabolically active aerobic anoxygenic phototrophic bacteria along an oligotrophic gradient in the Mediterranean Sea.* Biogeosciences, 2011. **8**(7): p. 1955-1970.
231. Lehours, A.-C., et al., *Summer distribution and diversity of aerobic anoxygenic phototrophic bacteria in the Mediterranean Sea in relation to environmental variables.* FEMS Microbiology Ecology, 2010. **74**(2): p. 397-409.
232. Yutin, N., M.T. Suzuki, and O. Béjà, *Novel Primers Reveal Wider Diversity among Marine Aerobic Anoxygenic Phototrophs.* Applied and Environmental Microbiology, 2005. **71**(12): p. 8958-8962.
233. Karr, E.A., et al., *Remarkable Diversity of Phototrophic Purple Bacteria in a Permanently Frozen Antarctic Lake.* Applied and Environmental Microbiology, 2003. **69**(8): p. 4910-4914.
234. Jiang, H., et al., *Abundance and diversity of aerobic anoxygenic phototrophic bacteria in saline lakes on the Tibetan plateau.* FEMS Microbiology Ecology, 2009. **67**(2): p. 268-278.
235. Yuan, H., et al., *Significant Role for Microbial Autotrophy in the Sequestration of Soil Carbon.* Applied and Environmental Microbiology, 2012. **78**(7): p. 2328-2336.
236. Allgaier, M., et al., *Aerobic Anoxygenic Photosynthesis in Roseobacter Clade Bacteria from Diverse Marine Habitats.* Applied and Environmental Microbiology, 2003. **69**(9): p. 5051-5059.
237. Yoshizawa, S., et al., *Diversity and functional analysis of proteorhodopsin in marine Flavobacteria.* Environmental Microbiology, 2012. **14**(5): p. 1240-1248.
238. Wu, L., Y. Cui, and S. Chen, *Diversity of nifH gene sequences in the sediments of South China Sea.* African Journal of Microbiology Research, 2011. **5**(32): p. 5972-5977.
239. Spiridonova, E.M., et al., *An Oligonucleotide Primer System for Amplification of the Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Genes of Bacteria of Various Taxonomic Groups.* Microbiology, 2004. **73**(3): p. 316-325.

240. Nanba, K., G.M. King, and K. Dunfield, *Analysis of Facultative Lithotroph Distribution and Diversity on Volcanic Deposits by Use of the Large Subunit of Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase.* Applied and Environmental Microbiology, 2004. **70**(4): p. 2245-2253.
241. Nagashima, K., et al., *Horizontal transfer of genes coding for the photosynthetic reaction centers of purple bacteria.* Journal of Molecular Evolution, 1997. **45**(2): p. 131-136.
242. Béjà, O., et al., *Unsuspected diversity among marine aerobic anoxygenic phototrophs.* Nature, 2002. **415**(6872): p. 630-633.
243. Poly, F., L.J. Monrozier, and R. Bally, *Improvement in the RFLP procedure for studying the diversity of nifH genes in communities of nitrogen fixers in soil.* Research in Microbiology, 2001. **152**(1): p. 95-103.
244. Markowitz, V.M., et al., *IMG: the integrated microbial genomes database and comparative analysis system.* Nucleic Acids Research, 2012. **40**(Database issue): p. D115-D122.
245. Tamura, K., et al., *MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0.* Molecular Biology and Evolution, 2013. **30**(12): p. 2725-2729.
246. Schloss, P.D., et al., *Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities.* Applied and Environmental Microbiology, 2009. **75**(23): p. 7537-7541.
247. Edgar, R.C., et al., *UCHIME improves sensitivity and speed of chimera detection.* Bioinformatics, 2011. **27**(16): p. 2194-2200.
248. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Molecular Systems Biology, 2011. **7**: p. 539.
249. Goujon, M., et al., *A new bioinformatics analysis tools framework at EMBL-EBI.* Nucleic Acids Research, 2010. **38**: p. W695-W699.
250. Price, M.N., P.S. Dehal, and A.P. Arkin, *FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments.* PLoS ONE, 2010. **5**(3): p. e9490.
251. Letunic, I. and P. Bork, *Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation.* Bioinformatics, 2007. **23**(1): p. 127-128.
252. Letunic, I. and P. Bork, *Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy.* Nucleic Acids Research, 2011. **39**: p. W475-W478.
253. Dixon, P., *VEGAN, a package of R functions for community ecology.* Journal of Vegetation Science, 2003. **14**(6): p. 927-930.
254. Kovaleva, O.L., et al., *Diversity of RuBisCO and ATP citrate lyase genes in soda lake sediments.* Fems Microbiology Ecology, 2011. **75**(1): p. 37-47.
255. Elsaied, H., H. Kimura, and T. Naganuma, *Composition of archaeal, bacterial, and eukaryal RuBisCO genotypes in three Western Pacific arc hydrothermal vent systems.* Extremophiles, 2007. **11**(1): p. 191-202.
256. Kojima, H., H. Fukuhara, and M. Fukui, *Community structure of microorganisms associated with reddish-brown iron-rich snow.* Systematic and Applied Microbiology, 2009. **32**(6): p. 429-437.
257. Feng, Y., et al., *Free-air CO2 enrichment (FACE) enhances the biodiversity of purple phototrophic bacteria in flooded paddy soil.* Plant and Soil, 2009. **324**(1-2): p. 317-328.
258. Feng, Y., et al., *Elevated Ground-Level O-3 Changes the Diversity of Anoxygenic Purple Phototrophic Bacteria in Paddy Field.* Microbial Ecology, 2011b. **62**(4): p. 789-799.
259. Tourova, T.P., et al., *Ribulose-1,5-bisphosphate carboxylase/oxygenase genes as a functional marker for chemolithoautotrophic halophilic sulfur-oxidizing bacteria in hypersaline habitats.* Microbiology, 2010. **156**(7): p. 2016-2025.
260. Tourova, T.P., et al., *Application of ribulose-1,5-bisphosphate carboxylase/oxygenase genes as molecular markers for assessment of the diversity of autotrophic microbial communities inhabiting the upper sediment horizons of the saline and soda lakes of the Kulunda Steppe.* Microbiology, 2011. **80**(6): p. 812-825.

