Functional diversity of microbial communities in the McMurdo Dry Valleys, Antarctica

Ting-Shyang (Sean) Wei

PhD

2016

Functional diversity of microbial communities in the McMurdo Dry Valleys, Antarctica

Ting-Shyang (Sean) Wei

A thesis submitted to

Auckland University of Technology

in fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

2016

School of Science

Auckland University of Technology

Table of Contents

Chapter 1 General introduction: global distribution, features and colonization of dryland environments 1
1.1 Definition and global distribution drylands 1
1.2 Life forms in dryland environments 4
1.3 Microbial life in drylands 5
1.3.1 Biological soil crusts 5
1.3.2 Epilithic colonization 6
1.3.3 Hypoliths 7
1.3.4 Cryptoendoliths and chasmoencoliths 9
1.4 Life under extreme aridity: hyper-arid deserts 10
1.4.1 Hyper-arid environments and stressors 10
1.4.2 Cryptic habitats 14
1.4.3 Microbial diversity in McMurdo Dry Valleys 15
1.4.4 Microbial functional roles in McMurdo Dry Valleys 19
1.5 Research gap in microbial ecology of Antarctica's McMurdo Dry Valleys 20
1.5.1 Microbial carbon and nitrogen cycling pathways 20
1.5.2 Stress response related to the Dry Valleys environments 22
1.6 The GeoChip DNA functional microarray 24
1.7 Research objectives 25

Chapter 2 Diversity of metabolic and stress-tolerance pathways in chasmo-endolithic and soil communities of the McMurdo Dry Valleys 35
2.1 Introduction 35
2.2 Material and methods 37
2.3 Results 38
2.3.1 GeoChip analysis and functional diversity 38
2.3.2 Biodiversity 39
2.3.3 Carbon utilization 40
2.3.4 Nitrogen utilization 40
2.3.5 Stress response 41
2.4 Discussion 41

Chapter 3 Diversity of metabolic and stress-tolerance pathways in hypolithic and soil communities of the McMurdo Dry Valleys 53
3.1 Introduction 53
3.2 Material and methods 54
3.3 Results 55
3.3.1 Carbon transformation pathways 55
3.3.2 Nitrogen transformation pathways 56
3.3.3 Stress response pathways 57
3.4 Discussion 58

Chapter 4 Potential biotic controls on endolithic, hypolithic and soil communities in the McMurdo Dry Valleys 68
4.1 Introduction 69
4.2 Material and methods 71
4.3 Results 72
4.3.1 Antibiotic resistance genes 72
4.3.2 bacteriophage diversity 73
4.4 Discussion 73

Chapter 5 Identification of molecular markers for genes encoding extracellular polymeric substance secretion pathways in desert cyanobacteria 86
5.1 Introduction 86
5.2 Material and methods 92
5.2.1 Cyanobacteria isolation and establishment of enriched cultures 92
5.2.2 Purification of cyanobacteria cultures 93
5.2.3 DNA extraction for cyanobacterial cell 94
5.2.4 Assessment of cyanobacteria culture composition and purity 94
5.2.4.1 Illumina 16S library preparation and sequencing 94
5.2.4.2 16S rRNA sequence analysis 96
5.2.5 Cyanobacteria metagenome analysis 96
5.2.5.1 Metagenome library preparation and sequencing 96
5.2.5.2 Cyanobacterial genome assembly and annotation 97
5.2.5.3 Identification of putative homologues of WZY-dependent pathway 98
5.2.3 Phylogenetic analysis of putative EPS homologues 98
5.3 Results 99
5.3.1 Cyanobacteria cultivation and purification 101
5.3.2 *De novo* metagenome assembly 99
5.3.3 Phylogeny analysis of putative WZY-dependent pathway homologues 101
5.4 Discussion 104

Chapter 6 Thesis conclusions 128

References 136

Appendices 160

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgement), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signed _____ Date_____

Acknowledgements

First of all, I would like to express my sincere gratitude to my primary supervisor Professor Stephen Pointing. Thank you for your continued support of my Ph.D studies and especially for your knowledgeable and flexible supervision. It has been pleasure working with you these past four years since I first started in your research group in Hong Kong University.

I also like to thank my secondary supervisor, Dr. Colleen Higgins. Thank you for your absolute patience and encouragement in addition to your bioinformatics expertise.

Besides my supervisors, I truly appreciate my thesis examiners: Prof. David Hopkins, Prof. Barbara Breen and Dr. Kimberly Rhode. Your insightful comments greatly strengthened the quality of this thesis.

My sincere appreciation also goes to our team in Institute of Applied Ecology New Zealand (AENZ) in Auckland University of Technology, generosity in experimental supports from Dr. Stephen Archer, Dr. Kevin Lee, Dr. Donna Lacap and Ms. Rachel Boyle. And thank you to the AUT staff, particularly Dr. Michelle Yoo, Dr. Brent Seale and Dr. Nicola Brasch for your never-ending encouragement.

Thanks for your company, my friends, Ruth, Jordan, Tom, Pete, Tim, and Jarrod.

And finally, thank you to my family for your endless support, especially my Dad who passed away before my Ph.D study. Without you, the path to my Ph.D would have been harder. Thank you and love you.