261. Yousuf, B., et al., *Application of targeted metagenomics to explore abundance and diversity of CO2-fixing bacterial community using cbbL gene from the rhizosphere of Arachis hypogaea.* Gene, 2012. **506**(1): p. 18-24.
262. Boyd, E.S., et al., *Chemolithotrophic Primary Production in a Subglacial Ecosystem.* Applied and Environmental Microbiology, 2014. **80**(19): p. 6146-6153.
263. Delwiche, C.F. and J.D. Palmer, *Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids.* Molecular Biology and Evolution, 1996. **13**(6): p. 873-882.
264. Dodds, W.K., D.A. Gudder, and D. Mollenhauer, *The Ecology Of Nostoc.* Journal of Phycology, 1995. **31**(1): p. 2-18.
265. Hawes, I., C. Howard-Williams, and W.F. Vincent, *Desiccation And Recovery Of Antarctic Cyanobacterial Mats.* Polar Biology, 1992. **12**(6-7): p. 587-594.
266. Jung, J.Y., et al., *Spectroscopic and photochemical analysis of proteorhodopsin variants from the surface of the Arctic Ocean.* FEBS Letters, 2008. **582**(12): p. 1679-1684.
267. Tank, M., V. Thiel, and J.F. Imhoff, *Phylogenetic relationship of phototrophic purple sulfur bacteria according to pufL and pufM genes.* International Microbiology, 2009. **12**(3): p. 175-185.
268. Feng, Y., et al., *Soil purple phototrophic bacterial diversity under double cropping (rice-wheat) with free-air CO2 enrichment (FACE).* European Journal of Soil Science, 2011c. **62**(4): p. 533-540.
269. Waidner, L.A. and D.L. Kirchman, *Diversity and Distribution of Ecotypes of the Aerobic Anoxygenic Phototrophy Gene pufM in the Delaware Estuary.* Applied and Environmental Microbiology, 2008. **74**(13): p. 4012-4021.
270. Masin, M., et al., *Distribution of aerobic anoxygenic phototrophs in temperate freshwater systems.* Environmental Microbiology, 2008. **10**(8): p. 1988-1996.
271. Temmerman, R., et al., *Culture-Independent Analysis of Probiotic Products by Denaturing Gradient Gel Electrophoresis.* Applied and Environmental Microbiology, 2003. **69**(1): p. 220-226.
272. Yutin, N., O. Béjà, and M.T. Suzuki, *The use of denaturing gradient gel electrophoresis with fully degenerate pufM primers to monitor aerobic anoxygenic phototrophic assemblages.* Limnology and Oceanography-Methods, 2008. **6**: p. 427-440.
273. Wood, S.A., et al., *Sources of edaphic cyanobacterial diversity in the Dry Valleys of Eastern Antarctica.* Isme Journal, 2008. **2**(3): p. 308-320.
274. Magalhaes, C., et al., *At limits of life: multidisciplinary insights reveal environmental constraints on biotic diversity in continental Antarctica.* PLoS One, 2012. **7**(9): p. e44578.
275. Tytgat, B., et al., *Bacterial community composition in relation to bedrock type and macrobiota in soils from the Sør Rondane Mountains, East Antarctica.* FEMS Microbiology Ecology, 2016. **92**(9): p. fiw126.
276. Fernández-Carazo, R., et al., *Cyanobacterial diversity for an anthropogenic impact assessment in the Sør Rondane Mountains area, Antarctica.* Antarctic Science, 2012. **24**(03): p. 229-242.
277. Tahon, G., et al., *Analysis of cbbL, nifH, and pufLM in Soils from the Sør Rondane Mountains, Antarctica, Reveals a Large Diversity of Autotrophic and Phototrophic Bacteria.* Microbial Ecology, 2016. **71**(1): p. 131-149.
278. Kusian, B. and B. Bowien, *Organization and regulation of cbb CO2 assimilation genes in autotrophic bacteria.* FEMS Microbiology Reviews, 1997. **21**(2): p. 135-155.
279. Ando, S., et al., *Detection of nifH Sequences in Sugarcane (Saccharum officinarum L.) and Pineapple (Ananas comosus [L.] Merr.).* Soil Science & Plant Nutrition, 2005. **51**(2): p. 303-308.
280. Edwards, U., et al., *Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA.* Nucleic Acids Research, 1989. **17**(19): p. 7843-7853.

281. Cleenwerck, I., et al., *Acetobacter ghanensis sp. nov., a novel acetic acid bacterium isolated from traditional heap fermentations of Ghanaian cocoa beans.* International Journal of Systematic and Evolutionary Microbiology, 2007. **57**(7): p. 1647-1652.
282. Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST.* Bioinformatics, 2010. **26**(19): p. 2460-2461.
283. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads.* EMBnet.journal, 2011. **17**(1): p. 10.
284. Edgar, R.C., *UPARSE: highly accurate OTU sequences from microbial amplicon reads.* Nature Methods, 2013. **10**(10): p. 996-998.
285. Schloss, P.D., *The Effects of Alignment Quality, Distance Calculation Method, Sequence Filtering, and Region on the Analysis of 16S rRNA Gene-Based Studies.* PLoS Computational Biology, 2010. **6**(7): p. 1-16.
286. DeSantis, T.Z., et al., *Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB.* Applied and Environmental Microbiology, 2006. **72**(7): p. 5069-5072.
287. McDonald, D., et al., *An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea.* ISME J, 2012. **6**(3): p. 610-618.
288. Wang, Q., et al., *Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy.* Applied and Environmental Microbiology, 2007. **73**(16): p. 5261-5267.
289. Li, W. and A. Godzik, *Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences.* Bioinformatics, 2006. **22**(13): p. 1658-1659.
290. Fu, L., et al., *CD-HIT: accelerated for clustering the next-generation sequencing data.* Bioinformatics, 2012. **28**(23): p. 3150-3152.
291. Joseph, S.J., et al., *Laboratory cultivation of widespread and previously uncultured soil bacteria.* Applied and Environmental Microbiology, 2003. **69**(12): p. 7210-7215.
292. Corinaldesi, C., R. Danovaro, and A. Dell'Anno, *Simultaneous Recovery of Extracellular and Intracellular DNA Suitable for Molecular Studies from Marine Sediments.* Applied and Environmental Microbiology, 2005. **71**(1): p. 46-50.
293. Zwart, G., et al., *Nearly identical 16S rRNA sequences recovered from lakes in North America and Europe indicate the existence of clades of globally distributed freshwater bacteria.* Systematic and Applied Microbiology, 1998. **21**(4): p. 546-556.
294. Giraud, E., et al., *Legumes symbioses: Absence of Nod genes in photosynthetic bradyrhizobia.* Science, 2007. **316**(5829): p. 1307-1312.
295. Wong, F.Y.K., et al., *Phylogenetic Analysis of Bradyrhizobium japonicum and Photosynthetic Stem-Nodulating Bacteria from Aeschynomene Species Grown in Separated Geographical Regions.* Applied and Environmental Microbiology, 1994. **60**(3): p. 940-946.
296. Moniz, M.B.J., et al., *Molecular phylogeny of Antarctic Prasiola (Prasiolales, Trebouxiophyceae) reveals extensive cryptic diversity.* Journal of Phycology, 2012. **48**(4): p. 940-955.
297. Lemieux, C., C. Otis, and M. Turmel, *Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae.* BMC Evolutionary Biology, 2014. **14**(1): p. 1-15.
298. Balarinová, K., et al., *Temperature-dependent growth rate and photosynthetic performance of Antarctic symbiotic alga Trebouxia sp. cultivated in a bioreactor.* Czech Polar Reports, 2013. **3**(1): p. 19-27.
299. Nash, T.H., *Lichen biology*. 1996, Cambridge: Cambridge University Press.
300. Henskens, F.L., T.G. Allan Green, and A. Wilkins, *Cyanolichens can have both cyanobacteria and green algae in a common layer as major contributors to photosynthesis.* Annals of Botany, 2012. **110**(3): p. 555-563.

301. Eckford, R., et al., *Free-Living Heterotrophic Nitrogen-Fixing Bacteria Isolated from Fuel-Contaminated Antarctic Soils.* Applied and Environmental Microbiology, 2002. **68**(10): p. 5181-5185.
302. Pointing, S.B., et al., *Highly specialized microbial diversity in hyper-arid polar desert.* Proceedings of the National Academy of Sciences, 2009. **106**(47): p. 19964-19969.
303. Van Horn, D.J., et al., *Factors Controlling Soil Microbial Biomass and Bacterial Diversity and Community Composition in a Cold Desert Ecosystem: Role of Geographic Scale.* Plos One, 2013. **8**(6): p. e66103.
304. Bakermans, C., et al., *Molecular characterization of bacteria from permafrost of the Taylor Valley, Antarctica.* Fems Microbiology Ecology, 2014. **89**(2): p. 331-346.
305. Kim, M., et al., *Highly Heterogeneous Soil Bacterial Communities around Terra Nova Bay of Northern Victoria Land, Antarctica.* Plos One, 2015. **10**(3): p. e0119966.
306. Butterfield, N.J., *Proterozoic photosynthesis – a critical review.* Palaeontology, 2015. **58**(6): p. 953-972.
307. Cardona, T., *Reconstructing the Origin of Oxygenic Photosynthesis: Do Assembly and Photoactivation Recapitulate Evolution?* Frontiers in Plant Science, 2016. **7**: p. 257.
308. Nomata, J., et al., *Dark-operative protochlorophyllide oxidoreductase generates substrate radicals by an iron-sulphur cluster in bacteriochlorophyll biosynthesis.* Scientific Reports, 2014. **4**: p. 5455.
309. Beale, S.I., *Enzymes of chlorophyll biosynthesis.* Photosynthesis Research, 1999. **60**(1): p. 43-73.
310. Do, H., et al., *Structure-based characterization and antifreeze properties of a hyperactive ice-binding protein from the Antarctic bacterium Flavobacterium frigoris PS1.* Acta Crystallographica Section D, 2014. **70**(4): p. 1061-1073.
311. Qin, Q.-L., et al., *Genome Sequence of Proteorhodopsin-Containing Sea Ice Bacterium Glaciecola punicea ACAM 611(T).* Journal of Bacteriology, 2012. **194**(12): p. 3267-3267.
312. Béjà, O., et al., *Proteorhodopsin phototrophy in the ocean.* Nature, 2001. **411**(6839): p. 786-789.
313. Williams, T.J., et al., *A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters.* Isme Journal, 2012. **6**(10): p. 1883-1900.
314. Yau, S., et al., *Metagenomic insights into strategies of carbon conservation and unusual sulfur biogeochemistry in a hypersaline Antarctic lake.* Isme Journal, 2013. **7**(10): p. 1944-1961.
315. Pedrós-Alió, C., *The Rare Bacterial Biosphere.* Annual Review of Marine Science, 2012. **4**(1): p. 449-466.
316. Logares, R., et al., *Patterns of Rare and Abundant Marine Microbial Eukaryotes.* Current Biology, 2014. **24**(8): p. 813-821.
317. Zhan, A., et al., *High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities.* Methods in Ecology and Evolution, 2013. **4**(6): p. 558-565.
318. Kauserud, H., et al., *High consistency between replicate 454 pyrosequencing analyses of ectomycorrhizal plant root samples.* Mycorrhiza, 2012. **22**(4): p. 309-315.
319. Zhan, A., et al., *Reproducibility of pyrosequencing data for biodiversity assessment in complex communities.* Methods in Ecology and Evolution, 2014. **5**(9): p. 881-890.
320. Pascual, J., et al., *Assessing Bacterial Diversity in the Rhizosphere of Thymus zygis Growing in the Sierra Nevada National Park (Spain) through Culture-Dependent and Independent Approaches.* PLOS ONE, 2016. **11**(1): p. e0146558.
321. Gotelli, N.J. and R.K. Colwell, *Estimating species richness*, in *Biological Diversity: Frontiers in Measurement and Assessment*, A.E. Magurran and B.J. McGill, Editors. 2010, Oxford University Press: Oxford. p. 39-54.
322. McMurdie, P.J. and S. Holmes, *Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible.* PLOS Computational Biology, 2014. **10**(4): p. e1003531.

323. Ferrera, I., et al., *Marked seasonality of aerobic anoxygenic phototrophic bacteria in the coastal NW Mediterranean Sea as revealed by cell abundance, pigment concentration and pyrosequencing of pufM gene.* Environmental Microbiology, 2014. **16**(9): p. 2953-2965.
324. Du, H., et al., *Real-time PCR for quantification of aerobic anoxygenic phototrophic bacteria based on pufM gene in marine environment.* Journal of Experimental Marine Biology and Ecology, 2006. **329**(1): p. 113-121.
325. Oz, A., et al., *Roseobacter-Like Bacteria in Red and Mediterranean Sea Aerobic Anoxygenic Photosynthetic Populations.* Applied and Environmental Microbiology, 2005. **71**(1): p. 344-353.
326. Van Trappen, S., J. Mergaert, and J. Swings, *Loktanella salsilacus gen. nov., sp. nov., Loktanella fryxellensis sp. nov. and Loktanella vestfoldensis sp. nov., new members of the Rhodobacter group, isolated from microbial mats in Antarctic lakes.* International Journal of Systematic and Evolutionary Microbiology, 2004. **54**(4): p. 1263-1269.
327. González, J.M. and M.A. Moran, *Numerical dominance of a group of marine bacteria in the alpha-subclass of the class Proteobacteria in coastal seawater.* Applied and Environmental Microbiology, 1997. **63**(11): p. 4237-42.
328. Wagner-Döbler, I. and H. Biebl, *Environmental Biology of the Marine Roseobacter Lineage.* Annual Review of Microbiology, 2006. **60**(1): p. 255-280.
329. Buchan, A., J.M. González, and M.A. Moran, *Overview of the Marine Roseobacter Lineage.* Applied and Environmental Microbiology, 2005. **71**(10): p. 5665-5677.
330. Liu, J.-H., et al., *Salinarimonas rosea gen. nov., sp. nov., a new member of the α-2 subgroup of the Proteobacteria.* International Journal of Systematic and Evolutionary Microbiology, 2010. **60**(1): p. 55-60.
331. Luo, G., et al., *Skermanella stibiiresistens sp. nov., a highly antimony-resistant bacterium isolated from coal-mining soil, and emended description of the genus Skermanella.* International Journal of Systematic and Evolutionary Microbiology, 2012. **62**(6): p. 1271-1276.
332. Igarashi, N., et al., *Horizontal Transfer of the Photosynthesis Gene Cluster and Operon Rearrangement in Purple Bacteria.* Journal of Molecular Evolution, 2000. **52**(4): p. 333-341.
333. Okubo, Y., H. Futamata, and A. Hiraishi, *Characterization of phototrophic purple nonsulfur bacteria forming colored microbial mats in a swine wastewater ditch.* Applied and Environmental Microbiology, 2006. **72**(9): p. 6225-6233.
334. Jezberova, J., J. Jezbera, and M.W. Hahn, *Insights into Variability of Actinorhodopsin Genes of the LG1 Cluster in Two Different Freshwater Habitats.* Plos One, 2013. **8**(7).
335. Wurzbacher, C., I. Salka, and H.-P. Grossart, *Environmental actinorhodopsin expression revealed by a new in situ filtration and fixation sampler.* Environmental Microbiology Reports, 2012. **4**(5): p. 491-497.
336. Sabehi, G., et al., *Novel Proteorhodopsin variants from the Mediterranean and Red Seas.* Environmental Microbiology, 2003. **5**(10): p. 842-849.
337. Sabehi, G., et al., *Different SAR86 subgroups harbour divergent proteorhodopsins.* Environmental Microbiology, 2004. **6**(9): p. 903-910.
338. Tahon, G., B. Tytgat, and A. Willems, *Diversity of Phototrophic Genes Suggests Multiple Bacteria May Be Able to Exploit Sunlight in Exposed Soils from the Sør Rondane Mountains, East Antarctica.* Frontiers in Microbiology, 2016. **7**(2026).
339. Daubin, V. and G.J. Szollosi, *Horizontal Gene Transfer and the History of Life.* Cold Spring Harbor Perspectives in Biology, 2016. **8**(4).
340. Mizrahi-Man, O., E.R. Davenport, and Y. Gilad, *Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs.* PLoS ONE, 2013. **8**(1): p. e53608.
341. Nichols, D., *Cultivation gives context to the microbial ecologist.* FEMS Microbiology Ecology, 2007. **60**(3): p. 351-357.
342. Prosser, J.I., *Dispersing misconceptions and identifying opportunities for the use of 'omics' in soil microbial ecology.* Nat Rev Micro, 2015. **13**(7): p. 439-446.