List of figures

Fig 1.1 Distribution of drylands under the UNEP scheme of desert classification 2
Fig 1.2 Different types of SRSCs communities 30
Fig 1.3 Relative abundance of recoverable bacterial phyla in hypolithic communities .. 31
Fig 1.4 Coloured bands of endolithic communities in a fractured sandstone 32
Fig 1.5 The landscape of the McMurdo Dry Valleys 33
Fig 1.6 The map of the McMurdo Dry Valleys 34
Fig. 2.1 Biodiversity estimates for chasmoendolith and soil communities using DNA gyrase subunit B gene (*gyrB*) as a phylogenetic marker 48
Fig 2.2 Taxa-function relationships for carbon cycling genes 49
Fig 2.3 Taxa-function relationships for nitrogen cycling genes 50
Fig 2.4 Taxa-function relationships for stress response genes 51
Fig 3.1 Distribution of pathways for carbon transformation 62
Fig 3.2 Overall abundance of different groups of carbon cycling genes 63
Fig 3.3 Distribution of pathways for nitrogen transformation 64
Fig 3.4 Overall abundance of different groups of nitrogen cycling genes 65
Fig 3.5 Distribution of pathways for stress response 66
Fig 3.6 Overall abundance of different groups of stress response genes 67
Fig S 3.1 Relative abundance of bacterial phyla in soil and hypolith communities 68
Fig 4.1 Categories antibiotic resistance genes recovered from Dry Valleys 80
Fig 4.2 Occurrence of antibiotic resistance genes from Dry Valleys 82
Fig 4.3 Categories of phage genomes recovered from Dry Valleys 83
Fig 4.4 Phage families recovered from specific substrates in Dry Valleys 84
Fig 5.1 The putative Wzy-dependent synthesis pathway 120
Fig 5.2 Morphology of cyanobacteria isolates from deserts 121
Fig 5.3 Flow cytometric analysis of the Dry Valley culture CBM 122
Fig 5.4 Maximum likelihood tree of Wzb homologues 123
Fig 5.5 Maximum likelihood tree of Wza homologues 124
Fig 5.6 Maximum likelihood tree of Wzx homologues 125
Fig 5.7 Maximum likelihood tree of Wzy homologues 126
Fig 5.8 Maximum likelihood tree of Wzc homologues 127
Fig 6.1 Diagram to illustrate overall carbon cycling in hypolithic, chasmoendolithic and soil communities in Miers Valley 133
Fig 6.2 Diagram to illustrate overall nitrogen cycling in hypolithic, chasmoendolithic and soil communities in Miers Valley in Miers Valley 134
Fig 6.3 Diagram to illustrate contigs identified as putative protein homologues involved in WZY-dependent and ABC transporter pathway 135

List of tables

Table 1.1 Subtypes of deserts classified by aridity index (AI)..... 2
Table 1.2 NDVI values for different cover types..... 3
Table 1.3 The features of hyper-arid desert: Atacama and Dry Valleys..... 28
Table 1.4 The major bacterial assemblage of niche communities in Dry Valleys..... 29
Table 2.1 Diversity metrics for chasmoendolith and soil communities..... 47
Table S2.1 Two-letter codes used to denote taxa in functional plots..... 50
Table 4.1 Functional loci targeting antibiotic resistance in the GeoChip..... 78
Table 4.2 Functional loci targeting phage in the GeoChip..... 79
Table 5.1 The data of sample site and growth conditions of cyanobacteria culture..... 111
Table 5.2 Pfam domains involved in Wzy-dependent pathway selected for analysis... 112
Table 5.3 The sample source of unicellular cyanobacteria strains..... 113
Table 5.4 The enriched culture composition and purification efficiency..... 117
Table 5.5 The assembly statistics of cyanobacteria cultures using SPAdes..... 118
Table 5.6 The assembly statistics of cyanobacteria cultures using MEGAHIT..... 119

List of acronyms

AI	Aridity index
ANAMMOX	Anaerobic ammonia oxidization
AOA	Ammonia-oxidizing archaea
AOB	Ammonia-oxidizing bacteria
BSCs	Biological soil crusts
DNAR	Dissimilatory nitrate reduction
EPS	Extracellular polymer substance
IM	Indices of moisture
MAP	Mean annual precipitation
MAT	Mean annual temperature
NGS	Next-generation sequencing
NVDI	Normalized difference vegetation index
OTUs	Operational taxonomic units
P	Annual precipitation
PET	Potential evapotranspiration
RH	Relative humidity
RPS	Released polysaccharides
RuBisCO	Ribulose-1, 5-bisphosphate carboxylase/oxygenase
SRSCs	Soil-and rock-surface communities
UNEP	United nation environmental programme

Abstract

In dryland ecosystems species richness for all domains decreases with increasing aridity. Several environmental stressors (desiccation, thermal and radiation stress) in hyper-arid drylands limit the complex life forms, thus microorganisms comprise most of the standing biomass and diversity in this particular ecosystem. Interestingly, these microorganisms colonize cryptic habitats that provide shelter from harsh environmental conditions. They form unique hypolithic, cryptoendolithic and chasmoendolithic communities beneath and within rocks and soil.

The microbial diversity and community structure in refugia of the McMurdo Dry Valleys, the coldest and driest dryland on Earth, has been extensively studied using 16S rRNA gene surveys. These have revealed photoautotrophic cyanobacteria dominate hypoliths and endoliths whereas Actinobacteria dominate soil communities. Despite this, very little is known about the functionality of these communities and how they respond to environmental stress. In the first part of this study, the GeoChip DNA microarray was used to interrogate carbon and nitrogen transformation pathways of chasmoendoliths, hypoliths and soil communities in a maritime-influenced location, Miers Valley. The chasmoendoliths, and hypoliths were identified as the potential primary production sites since cyanobacterial *rubisco* signatures were not commonly recovered from soils. Other forms of *rubisco* originated from Proteobacteria and archaea were identified, suggesting that chemoautotrophic pathways also contributed to carbon fixation. All communities supported diverse carbohydrate transformation pathways. However, soil communities supported significantly greater aromatic carbon utilization genes than hypoliths and chasmoendoliths, and this was related to the recalcitrant 'legacy' carbon stored in Antarctic soils.

For nitrogen fixation, all communities displayed the full suite of genes involved in nitrogen transformations. Soil communities generally supported slightly higher

abundance of proteobacterial nitrogenase than chasmoendoliths and hypoliths, and so soil may be more important than rock as a site for microbial nitrogen fixation. Stress response pathways were also identified, and soil communities displayed higher diversity and abundance of environmental stress response pathways compared to more cryptic rock communities. The high throughput 16S rRNA gene sequencing showed Cyanobacteria (46%), Actinobacteria (31%), Proteobacteria (25%) and were relatively dominant in hypolithic and soil communities, and these phyla also displayed the greatest functional diversity. This suggests that the metabolic plasticity for these taxa may be an important factor in their role as keystone species in these communities.

In addition, GeoChip analyses also identified possible sources of community regulation in the Dry Valleys where grazing and predation are minimal or non-existent. Widespread occurrence of antibiotic resistance genes indicated that inter-specific competition may be important among these diverse communities, and phage signatures indicated a potential bottom-up control via phage lysis of microbial cells.

In a metagenomic study, protein sequences involved in extracellular polymer substances (EPS) synthesis (Wzy-dependent pathway) of dryland cyanobacteria were identified by *de novo* sequencing on the Miseq platform, since EPS played important roles in bio-weathering, nutrient repositories and stress response. Together with other cyanobacterial EPS synthesis protein sequences, phylogenetic analyses showed that these proteins derived different environments were not correlated to their habitats, indicating these genes were highly conserved and that EPS production might be induced and regulated under specific conditions and mechanisms.

Chapter 1: Literature review: global distribution, features and colonization of dryland environments

1.1 Definition and global distribution of drylands

Dryland, desert and arid zone are the interchangeable terms used to describe a region that experiences prolonged moisture deficiency (Thomas, 2011). Drylands occur on all continents and comprise nearly half (up to 6 billions hectare, 41.3%) of the global landmass (Safriel et al., 2005). The occurrence of aridity can be traced back 1.8 billion years by records of dune sediments preserved in solid rocks (Glennie, 1987). Defining a dryland is complicated as many aspects must be considered, for example, climate, geomorphology, hydrology, vegetation and fauna. Several studies have indicated that aridity or humidity are critical contributors to control biological colonization of the dryland biome (Bugmann & Solomon, 2000; Olson et al., 2001; Pointing, Warren-Rhodes, Lacap, Rhodes, & McKay, 2007; Woodward, Lomas, & Kelly, 2004). Meigs (1953) classified arid environments globally (except the polar regions) into three types by using indices of moisture availability (IM): hyper-arid deserts (at least 12 consecutive months without rainfall); arid deserts (average annual rainfall between 25 to 200 mm) and semiarid lands (average annual rainfall between 250 and 500mm). In 1992 The United Nation Environmental Programme (UNEP) used meterological data from 2000 stations over a fixed time period from 1951-1980 to define dryland as an area with an aridity index (AI) less than 0.65. AI is based on the ratio of annual precipitation from rainfall (P) to potential evapotranspiration (PET) from surfaces and plants (UNEP, 1992). Furthermore, different subtypes of dryland can be categorized by different ranges of AI (Fig 1.1 & Table 1.1). For example, the Atacama Desert is classified as hyper-arid desert (AI <0.05) because the precipitation was less than 10 mm per year but the potential evapotranspiration reached up to 1500 mm per year (McKay

et al., 2004; Risacher et al., 2003). Another delineation is made based upon long-term macroclimate data, and is termed the Köppen climate classification. To determine whether a location has a desert climate, the mean annual precipitation (MAP) and precipitation threshold are considered. Different boundaries of mean annual temperature (MAT) divide desert climate into hot desert (MAT ≥ 18°C) and cold desert climate (MAT <18°C) (Peel, Finlayson, & McMahon, 2007).

Table 1.1 Subtypes of deserts classified by aridity index (AI).

Subtypes	Aridity Index (P/PET)	Share of Global* (%)
Hyper-arid	<0.05	6.6
Arid	0.05-0.20	10.6
Semiarid	0.20-0.50	15.2
Dry subhumid	0.50-0.65	8.7

* Data adapted from Safriel et al (2005)