343. Vasileva-Tonkova, E., et al., *Ecophysiological properties of cultivable heterotrophic bacteria and yeasts dominating in phytocenoses of Galindez Island, maritime Antarctica.* World Journal of Microbiology & Biotechnology, 2014. **30**(4): p. 1387-1398.
344. Ferrer, M., et al., *Mining enzymes from extreme environments.* Current Opinion in Microbiology, 2007. **10**(3): p. 207-214.
345. Tamames, J., et al., *Environmental distribution of prokaryotic taxa.* BMC Microbiology, 2010. **10**(1): p. 85.
346. Storelli, N., et al., *CO2 assimilation in the chemocline of Lake Cadagno is dominated by a few types of phototrophic purple sulfur bacteria.* Fems Microbiology Ecology, 2013. **84**(2): p. 421-432.
347. Wahlund, T.M., et al., *A thermophilic green sulfur bacterium from new-zealand hot-springs, Chlorobium tepidum sp. nov.* Archives of Microbiology, 1991. **156**(2): p. 81-90.
348. Hirose, S., et al., *Diversity of Purple Phototrophic Bacteria, Inferred from pufM Gene, within Epilithic Biofilm in Tama River, Japan.* Microbes and Environments, 2012. **27**(3): p. 327-329.
349. Tayeh, M.A. and M.T. Madigan, *Malate-dehydrogenase in phototrophic purple bacteria: purification, molecular weight, and quaternary structure.* Journal of Bacteriology, 1987. **169**(9): p. 4196-4202.
350. Cho, J.-C. and S.J. Giovannoni, *Cultivation and Growth Characteristics of a Diverse Group of Oligotrophic Marine Gammaproteobacteria.* Applied and Environmental Microbiology, 2004. **70**(1): p. 432-440.
351. Yurkov, V.V. and J.T. Beatty, *Aerobic anoxygenic phototrophic bacteria.* Microbiology and Molecular Biology Reviews, 1998. **62**(3): p. 695-+.
352. Cameron, R.E., *Abundance of Microflora in Soils of Desert Regions.* NASA Technical Reports, 1969. **32**(1378): p. 1-16.
353. Ranganayaki, S. and C. Mohan, *Effect of sodium molybdate on microbial fixation of nitrogen.* Zeitschrift für allgemeine Mikrobiologie, 1981. **21**(8): p. 607-610.
354. Jensen, H.L., *Nitrogen fixation in leguminous plants. I. General characters of root-nodule bacteria isolated from species of Medicago and Trifolium in Australia.* Proceedings of The Linnean Society of New South Wales, 1942. **67**: p. 98--108.
355. Benemann, J.R., et al., *The vanadium effect in nitrogen fixation by azotobacter.* Biochimica et Biophysica Acta (BBA) - General Subjects, 1972. **264**(1): p. 25-38.
356. Bellenger, J.P., T. Wichard, and A.M.L. Kraepiel, *Vanadium Requirements and Uptake Kinetics in the Dinitrogen-Fixing Bacterium Azotobacter vinelandii.* Applied and Environmental Microbiology, 2008. **74**(5): p. 1478-1484.
357. Wieme, A.D., et al., *Effects of Growth Medium on Matrix-Assisted Laser Desorption–Ionization Time of Flight Mass Spectra: a Case Study of Acetic Acid Bacteria.* Applied and Environmental Microbiology, 2014. **80**(4): p. 1528-1538.
358. Niemann, S., et al., *Evaluation of the resolving power of three different DNA fingerprinting methods to discriminate among isolates of a natural Rhizobium meliloti population.* Journal of Applied Microbiology, 1997. **82**(4): p. 477-484.
359. Coenye, T., et al., *Classification of Alcaligenes faecalis-like isolates from the environment and human clinical samples as Ralstonia gilardii sp. nov.* International Journal of Systematic Bacteriology, 1999. **49**: p. 405-413.
360. Marusina, A.I., et al., *A System of Oligonucleotide Primers for the Amplification of nifH Genes of Different Taxonomic Groups of Prokaryotes.* Microbiology, 2001. **70**(1): p. 73-78.
361. Chun, J., et al., *EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences.* International Journal of Systematic and Evolutionary Microbiology, 2007. **57**(10): p. 2259-2261.
362. Kahm, M., et al., *grofit: Fitting Biological Growth Curves with R.* Journal of Statistical Software, 2010. **33**(7): p. 1-21.

363. Schloss, P.D. and J. Handelsman, *Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness.* Applied and Environmental Microbiology, 2005. **71**(3): p. 1501-1506.
364. Lee, K.C.Y., et al., *Phylogenetic Delineation of the Novel Phylum Armatimonadetes (Former Candidate Division OP10) and Definition of Two Novel Candidate Divisions.* Applied and Environmental Microbiology, 2013. **79**(7): p. 2484-2487.
365. Aislabie, J.M., et al., *Bacterial composition of soils of the Lake Wellman area, Darwin Mountains, Antarctica.* Extremophiles, 2013. **17**(5): p. 775-786.
366. Eloe-Fadrosh, E.A., et al., *Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs.* Nature Communications, 2016. **7**.
367. Boeuf, D., et al., *Summer community structure of aerobic anoxygenic phototrophic bacteria in the western Arctic Ocean.* Fems Microbiology Ecology, 2013. **85**(3): p. 417-432.
368. Cottrell, M.T., A. Mannino, and D.L. Kirchman, *Aerobic Anoxygenic Phototrophic Bacteria in the Mid-Atlantic Bight and the North Pacific Gyre.* Applied and Environmental Microbiology, 2006. **72**(1): p. 557-564.
369. Gunnigle, E., et al., *Draft genomic DNA sequence of the multi-resistant Sphingomonas sp. strain AntH11 isolated from an Antarctic hypolith.* FEMS Microbiology Letters, 2015. **362**(8).
370. Busse, H.-J., et al., *Sphingomonas aurantiaca sp. nov., Sphingomonas aerolata sp. nov. and Sphingomonas faeni sp. nov., air- and dustborne and Antarctic, orange-pigmented, psychrotolerant bacteria, and emended description of the genus Sphingomonas.* International Journal of Systematic and Evolutionary Microbiology, 2003. **53**(5): p. 1253-1260.
371. Hiraishi, A., H. Kuraishi, and K. Kawahara, *Emendation of the description of Blastomonas natatoria (Sly 1985) Sly and Cahill 1997 as an aerobic photosynthetic bacterium and reclassification of Erythromonas ursincola Yurkov et al. 1997 as Blastomonas ursincola comb. nov.* International Journal of Systematic and Evolutionary Microbiology, 2000. **50**(3): p. 1113-1118.
372. Salka, I., et al., *The Draft Genome Sequence of Sphingomonas sp. Strain FukuSWIS1, Obtained from Acidic Lake Grosse Fuchskuhle, Indicates Photoheterotrophy and a Potential for Humic Matter Degradation.* Genome Announcements, 2014. **2**(6): p. e01183-14.
373. Baraniecki, C.A., J. Aislabie, and J.M. Foght, *Characterization of Sphingomonas sp. Ant 17, an Aromatic Hydrocarbon-Degrading Bacterium Isolated from Antarctic Soil.* Microbial Ecology, 2002. **43**(1): p. 44-54.
374. Huang, J., et al., *Sphingomonas yantingensis sp. nov., a mineral-weathering bacterium isolated from purplish paddy soil.* International Journal of Systematic and Evolutionary Microbiology, 2014. **64**(3): p. 1030-1034.
375. Hortnagl, P., M.T. Perez, and R. Sommaruga, *Living at the border: A community and single-cell assessment of lake bacterioneuston activity.* Limnology and Oceanography, 2010. **55**(3): p. 1134-1144.
376. Onofri, S., et al., *Antarctic microfungi as models for exobiology.* Planetary and Space Science, 2004. **52**(1–3): p. 229-237.
377. Romanovskaya, V., et al., *Microbial diversity in terrestrial Antarctic biotopes.* Ukr Antarct J, 2009. **8**: p. 364-369.
378. Vuilleumier, S., et al., *Methylobacterium Genome Sequences: A Reference Blueprint to Investigate Microbial Metabolism of C1 Compounds from Natural and Industrial Sources.* PLOS ONE, 2009. **4**(5): p. e5584.
379. Hirsch, P., et al., *Hymenobacter roseosalivarius gen. nov., sp. nov. from continental Antarctic soils and sandstone: Bacteria of the Cytophaga/Flavobacterium/Bacteroides line of phylogenetic descent.* Systematic and Applied Microbiology, 1998. **21**(3): p. 374-383.
380. Klassen, J.L. and J.M. Foght, *Characterization of Hymenobacter isolates from Victoria Upper Glacier, Antarctica reveals five new species and substantial non-vertical evolution within this genus.* Extremophiles, 2011. **15**(1): p. 45-57.