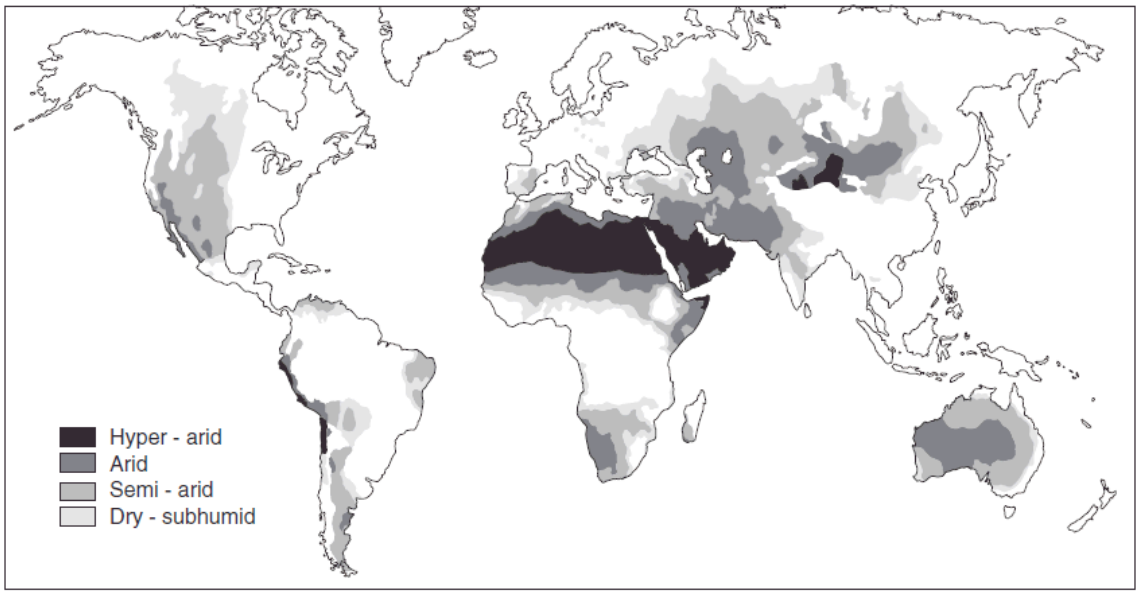


Fig 1.1 Distribution of drylands under the UNEP scheme of desert classification (UNEP, 1992)

In addition to climatic classifications, a vegetation index is another delineation to classify arid environments. From an ecological perspective, the primary producers in

drylands can be categorized into three vegetation groups. A) biological soil crusts: composed of cyanobacteria, algae, lichens, and mosses that are photosynthetically active once the water and light are available; B) annuals: plants that are not active during the dry seasons but blooming and sprouting after encountering seasonal or annual rainfall; C) perennials: plants that reduce transpiration surface and develop deep roots to allow them to survive during the dry seasons (Dall'Olmo & Karnieli, 2010). The normalized difference vegetation index (NVDI) is obtained from a global scale in real time from satellite remote sensing images, quantifying the photosynthetic capacity of vegetation (Wessels, Prince, Frost, & van Zyl, 2004). The index value is between -1 to 1 (Table 1.2). Water typically has a value less than 0; bare is soil between 0 to 0.1, and vegetation is over 0.1 (Lillesand & Kiefer, 1994). UNEP (2006) further defines drylands as geographical regions with high aridity (AI < 0.2) and composed of large areas of bare soil with low vegetation cover.

Table 1.2 NDVI values for different cover types

	Dense vegetation	Dry bare soil	Clouds	Snow and ice	Water
NDVI	0.7	0.025	0.002	-0.046	-0.256

Global distribution of hot arid deserts mainly lies in two belts 25° to 35° north and south of the Equator. This pattern of distribution correlates with the causes of aridity, which result from climatic, topographic and oceanographic factors that prevent moisture and precipitation. Tropical high-pressure belts carry concentrated zones of descending stable air, which brings consistently little and unstable precipitation to these areas. In central Asia, due to the large continent, moisture-bearing winds from distant ocean hardly travel into the central continent. Other topographic effects, such as the rain shadow of mountain barriers, could enhance these causative factors for aridity. This

effect can be observed in coastal South America, western North America by the Rocky Mountains and Australia by the north-south orientation of the Great Dividing Range; cold ocean currents cause low sea-surface evaporation, high atmospheric humidity and low precipitation that contribute to the western coastal deserts in South America, South Africa and Australia (Thomas, 2011).

The above classification system does not include cold deserts in polar latitudes. Although polar climate was defined in the Köppen system (temperature of the hottest month is less than 18°C), the precipitation data was missing due to lack of precipitation stations at inland Antarctica and certain regions within the High Artic circle (Peel et al., 2007). Polar deserts comprise 5 x 10^6 km^2 (approximately 3.4 %) of the Earth's surface, covering a substantial part of High Arctic and ice-free regions of Antarctic continent (Cockell & Stokes, 2006). The McMurdo Dry Valleys are a largely ice-free polar desert located in Antarctica continent. Apart from low precipitation, environmental stress such as extreme temperature fluctuation, high incidence of solar radiation (Doran, McKay, et al., 2002) and oligotrophy (low organic carbon source) (Cowan & Tow, 2004; Wynn-Williams, 1990) are the main challenges for life.

1.2 Life in dryland environments

The availability of water is the main factor that constrains distribution and diversity of life. Species richness decreases as aridity declines (Gaston, 2000). For example, the species richness of mammals in the hyper-arid region of the Sahara and Atacama deserts was roughly five times lower than Amazonian rain forests, and in polar region mammalian species richness was approximately zero (Olson et al., 2001). The species richness of vascular plants also presented similar patterns (Kier et al., 2005). In dry subhumid drylands, the major biome is forest whereas in semiarid drylands it is grassland (Safriel et al., 2005). One of the most distinct features of semi-arid dryland

that influences plant growth is the seasonal or sparse rainfall (Schmidt & Karnieli, 2000). Most of the deserts in North America are semi-arid, receiving rainfall during winter and spring from the Pacific Ocean. The vegetation is primary steppe, bursage and perennial grassland (Laity, 2008). In arid and hyper-arid deserts higher plants are largely absent, and biological covers are dominated by small non-vascular plants: lichens and mosses (Belnap, 2003; West, 1990). Soil fauna, such as invertebrates, that are abundant in semi-arid (grassland) drylands are significantly reduced in arid deserts where microorganisms are more dominant (Fierer, Strickland, Liptzin, Bradford, & Cleveland, 2009). In hyper-arid deserts, for example in certain parts of Atacama, annual precipitation is less than 10 mm (Crits-Christoph et al., 2013; McKay et al., 2004), which is not sufficient to support plant growth. Thus, vegetation cover is almost lacking from sea level to 2500m in this area (Azua-Bustos, Urrejola, & Vicuña, 2012; Laity, 2008).

1.3 Microbial life in drylands

In deserts, small poikilohydric life forms are organisms that have adapted to moisture stress (desiccation). They colonize layers under or on the surface of soil and rocky substrates, are also called soil- and rock-surface communities (SRSCs) (Pointing & Belnap, 2012). The components of SRSCs are primarily cyanobacteria, chlorophytes, fungi, heterotrophic bacteria, lichens and mosses (Pointing & Belnap, 2012). The SRSC communities have been categorized into five groups: biological soil crusts, epiliths, hypoliths, cryptoendoliths and chasmoendoliths (Fig 1.2) (Golubic, Friedmann, & Schneider, 1981).

1.3.1 Biological soil crusts

Biological soil crusts (BSCs) are complex microbial communities colonizing the soil

surface. Due to their low requirement of moisture and the disadvantage of light competition with vascular plants, BSCs occur where the vegetation cover is sparse or absent and the water availability is limited. Up to 70% of the surface of semi-arid and arid deserts are covered by BSCs. The BSCs include lichens (*Collema* spp.) and mosses (*Bryum* spp. and *Tortula* spp.). The dominant filamentous cyanobacteria *Microcoleus* dwell within the upper layers whereas the diazotrophic genus *Nostoc* inhabits the lower soil interface layer. Other microorganisms, such as coccoid algae, heterotrophic bacteria and fungi are also identified in BSCs (Belnap, Büdel, & Lange, 2003). BSCs occur in hot and cold deserts with different external morphology. Smooth and rugose crusts occur in hyper-arid and arid hot deserts as dominant cyanobacteria bind the sand grains, creating smoother soil surface. The main difference is that rugose crusts contain sparse mosses and lichens; in contrast, rolling and pinnacled crusts are found in less arid and cooler deserts where soils frost-heave and the moisture supports growth of mosses and lichens (Belnap, 2003). The internal structure of BSCs may be related to soil types, climate regime and distance from plants (Belnap, 2003; Pointing & Belnap, 2012). For example, gypsum soils in Colorado Plateau desert support higher abundance of Actinobacteria and Proteobacteria whereas sandstone and shale soils support higher abundance of cyanobacteria (Steven, Gallegos-Graves, Belnap, & Kuske, 2013).

1.3.2 Epilthic colonization

Epiliths are the communities colonized on the bare rock surface and their distribution is ubiquitous, including desert environments. Due to the direct exposure to desiccation, biofilms are important to epilithic communities since they can retain water and facilitate access to atmospheric water vapour. Extracellular polymeric substance-secreting cyanobacteria and fungi are the key components of biofilms. Other heterotrophic, chemolithotrophic and chemoorganotrophic bacteria are also identified (Gorbushina,

2007). Epilithic lichens (foliose and crustose) and mosses occur when the environment is less extreme: Chihuahuan, Sonoran, and Mohave semi-arid deserts in North America (Nash, T. H. III, White, S. L. Marsh, 1977), milder lower latitude Antarctic regions, such as Princess Elizabeth Land and Mac Robertson Land (T. Makhalanyane, Pointing, & Cowan, 2014) and moisture-sufficient parts of McMurdo Dry Valleys (De Los Ríos, Wierzchos, & Ascaso, 2014). The microbial composition of lichen-dominated epiliths is primarily photobionts and lichen-forming fungi (De Los Ríos et al., 2014). Epilithic communities likely support other diverse taxa, their diversity was mostly identified by now-dated molecular approaches with recognized limitations in diversity estimation (De Los Ríos et al., 2014; Gorbushina, 2007; Wierzchos, de los Ríos, & Ascaso, 2012). Application of metagenomic sequencing will more accurately reveal the community structures in future.

1.3.3 Hypoliths

Most extreme drylands are typified by desert pavement, a continuous terrain with pebbles or stones embedded (Laity, 2008). Since extreme open soils do not support biological soil crusts (Bowker et al., 2016), the translucent marble or quartz embedded in the soils are colonized, supporting hypolithic communities on the ventral surface of stones (Warren-Rhodes et al., 2007; Weber et al., 2013; Wong et al., 2010). Hypoliths occur in all deserts (Bahl et al., 2011; Caruso et al., 2011) but they are particularly important to hyper-arid deserts where cyanobacteria-dominated hypoliths are the major sites of productivity in such an oligotrophic environment (Warren-Rhodes et al., 2006). This is illustrated by studies in the Arctic demonstrating hypolithic productivity was about 0.8 g/m^2 per year, which is similar to the productivity from plants, lichens and bryophytes combined (Cockell & Stokes, 2004). This dominance by hypoliths in extreme desert productivity was also recorded in the high Tibetan plateau (Wong et al.,

2010). Hypoliths are also emerging as important sites for nitrogen fixation, as demonstrated recently from Antarctic hypoliths (Cowan, Sohm, et al., 2011) and this may translate to a relevance for hypoliths as part of productivity and nitrogen fixation by the global cryptogamic cover (Elbert et al., 2012).