381. Huang, Y., et al., *Novel acsF gene primers revealed a diverse phototrophic bacterial population including Gemmatimonadetes in the Lake Taihu.* Applied and Environmental Microbiology, 2016.
382. Srinivasan, S., et al., *Hymenobacter terrae sp. nov., a Bacterium Isolated from Soil.* Current Microbiology, 2015. **70**(5): p. 643-650.
383. Mevs, U., et al., *Modestobacter multiseptatus gen. nov., sp nov., a budding actinomycete from soils of the Asgard Range (Transantarctic Mountains).* International Journal of Systematic and Evolutionary Microbiology, 2000. **50**: p. 337-346.
384. Peeters, K., et al., *Heterotrophic bacterial diversity in aquatic microbial mat communities from Antarctica.* Polar Biology, 2012. **35**(4): p. 543-554.
385. Babalola, O.O., et al., *Phylogenetic analysis of actinobacterial populations associated with Antarctic Dry Valley mineral soils.* Environmental Microbiology, 2009. **11**(3): p. 566-576.
386. Saul, D.J., et al., *Hydrocarbon contamination changes the bacterial diversity of soil from around Scott Base, Antarctica.* FEMS Microbiology Ecology, 2005. **53**(1): p. 141-155.
387. Shivaji, S., et al., *Antarctic ice core samples: culturable bacterial diversity.* Research in Microbiology, 2013. **164**(1): p. 70-82.
388. Gtari, M., et al., *Contrasted resistance of stone-dwelling Geodermatophilaceae species to stresses known to give rise to reactive oxygen species.* Fems Microbiology Ecology, 2012. **80**(3): p. 566-577.
389. Normand, P., et al., *Genome Sequence of Radiation-Resistant Modestobacter marinus Strain BC501, a Representative Actinobacterium That Thrives on Calcareous Stone Surfaces.* Journal of Bacteriology, 2012. **194**(17): p. 4773-4774.
390. White, P.L., D.D. Wynn-Williams, and N.J. Russell, *Diversity of thermal responses of lipid composition in the membranes of the dominant culturable members of an Antarctic fellfield soil bacterial community.* Antarctic Science, 2000. **12**(3): p. 386-393.
391. Rathgeber, C., et al., *The photosynthetic apparatus and photoinduced electron transfer in the aerobic phototrophic bacteria Roseicyclus mahoneyensis and Porphyrobacter meromictius.* Photosynthesis Research, 2012. **110**(3): p. 193-203.
392. Nishimura, K., et al., *Expression of the puf Operon in an Aerobic Photosynthetic Bacterium, Roseobacter denitrificans.* Plant and Cell Physiology, 1996. **37**(2): p. 153-159.
393. Lami, R., et al., *Light-dependent growth and proteorhodopsin expression by Flavobacteria and SAR11 in experiments with Delaware coastal waters.* Environmental Microbiology, 2009. **11**(12): p. 3201-3209.
394. Woese, C.R. and G.E. Fox, *Phylogenetic structure of the prokaryotic domain: The primary kingdoms.* Proceedings of the National Academy of Sciences, 1977. **74**(11): p. 5088-5090.
395. Brown, C.T., et al., *Unusual biology across a group comprising more than 15% of domain Bacteria.* Nature, 2015. **523**(7559): p. 208-211.
396. Woese, C.R., *Bacterial evolution.* Microbiological Reviews, 1987. **51**(2): p. 221-271.
397. Marx, V., *Microbiology: the return of culture.* Nat Meth, 2017. **14**(1): p. 37-40.
398. Dickson, I., *Gut microbiota: Culturomics: illuminating microbial dark matter.* Nat Rev Gastroenterol Hepatol, 2017. **14**(1): p. 3-3.
399. Youssef, N.H., et al., *Partial Genome Assembly for a Candidate Division OP11 Single Cell from an Anoxic Spring (Zodletone Spring, Oklahoma).* Applied and Environmental Microbiology, 2011. **77**(21): p. 7804-7814.
400. de la Torre, J.R., et al., *Microbial Diversity of Cryptoendolithic Communities from the McMurdo Dry Valleys, Antarctica.* Applied and Environmental Microbiology, 2003. **69**(7): p. 3858-3867.
401. Zielińska, S., et al., *The Arctic Soil Bacterial Communities in the Vicinity of a Little Auk Colony.* Frontiers in Microbiology, 2016. **7**(1298).
402. Chen, Y., et al., *Mulching practices altered soil bacterial community structure and improved orchard productivity and apple quality after five growing seasons.* Scientia Horticulturae, 2014. **172**: p. 248-257.

403. Suen, G., et al., *An Insect Herbivore Microbiome with High Plant Biomass-Degrading Capacity.* PLoS Genetics, 2010. **6**(9): p. e1001129.
404. Costello, E.K., et al., *Fumarole-Supported Islands of Biodiversity within a Hyperarid, High-Elevation Landscape on Socompa Volcano, Puna de Atacama, Andes.* Applied and Environmental Microbiology, 2009. **75**(3): p. 735-747.
405. Delmotte, N., et al., *Community proteogenomics reveals insights into the physiology of phyllosphere bacteria.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(38): p. 16428-16433.
406. Dunbar, J., et al., *Empirical and Theoretical Bacterial Diversity in Four Arizona Soils.* Applied and Environmental Microbiology, 2002. **68**(6): p. 3035-3045.
407. Graff, A. and R. Conrad, *Impact of flooding on soil bacterial communities associated with poplar Populus sp. trees.* FEMS Microbiology Ecology, 2005. **53**(3): p. 401-415.
408. Wong, F.K.Y., et al., *Hypolithic Microbial Community of Quartz Pavement in the High-Altitude Tundra of Central Tibet.* Microbial Ecology, 2010. **60**(4): p. 730-739.
409. Walker, J.J. and N.R. Pace, *Phylogenetic Composition of Rocky Mountain Endolithic Microbial Ecosystems.* Applied and Environmental Microbiology, 2007. **73**(11): p. 3497-3504.
410. Lin, X., et al., *Vertical stratification of subsurface microbial community composition across geological formations at the Hanford Site.* Environmental Microbiology, 2012. **14**(2): p. 414-425.
411. Chong, C., et al., *Assessment of soil bacterial communities on Alexander Island (in the maritime and continental Antarctic transitional zone).* Polar Biology, 2012. **35**(3): p. 387-399.
412. Eichorst, S.A. and C.R. Kuske, *Identification of Cellulose-Responsive Bacterial and Fungal Communities in Geographically and Edaphically Different Soils by Using Stable Isotope Probing.* Applied and Environmental Microbiology, 2012. **78**(7): p. 2316-2327.
413. Zhou, J., et al., *Bacterial phylogenetic diversity and a novel candidate division of two humid region, sandy surface soils.* Soil Biology and Biochemistry, 2003. **35**(7): p. 915-924.
414. Lesaulnier, C., et al., *Elevated atmospheric CO2 affects soil microbial diversity associated with trembling aspen.* Environmental Microbiology, 2008. **10**(4): p. 926-941.
415. Marshall Hathaway, J.J., et al., *Comparison of Bacterial Diversity in Azorean and Hawai'ian Lava Cave Microbial Mats.* Geomicrobiology journal, 2014. **31**(3): p. 205-220.
416. Kim, H.M., et al., *Bacterial community structure and soil properties of a subarctic tundra soil in Council, Alaska.* FEMS Microbiology Ecology, 2014. **89**(2): p. 465-475.
417. Li, Q., et al., *Distribution and Diversity of Bacteria and Fungi Colonization in Stone Monuments Analyzed by High-Throughput Sequencing.* PLOS ONE, 2016. **11**(9): p. e0163287.
418. Vidal-Romaní, J.R., et al., *Bioweathering related to groundwater circulation in cavities of magmatic rock massifs.* Environmental Earth Sciences, 2015. **73**(6): p. 2997-3010.
419. Vuono, D.C., et al., *Disturbance opens recruitment sites for bacterial colonization in activated sludge.* Environmental Microbiology, 2016. **18**(1): p. 87-99.
420. Sun, Y., et al., *Microbial community analysis in biocathode microbial fuel cells packed with different materials.* AMB Express, 2012. **2**: p. 21-21.
421. Angenent, L.T., et al., *Molecular identification of potential pathogens in water and air of a hospital therapy pool.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(13): p. 4860-4865.
422. Baron, J.L., et al., *Shift in the Microbial Ecology of a Hospital Hot Water System following the Introduction of an On-Site Monochloramine Disinfection System.* PLOS ONE, 2014. **9**(7): p. e102679.
423. Sapp, M., et al., *Metabarcoding of Bacteria Associated with the Acute Oak Decline Syndrome in England.* Forests, 2016. **7**(5): p. 95.
424. Lu, S., et al., *Insights into the Structure and Metabolic Function of Microbes That Shape Pelagic Iron-Rich Aggregates ("Iron Snow").* Applied and Environmental Microbiology, 2013. **79**(14): p. 4272-4281.