Hypolithic community structure is dominated by the Cyanobacteria, as demonstrated for the Atacama desert (Azúa-Bustos et al., 2011; Warren-Rhodes et al., 2006), northern Australian desert (Tracy et al., 2010), South African Knersvlakte desert (Weber et al., 2013), Namib desert (Stomeo et al., 2013), American Mojave desert (Schlesinger et al., 2003) and Chinese Taklimakan desert (Pointing et al., 2007). The Cyanobacteria-dominated hypoliths are also present in Tibetan tundra (Wong et al., 2010) and polar deserts in the Artic (Cockell & Stokes, 2004, 2006) and Antarctic Dry Valleys (Cowan, Pointing, et al., 2011; Khan et al., 2011; Pointing et al., 2009). However, the major cyanobacterial taxonomy is quite different between polar and non-polar deserts (Fig 1.3). Hypoliths from warmer non-polar deserts are dominated by the cyanobacterial genus *Chroococcidiopsis* and support greater overall richness (Bahl et al., 2011; Lacap, Warren-Rhodes, McKay, & Pointing, 2011; Warren-Rhodes et al., 2006, 2007), whereas those from extreme cold and polar deserts support mostly filamentous oscillatorian cyanobacteria and lower overall richness (Cockell & Stokes, 2006; Khan et al., 2011; Pointing et al., 2009; Wong et al., 2010). In addition to photoautotrophs, the Chloroflexi, a phylum of photoheterotrophic bacteria, may dominate in some red-pigmented hypoliths in the Atacama desert (Lacap et al., 2011). Hypoliths, especially in hot deserts, support higher abundance of heterotrophic bacteria. Actinobacteria, Alpha Proteobacteria, and Gamma Proteobacteria are ubiquitous and so these keystone taxa may be critical to hypolithic community (Chan et al., 2012). Other taxa appear to be more specific to individual desert only. For example, Deinococci occur more frequently in hot versus polar desert hypoliths (Lacap et al., 2011; Pointing

et al., 2009, 2007). Acidobacteria and Bacteroidetes only comprise minor fractions of the hypolithic community (Azúa-Bustos et al., 2011; Lacap et al., 2011; Makhalanyane, Valverde, Birkeland, et al., 2013; Makhalanyane, Valverde, Lacap, et al., 2013; Pointing et al., 2009). The archaea were absent or a minor component of hypoliths (Pointing et al., 2009; Wong et al., 2010).

1.3.4 Cryptoendoliths and chasmoendoliths

Endoliths are microorganisms colonizing just a few millimetres below the surface of porous rocks, such as sandstone, limestone and weathered granite (Pointing & Belnap, 2012). Cryptoendoliths are communities inhabiting the pore spaces of rocks, whereas chasmoendoliths occupy cracks and fissures in rocks (E. Imre Friedmann, 1980). Collectively cryptoendoliths and chasmoendoliths are referred to as endoliths. The first discovery of endoliths was in the Antarctic Dry Valleys where the environment had previously been assumed to be lifeless (Friedmann & Ocampo, 1976). They are now known to commonly occur in cold and hot deserts globally (Wierzchos et al., 2012). The community structure of endoliths is dominated by cyanobacteria or lichens, and associated with pigmented microbial layers. In general terms heavily weathered sandstone and granite substrates support lichen cryptoendoliths whilst relatively unweathered rock substrates support cyanobacterial cryptoendoliths. Both types display a layered structure that also includes fungi, green algae and heterotrophic bacteria (Fig 1.4) (De Los Ríos et al., 2014; Friedmann & Ocampo, 1976; Wierzchos et al., 2012). In lichen-dominated cryptoendoliths, the dominant phyla are chlorophyte algae and Ascomycetes fungi associated with very minor heterotrophic Actinobacteria and Proteobacteria (De la Torre, Goebel, Friedmann, & Pace, 2003; De Los Ríos et al., 2014; Wierzchos et al., 2012). The primary component of cyanobacterium-dominated cryptoendoliths is cyanobacteria, especially the genera *Chroococcidiopsis* (Dong, Rech,

Jiang, Sun, & Buck, 2007; E. Imre Friedmann, 1980; Pointing et al., 2009) and *Nostoc* (Ziolkowski et al., 2013). Chasmoendoliths are also dominated by cyanobacteria (*Chroococcidiopsis* and *Oscillatoriales*) in hyper-arid polar deserts (DiRuggiero et al., 2013; Pointing et al., 2009; Yung et al., 2014). Heterotrophic bacteria, such as Actinobacteria and Proteobacteria were largely recovered in cyanobacterium-dominated endoliths (DiRuggiero et al., 2013; Dong et al., 2007; Pointing et al., 2009; Yung et al., 2014; Ziolkowski et al., 2013). Archaeal signatures were also recovered in certain hot (Wong et al., 2009) and polar deserts (Yung et al., 2014; Ziolkowski et al., 2013).

1.4 Life under extreme aridity: hyper-arid deserts

1.4.1 Hyper-arid environments and stressors

Environmental stress such as low water availability, low or high temperature and high incidence of ultraviolet irradiation (UV) are the major stressors in dryland environments, especially in hyper-arid deserts (Table 1.3). Water is critical to cellular mechanisms and stability (Billi & Potts, 2002), therefore, microbial life that has not adapted to desiccation is unable to survive in desert environments. Temperature extremes also present intense challenges. Direct stresses, such as heat and cold shocks, and temperature fluctuation (freeze-thaw cycles), commonly occurs in hyper-arid deserts (Pointing & Belnap, 2012). In general, clear sky condition is one of the climatic features of deserts, resulting in high incidence of UV on desert surface. Clouds generally reduce surface UV level based on their amount, coverage, morphology and composition (water droplet, ice crystal and tropospheric ozone) (Cordero et al., 2014).

The Atacama Desert s classified as a hyper-arid desert locating between Peru and Chile. In its dries portion, the annual rainfall was less than 10 mm (Crits-Christoph et al., 2013; McKay et al., 2004) and in certain regions, such as Arica and Yungay

province the precipitation was even lesser than 1 mm (Crits-Christoph et al., 2013; Laity, 2008), resulting in soil water content less than 3% (Lester, Satomi, & Ponce, 2007). High aridity in Atacama is due to the following. First, based on the sedimentary data, a fluctuating climate in Atacama between semi-arid and arid occurred 150 millions ago. However, Atacama located at the dry subtropical climate belt, and a cold, upwelling current reduce this climatic fluctuation. Thus, Atacama switched to hyper-arid desert between 4 to 3 millions ago (Hartley, 2003). Second, the central Andean mountain formed between 9 to 15 millions ago and no significant latitudinal movement afterwards (Hartley, 2003). This high latitude Andean mountain blocked the moisture-containing current from Atlantic Ocean due to rainshadow effect on Atacama (Houston & Hartley, 2003). However, as a coastal desert, fog-derived moisture is the major water input to support the coastal and inland ecosystem where the rainfall events are sparse or absent. The water volume collected from fog ranged from 1.46 to 6.3 L/m^2 per day (Cáceres et al., 2007; Larrain et al., 2002) and was sufficient enough to support the predominant growth of cyanobacteria and colonization of hypoliths (Azúa-Bustos et al., 2011; Warren-Rhodes et al., 2006). Water availability is also determined by the presence of solutes in environments. High levels of salt deposit is another feature of Atacama desert surface because of the volcanic activity, rocks weathering, and dry fallout from the atmosphere over millions of years (Ewing et al., 2006; Witherow et al., 2006). The sodium and nitrate concentration in soil was 3498 mg/kg and 37 mg/kg (Crits-Christoph et al., 2013; Orlando, Carú, Pommerenke, & Braker, 2012); whereas in the Namib desert was only 1634 mg/kg and 6.9 mg/kg, respectively (Ronca, Ramond, Jones, Seely, & Cowan, 2015). The high-salt environments with water activity (a_w) of 0.9 or less would cause osmotic stress to microorganisms (Grant, 2004). The southern hemisphere generally receives more UV radiation than the northern hemisphere since the total amount of ozone is much lower (McKenzie, Smale, Bodeker, & Claude, 2003;

World Meteorological Organization, 2011). Therefore, the highest surface UV is expected to occur in summer at prevalent cloudless area in south hemisphere. The UV index in Chajnantor Plateau, Atacama was from 7 to 20 (Cordero et al., 2014) whereas in Arizona desert the average UV index was only 6.3 (Svoma et al., 2011).

Another well-known hyper-arid desert is the McMurdo Dry Valleys, a polar desert in Antarctica. Antarctica is the fifth largest continent covering 13 million km^2 and representing about 10% of the world's landmass. In Antarctica more than 98% of the landmass is covered by permanent glacial ice sheets (Bargagli, 2005), but interestingly around 0.4% of the continent is ice-free (Cary, McDonald, Barrett, & Cowan, 2010). These ice-free areas include the Ellsworth, part of the Transantarctic Mountains, the North Victoria Land Mountains, coastline of Antarctica, Antarctic Peninsula, Vestfold Hills and the McMurdo Dry Valleys (Cary et al., 2010). The Dry Valleys were largely carved out by glaciers that have retreated approximately 4.6 million years ago (Doran, Wharton, & Lyons, 1994; Prentice, Kleman, & Stroeven, 2013), composed of fourteen individual valleys. The Dry Valleys located between the Polar Plateau and the Ross Sea in southern Victoria Land that are designated as Antarctic Specially Managed Area (ASMA) to reflect their environmental significance (Scientific Committee on Antarctic Research, 2004). The geography of the Dry Valleys is comprised of perpetually primitive ice-covered lakes, periodic water streams from melting snow and arid mineral soils surrounded by barren mountains (Fig 1.5) (Cary et al., 2010). The Dry Valleys exist because the Transantarctic Mountains block most of the snow flow from the East Antarctic Ice Sheet toward McMurdo Sound. Furthermore, ablation (the melting of snow or ice that runs off the glacier, evaporation, sublimation, carving, or erosive removal of snow or ice by wind) exceeds the accumulation throughout the year (Fountain et al., 1999).

The Dry Valleys experience extremely low and also a broad range of air

temperatures: mean annual temperatures range from -27 to -10 °C, and in summer temperatures can reach up to 10 °C while in winter can be as low as -60 °C (Cary et al., 2010; Doran, McKay, et al., 2002). Extreme temperature fluctuation also occurs in the Dry Valleys. Air temperatures reached above freezing point (0.6 °C) and then could drop to below freezing point in just few minutes (McKay & Friedmann, 1985). Daily freeze-thaw events were also recorded in soil, occurring up to 79 times during the austral summer (Aislabie et al., 2006).