425. Brodie, E.L., et al., *Urban aerosols harbor diverse and dynamic bacterial populations.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(1): p. 299-304.
426. Grice, E.A., et al., *Longitudinal shift in diabetic wound microbiota correlates with prolonged skin defense response.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(33): p. 14799-14804.
427. Kong, H.H., et al., *Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis.* Genome Research, 2012. **22**(5): p. 850-859.
428. Grice, E.A., et al., *Topographical and Temporal Diversity of the Human Skin Microbiome.* Science (New York, N.Y.), 2009. **324**(5931): p. 1190-1192.
429. Carlier, A., et al., *Draft genome and description of Orrella dioscoreae gen. nov. sp. nov., a new species of Alcaligenaceae isolated from leaf acumens of Dioscorea sansibarensis.* Systematic and Applied Microbiology, 2017. **40**(1): p. 11-21.
430. Markowitz, V.M., et al., *IMG ER: a system for microbial genome annotation expert review and curation.* Bioinformatics, 2009. **25**(17): p. 2271-2278.
431. Overbeek, R., et al., *The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST).* Nucleic Acids Research, 2014. **42**(Database issue): p. D206-D214.
432. Arndt, D., et al., *PHASTER: a better, faster version of the PHAST phage search tool.* Nucleic Acids Research, 2016. **44**(Web Server issue): p. W16-W21.
433. Grissa, I., G. Vergnaud, and C. Pourcel, *CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats.* Nucleic Acids Research, 2007. **35**(Web Server issue): p. W52-W57.
434. Chen, I.M.A., et al., *IMG/M: integrated genome and metagenome comparative data analysis system.* Nucleic Acids Research, 2016. **45**(D1): p. D507-D516.
435. MacFaddin, J.F., *Biochemical tests for identification of medical bacteria*. 2nd ed. ed. 1980, Baltimore (Md.): Williams & Wilkins Co. 527.
436. Spurr, A.R., *A low-viscosity epoxy resin embedding medium for electron microscopy.* Journal of Ultrastructure Research, 1969. **26**(1–2): p. 31-43.
437. Mergaert, J., L. Verdonck, and K. Kersters, *Transfer of Erwinia ananas (synonym, Erwinia uredovora) and Erwinia stewartii to the Genus Pantoea emend. as Pantoea ananas (Serrano 1928) comb. nov. and Pantoea stewartii (Smith 1898) comb. nov., Respectively, and Description of Pantoea stewartii subsp. indologenes subsp. nov.* International Journal of Systematic and Evolutionary Microbiology, 1993. **43**(1): p. 162-173.
438. Mohammadi, M., L. Burbank, and M.C. Roper, *Biological Role of Pigment Production for the Bacterial Phytopathogen Pantoea stewartii subsp. stewartii.* Applied and Environmental Microbiology, 2012. **78**(19): p. 6859-6865.
439. Henriques, M., A. Silva, and J. Rocha, *Extraction and quantification of pigments from a marine microalga: a simple and reproducible method*, in *Communicating Current Research and Educational Topics and Trends in Applied Microbiology*, A. Méndez-Vilas, Editor. 2007, FORMATEX: Spain.
440. Hug, L.A., et al., *A new view of the tree of life.* Nature Microbiology, 2016. **1**: p. 16048.
441. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea.* Nature reviews. Genetics, 2010. **11**(3): p. 181-190.
442. Maeder, C. and D.E. Draper, *A Small Protein Unique to Bacteria Organizes rRNA Tertiary Structure Over an Extensive Region of the 50 S Ribosomal Subunit.* Journal of Molecular Biology, 2005. **354**(2): p. 436-446.
443. Akanuma, G., et al., *Inactivation of Ribosomal Protein Genes in Bacillus subtilis Reveals Importance of Each Ribosomal Protein for Cell Proliferation and Cell Differentiation.* Journal of Bacteriology, 2012. **194**(22): p. 6282-6291.

444. Lecompte, O., et al., *Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale.* Nucleic Acids Research, 2002. **30**(24): p. 5382-5390.
445. Yarza, P., et al., *Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences.* Nat Rev Micro, 2014. **12**(9): p. 635-645.
446. Spring, S., et al., *Characterization of the first cultured representative of Verrucomicrobia subdivision 5 indicates the proposal of a novel phylum.* The ISME Journal, 2016. **10**(12): p. 2801-2816.
447. Squina, F.M. and A.Z. Mercadante, *Influence of nicotine and diphenylamine on the carotenoid composition of Rhodotorula strains.* Journal of Food Biochemistry, 2005. **29**(6): p. 638-652.
448. Ishikawa, S., et al., *A new FtsZ-interacting protein, YlmF, complements the activity of FtsA during progression of cell division in Bacillus subtilis.* Molecular Microbiology, 2006. **60**(6): p. 1364-1380.
449. Chastain, C.J., et al., *Functional evolution of C-4 pyruvate, orthophosphate dikinase.* Journal of Experimental Botany, 2011. **62**(9): p. 3083-3091.
450. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013.* Nucleic Acids Research, 2014. **42**(Database issue): p. D490-D495.
451. Roberts, A.W., E.M. Roberts, and C.H. Haigler, *Moss cell walls: structure and biosynthesis.* Frontiers in Plant Science, 2012. **3**: p. 166.
452. Nobles, D.R., D.K. Romanovicz, and R.M. Brown, *Cellulose in Cyanobacteria. Origin of Vascular Plant Cellulose Synthase?* Plant Physiology, 2001. **127**(2): p. 529-542.
453. Jonas, R. and L.F. Farah, *Production and application of microbial cellulose.* Polymer Degradation and Stability, 1998. **59**(1): p. 101-106.
454. Domozych, D.S., et al., *The Cell Walls of Green Algae: A Journey through Evolution and Diversity.* Frontiers in plant science, 2012. **3**: p. 82.
455. Zhao, C., et al., *High-yield production of extracellular type-I cellulose by the cyanobacterium Synechococcus sp. PCC 7002.* Cell Discovery, 2015. **1**: p. 15004.
456. Iriki, Y., et al., *Xylan of Siphonaceous Green Algae.* Nature, 1960. **187**(4731): p. 82-83.
457. Lembre, P., C.c. Lorentz, and P.D. Martino, *Exopolysaccharides of the Biofilm Matrix: A Complex Biophysical World*, in *The Complex World of Polysaccharides*. 2012, InTech: Rijeka. p. Ch. 0.
458. Michael, V., et al., *Biofilm plasmids with a rhamnose operon are widely distributed determinants of the 'swim-or-stick' lifestyle in roseobacters.* The ISME Journal, 2016. **10**(10): p. 2498-2513.
459. Biele, J., et al., *The landing(s) of Philae and inferences about comet surface mechanical properties.* Science, 2015. **349**(6247).
460. Convey, P., et al., *The spatial structure of Antarctic biodiversity.* Ecological Monographs, 2014. **84**(2): p. 203-244.
461. Laybourn-Parry, J. and D.A. Pearce, *The biodiversity and ecology of Antarctic lakes: models for evolution.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2007. **362**(1488): p. 2273-2289.
462. IPCC, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. 2013, Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. 1535.
463. Quayle, W.C., et al., *Extreme Responses to Climate Change in Antarctic Lakes.* Science, 2002. **295**(5555): p. 645-645.
464. Lyons, W.B., et al., *Antarctic Lake Systems and Climate Change*, in *Trends in Antarctic Terrestrial and Limnetic Ecosystems: Antarctica as a Global Indicator*, D.M. Bergstrom, P. Convey, and A.H.L. Huiskes, Editors. 2006, Springer Netherlands: Dordrecht. p. 273-295.
465. Nydahl, A.C., et al., *Toxicity of fuel-contaminated soil to Antarctic moss and terrestrial algae.* Environmental Toxicology and Chemistry, 2015. **34**(9): p. 2004-2012.

466. Rayner, J.L., et al., *Petroleum-hydrocarbon contamination and remediation by microbioventing at sub-Antarctic Macquarie Island.* Cold Regions Science and Technology, 2007. **48**(2): p. 139-153.
467. Frenot, Y., et al., *Biological invasions in the Antarctic: extent, impacts and implications.* Biological Reviews, 2005. **80**(1): p. 45-72.
468. Cowan, D.A., et al., *Non-indigenous microorganisms in the Antarctic: assessing the risks.* Trends in Microbiology, 2011. **19**(11): p. 540-548.
469. Sanford, R.A., et al., *Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(48): p. 19709-19714.
470. Fredriksson, N.J., M. Hermansson, and B.-M. Wilén, *The Choice of PCR Primers Has Great Impact on Assessments of Bacterial Community Diversity and Dynamics in a Wastewater Treatment Plant.* PLOS ONE, 2013. **8**(10): p. e76431.
471. Tytgat, B., *Distribution and characterization of bacterial communities in diverse Antarctic ecosystems by high-troughput sequencing*, in *Department of Biology, Department of Biochemistry and microbiology*. 2016, Ghent University: Ghent, Belgium.
472. Vos, M., et al., *A Comparison of rpoB and 16S rRNA as Markers in Pyrosequencing Studies of Bacterial Diversity.* PLOS ONE, 2012. **7**(2): p. e30600.
473. Sharpton, T.J., *An introduction to the analysis of shotgun metagenomic data.* Frontiers in Plant Science, 2014. **5**: p. 209.
474. Sharpton, T.J., et al., *PhylOTU: A High-Throughput Procedure Quantifies Microbial Community Diversity and Resolves Novel Taxa from Metagenomic Data.* PLoS Computational Biology, 2011. **7**(1): p. e1001061.
475. Schloss, P.D. and J. Handelsman, *A statistical toolbox for metagenomics: assessing functional diversity in microbial communities.* BMC Bioinformatics, 2008. **9**: p. 34-34.
476. Mavromatis, K., et al., *Use of simulated data sets to evaluate the fidelity of metagenomic processing methods.* Nat Meth, 2007. **4**(6): p. 495-500.
477. Mende, D.R., et al., *Assessment of Metagenomic Assembly Using Simulated Next Generation Sequencing Data.* PLOS ONE, 2012. **7**(2): p. e31386.
478. Degnan, P.H. and H. Ochman, *Illumina-based analysis of microbial community diversity.* Isme Journal, 2012. **6**(1): p. 183-194.
479. Kunin, V., et al., *A Bioinformatician's Guide to Metagenomics.* Microbiology and Molecular Biology Reviews, 2008. **72**(4): p. 557-578.
480. Wang, Q., et al., *Xander: employing a novel method for efficient gene-targeted metagenomic assembly.* Microbiome, 2015. **3**: p. 32.
481. Oulas, A., et al., *Metagenomics: Tools and Insights for Analyzing Next-Generation Sequencing Data Derived from Biodiversity Studies.* Bioinformatics and Biology Insights, 2015. **9**: p. 75-88.
482. Orellana, L.H., et al., *Detecting Nitrous Oxide Reductase (nosZ) Genes in Soil Metagenomes: Method Development and Implications for the Nitrogen Cycle.* Mbio, 2014. **5**(3).
483. Handelsman, J., *Metagenomics: Application of Genomics to Uncultured Microorganisms.* Microbiology and Molecular Biology Reviews, 2004. **68**(4): p. 669-685.
484. Frank, J.A., et al., *Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data.* Scientific Reports, 2016. **6**: p. 25373.
485. Sharma, S. and J. Vakhlu, *Metagenomics as advanced screening methods for novel microbial metabolites*, in *Microbial Biotechnology: Progress and Trends*, F.D. Harzevili and H. Chen, Editors. 2014, CRC Press. p. 379.
486. Lou, D.I., et al., *High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(49): p. 19872-19877.
487. Victoria Wang, X., et al., *Estimation of sequencing error rates in short reads.* BMC Bioinformatics, 2012. **13**(1): p. 185.

488. He, Y., et al., *Stability of operational taxonomic units: an important but neglected property for analyzing microbial diversity.* Microbiome, 2015. **3**(1): p. 20.
489. Powell, J.R. and B.A. Sikes, *Method or madness: does OTU delineation bias our perceptions of fungal ecology?* New Phytologist, 2014. **202**(4): p. 1095-1097.
490. Schirmer, M., et al., *Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data.* BMC Bioinformatics, 2016. **17**(1): p. 125.
491. Schirmer, M., et al., *Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform.* Nucleic Acids Research, 2015. **43**(6): p. e37-e37.
492. Edgar, R.C., *UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing.* bioRxiv, 2016.
493. He, Z., et al., *Estimating DNA polymorphism from next generation sequencing data with high error rate by dual sequencing applications.* BMC Genomics, 2013. **14**(1): p. 535.
494. Nguyen, N.H., et al., *Parsing ecological signal from noise in next generation amplicon sequencing.* New Phytologist, 2015. **205**(4): p. 1389-1393.
495. Shade, A., et al., *Culturing captures members of the soil rare biosphere.* Environmental Microbiology, 2012. **14**(9): p. 2247-2252.
496. Bodelier, P.L.E., et al., *Microbial minorities modulate methane consumption through niche partitioning.* ISME J, 2013. **7**(11): p. 2214-2228.
497. Sjöstedt, J., et al., *Recruitment of Members from the Rare Biosphere of Marine Bacterioplankton Communities after an Environmental Disturbance.* Applied and Environmental Microbiology, 2012. **78**(5): p. 1361-1369.
498. Caporaso, J.G., et al., *The Western English Channel contains a persistent microbial seed bank.* ISME J, 2012. **6**(6): p. 1089-1093.
499. Mysara, M., et al., *IPED: a highly efficient denoising tool for Illumina MiSeq Paired-end 16S rRNA gene amplicon sequencing data.* BMC Bioinformatics, 2016. **17**: p. 192.
500. Schmidt, T.S.B., J.F. Matias Rodrigues, and C. von Mering, *Limits to robustness and reproducibility in the demarcation of operational taxonomic units.* Environmental Microbiology, 2015. **17**(5): p. 1689-1706.
501. Feng, Y., et al., *A phototrophy-driven microbial food web in a rice soil.* Journal of Soils and Sediments, 2011d. **11**(2): p. 301-311.
502. Gollner, S., et al., *Diversity of Meiofauna from the 9° 50′ N East Pacific Rise across a Gradient of Hydrothermal Fluid Emissions.* PLOS ONE, 2010. **5**(8): p. e12321.
503. Lagier, J.-C., et al., *Culture of previously uncultured members of the human gut microbiota by culturomics.* Nature Microbiology, 2016. **1**: p. 16203.
504. Tandina, F., et al., *Use of MALDI-TOF MS and culturomics to identify mosquitoes and their midgut microbiota.* Parasites & Vectors, 2016. **9**(1): p. 495.
505. Cadieux, B., et al., *Role of Lipase from Community-Associated Methicillin-Resistant Staphylococcus aureus Strain USA300 in Hydrolyzing Triglycerides into Growth-Inhibitory Free Fatty Acids.* Journal of Bacteriology, 2014. **196**(23): p. 4044-4056.
506. Soora, M. and H. Cypionka, *Light Enhances Survival of Dinoroseobacter shibae during Long-Term Starvation.* PLOS ONE, 2014. **8**(12): p. e83960.
507. Connon, S.A. and S.J. Giovannoni, *High-Throughput Methods for Culturing Microorganisms in Very-Low-Nutrient Media Yield Diverse New Marine Isolates.* Applied and Environmental Microbiology, 2002. **68**(8): p. 3878-3885.
508. Davis, K.E.R., S.J. Joseph, and P.H. Janssen, *Effects of Growth Medium, Inoculum Size, and Incubation Time on Culturability and Isolation of Soil Bacteria.* Applied and Environmental Microbiology, 2005. **71**(2): p. 826-834.
509. Janssen, P.H., et al., *Improved Culturability of Soil Bacteria and Isolation in Pure Culture of Novel Members of the Divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia.* Applied and Environmental Microbiology, 2002. **68**(5): p. 2391-2396.
510. Rappe, M.S., et al., *Cultivation of the ubiquitous SAR11 marine bacterioplankton clade.* Nature, 2002. **418**(6898): p. 630-633.

511. Okubo, T., et al., *Complete Genome Sequence of Bradyrhizobium sp. S23321: Insights into Symbiosis Evolution in Soil Oligotrophs.* Microbes and Environments, 2012. **27**(3): p. 306-315.
512. King, G.M. and C.F. Weber, *Distribution, diversity and ecology of aerobic CO-oxidizing bacteria.* Nature Reviews Microbiology, 2007. **5**(2): p. 107-118.
513. Pearce, D.A., et al., *Metagenomic analysis of a southern maritime Antarctic soil.* Frontiers in Microbiology, 2012. **3**.
514. Pearce, D., et al., *Preliminary Analysis of Life within a Former Subglacial Lake Sediment in Antarctica.* Diversity, 2013. **5**(3): p. 680.
515. Soo, R.M., et al., *Microbial biodiversity of thermophilic communities in hot mineral soils of Tramway Ridge, Mount Erebus, Antarctica.* Environmental Microbiology, 2009. **11**(3): p. 715-728.
516. Lee, K.C.Y., et al., *Genomic analysis of Chthonomonas calidirosea, the first sequenced isolate of the phylum Armatimonadetes.* The ISME Journal, 2014. **8**(7): p. 1522-1533.
517. Kelly, D.P. and J.C. Murrell, *Microbial metabolism of methanesulfonic acid.* Archives of Microbiology, 1999. **172**(6): p. 341-348.
518. Moosvi, S.A., et al., *Isolation and properties of methanesulfonate-degrading Afipia felis from Antarctica and comparison with other strains of A. felis.* Environmental Microbiology, 2005. **7**(1): p. 22-33.
519. Nichols, D., et al., *Use of Ichip for High-Throughput In Situ Cultivation of "Uncultivable" Microbial Species.* Applied and Environmental Microbiology, 2010. **76**(8): p. 2445-2450.
520. Gross, A., et al., *Technologies for Single-Cell Isolation.* International Journal of Molecular Sciences, 2015. **16**(8): p. 16897-16919.
521. Giovannoni, S. and U. Stingl, *The importance of culturing bacterioplankton in the 'omics' age.* Nature Reviews Microbiology, 2007. **5**(10): p. 820-826.
522. Tuorto, S.J., et al., *Bacterial genome replication at subzero temperatures in permafrost.* The ISME Journal, 2014. **8**(1): p. 139-149.
523. Panikov, N.S., et al., *Microbial activity in soils frozen to below −39 °C.* Soil Biology and Biochemistry, 2006. **38**(4): p. 785-794.
524. Junge, K., et al., *Bacterial incorporation of leucine into protein down to −20 °C with evidence for potential activity in sub-eutectic saline ice formations.* Cryobiology, 2006. **52**(3): p. 417-429.
525. Berg, I.A., *Ecological Aspects of the Distribution of Different Autotrophic CO(2) Fixation Pathways.* Applied and Environmental Microbiology, 2011. **77**(6): p. 1925-1936.
526. Murata, N., et al., *Photoinhibition of photosystem II under environmental stress.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 2007. **1767**(6): p. 414-421.

Curriculum vitae

Personalia

Guillaume Tahon
Gentstraat 77 B4, 9800 Deinze, Belgium
Born on May 23rd 1987, Veurne, Belgium
Guillaume.Tahon@telenet.be
+32 (0) 472.33.51.45

Educational background

2012-Present	PhD – LM-UGent, Faculty of Sciences, Ghent University Dissertation: The role of photoheterotrophic and chemoautotrophic prokaryotes in the microbial food web in terrestrial Antarctica: a cultivation approach combined with functional analysis. Promotor: Prof. Dr. Anne Willems
2006-2012	Hogeschool Gent, Ghent, Belgium Degree of Master of Science in Industrial Engineering Dissertation: Diversity of Lactic Acid Bacteria Associated with the Production of *Dua Muoi*, Vietnamese Fermented Vegetables. Promotor: Prof. Dr. Peter Vandamme
2005-2006	Ghent University, Ghent, Belgium 1st bachelor Bioengineering
1999-2005	Sint-Aloysiuscollege, Diksmuide, Belgium Degree Secondary Education Sciences-Mathematics

Language proficiencies

Dutch – native speaker
French – excellent command
English – excellent command
German – basic communication skills

Scientific output

A1 publications

Tahon G., Tytgat B., Lebbe L., Carlier A. and Willems A. (2017) *Abditibacterium utsteinense* sp. nov. the first cultivated representative of the bacterial candidate phylum FBP. In preparation.

Tahon G. and Willems A. (2017) Isolation and characterization of aerobic anoxygenic phototrophs from exposed soils from the Sør Rondane Mountains, East Antarctica. Systematic and Applied Microbiology, under review.

Tahon G., Tytgat B. and Willems A. (2017) Diversity of key genes for primary production and diazotrophy in soils from the Sør Rondane Mountains, East-Antarctica, revealed by amplicon sequencing. BMC Microbiome, under review.

Tahon G., Tytgat B. and Willems A. (2016) Diversity of Phototrophic Genes Suggests Multiple Bacteria May Be Able to Exploit Sunlight in Exposed Soils from the Sør Rondane Mountains, East Antarctica. Frontiers in Microbiology 7: 2026, doi: 10.3389/fmicb.2016.02026

Tahon G., Tytgat B., Stragier P. and Willems A. (2016) Analysis of *cbbL*, *nifH*, and *pufLM* in Soils from the Sør Rondane Mountains, Antarctica, Reveals a Large Diversity of Autotrophic and Phototrophic Bacteria". Microbial Ecology 2016;71: 131-49, doi: 10.1007/s00248-015-0704-6

Oral Presentations

Tahon G., Tytgat B. and Willems A. (2017) Diversity of phototrophic bacteria in terrestrial samples from the Sør Rondane Mountains, East Antarctica. XIIth SCAR Biology Symposium, July 10-14, 2017, Leuven, Belgium.

Tahon G., Tytgat B., Peeters K. and Willems A. (2016) Shining a light on exposed high-altitude Antarctic ecosystems provides a clearer view on the diversity of phototrophic bacteria. BNCGG-BNCAR Antarctica Symposium, April 29th, 2016, Brussels, Belgium.

Tahon G., Tytgat B. and Willems A. (2015) Diversity of *cbbL*, *nifH* and *pufLM* genes in soils around the Princess Elizabeth Station, Sør Rondane Mountains, Antarctica. The 6th International Conference on Polar and Alpine Microbiology, September 6-10, 2015, České Budějovice, Czech Republic.

Poster presentations

Tahon G., Tytgat B. and Willems A. (2015) Life on the frozen continent: diversity of RuBisCO, *nifH* and *pufLM* genes in soils around the Princess Elisabeth Station, Sør Rondane Mountains, Antarctica. The 6th Congress of European Microbiologists, June 7-11, 2015, Maastricht, The Netherlands.

Tahon G., Stragier P. and Willems A. (2013) Novel insights in RuBisCO genes of terrestrial samples from Antarctica. BSM symposium - Microbial Diversity for Science and Industry, November 26-27, 2013, Brussels, Belgium.

Tahon G., Stragier P., Becker A. and Willems A. (2012) Characterization of prokaryotes in environmental samples from terrestrial Antarctica. BSM symposium - Posttranscriptional regulation and epigenetics in microorganisms, November 30th, 2012, Brussels, Belgium.

Specialist courses (all at UGent)

2014	Advanced Academic English: Effective Slide Design
2013	Microbial Evolution and Diversity Personal Effectiveness
2012	Molecular Microbial Ecology Microbial Genomics Ecological Applications of Biomarkers in Aquatic Food Web Studies Creative Thinking

Awarded grants and prizes

2015	Second best oral presentation of early career scientists at the 6th International Conference on Polar and Alpine Microbiology, September 6-10, 2015, České Budějovice, Czech Republic. Title: Diversity of *cbbL*, *nifH* and *pufLM* genes in soils around the Princess Elizabeth Station, Sør Rondane Mountains, Antarctica. FEMS, Federation of European Microbiological Society FEMS meeting attendance grant for the 6th Congress of European Microbiologists, June 7-11, 2015, Maastricht, The Netherlands.
2014	SCAR, Scientific Committee on Antarctic Research SCAR meeting attendance grant for the AnT-ERA workshop on "Molecular and genetic advances to understanding evolution and biodiversity in the polar regions - The legacy of EBA", October 2-3, 2014, Naples, Italy.