Low precipitation, low humidity, strong katabatic winds and high sublimation rate are the key factors leading to long periods of desiccation in Dry Valleys (Bargagli, 2005; MacClune, Fountain, Kargel, & MacAyeal, 2003; Wynn-Williams, 1990). The water content in soil ranged between 0.5 to 2 % (Lee et al., 2012). Precipitation in the Dry Valley is primarily from snow (Wynn-Williams, 2002) although rainfall occur occasionally in summer but rarely reaches the valley floor (Fountain, Nylen, Monaghan, Basagic, & Bromwich, 2010). Several reports have indicated that precipitation is excessively low, generally less than 15 cm water-equivalent per year (Cowan & Tow, 2004; Witherow et al., 2006). Annual snow fall ranged from 3 to 50 mm with the highest values near the coast and decreased gradually at inland site (Fountain et al., 2010). Furthermore, relative humidity (RH) decreases with distance from the coast (74% at Explorer's Cove) to inland valley (55% at Wright Valley) and in winter the RH could be less than 12% (Nylen, Fountain, & Doran, 2004). All these phenomena are explained by strong and persistent katabatic winds, with average annual speed of 2.5 to 4.1m per hour (Doran, McKay, et al., 2002). Drier winds descend from the Antarctica plateau down to the valley floor, causing aridity in inland region (Fountain et al., 2010). Also, persistent winds reduce the RH, especially in winter, when the occurrence of katabatic winds was five times higher than in summer (Nylen et al., 2004). Sublimation has been indicated as main factors resulting in long-term desiccation in the Dry Valleys.

Sublimation rate was up to one order of magnitude greater than snow accumulation (MacClune et al., 2003).

Oligotrophy is another extreme in this ecosystem. Organic carbon in soil was measured, ranged from 0.1 to 0.96 g/kg among several sites in Dry Valleys (Hopkins et al., 2009; Lee et al., 2012; Pointing et al., 2009), which was relatively low compared to other deserts such as the Sonoran desert in Arizona US (5 to 25 g/kg) and High Arctic Desert in north-eastern Greenland (51 to 56 g/kg) (Cary et al., 2010; Matsumoto et al., 1983). Generally, the microbial abundance is correlated with organic matters reflect the availability of nutrient sources to heterotrophic microbes (Horowitz et al., 1969).

There are no reports on the incidence of UV radiation on the Dry Valleys surface. However, due to the ozone hole above Antarctica continent, McMurdo received 30% to 60% more of UV radiation based on data collected from October and November in 1991 to 2006. It was estimated that the increasing UV level might accelerate surface warming over the Antarctica and cause the ice and snow melting. Thus, lead to the direct exposure of organisms under the UV radiation (World Meteorological Organization, 2011). In Dry Valleys the solar radiation has increased 8.1 W/m^2 per decade and this increase has resulted from decreasing winds and cloudiness (Doran, Priscu, et al., 2002).

1.4.2 Cryptic habitats

Vascular plants and complex metazoans are unable to survive in the McMurdo Dry Valleys under harsh abiotic stress (Boyd, Rothenbg, & Boyd, 1970; Cowan & Tow, 2004; Horowitz, Cameron, & Hubbard, 1972; Pointing et al., 2015), and even the epilithic crustose lichens were sparse or absent (De Los Ríos et al., 2014; Friedmann, 1982; Makhalanyane, Pointing, & Cowan, 2014). To cope with extreme environmental stress, microorganisms in the Dry Valleys have developed the stress tolerance strategies

and the first step is stress avoidance (Pointing & Belnap, 2012). They take refuge by colonizing the underside of rocky substrates (hypoliths) or within the rock interior (endoliths), whereby these cryptic habitats provide a relatively stable micro-environment that protects them from thermal, desiccation and UV stresses (Pointing & Belnap, 2012; Wierzchos et al., 2012). For example, it has been measured that temperature inside the rocky substrates were elevated above freezing point 10 to 12 hours per day compared to the surrounding air due to the solar gain from the rocks (McKay & Friedmann, 1985). Furthermore, these rocky habitats obtained moisture from dew or rime deposition as a result of thermal differences between substrates and air (Pointing et al., 2015), with up to 0.6 mm of moistures deposited above endolithic layers daily (Büdel, Bendix, Bicker, & Allan Green, 2008). Rocky substrates also effectively filtered harmful UV irradiance while providing sufficient photosynthetically active radiation (PAR) for photosynthetic autotrophs (Chan et al., 2012; Cowan, Pointing, et al., 2011).

In addition to these buffers from rocks, pigmented fungi and cyanobacteria can form a microbial "cabana", that secrete "sunscreen" compounds (melanins, mycosporines and scytonemin) and extracellular polymer substances (EPS), providing an extra layer of protection from UV radiation and retaining the moisture for community layers below (De Los Ríos et al., 2014; Pointing & Belnap, 2012). Thus, microbial communities in cryptic niches comprise the most abundant standing biomass in this ecosystem (Pointing et al., 2015; Pointing, Bollard-Breen, & Gillman, 2014).

1.4.3 microbial diversity in McMurdo Dry Valleys

Early studies showed that the soil from the Dry Valleys were sterile (Horowitz et al., 1969, 1972) or contained extremely limited microbial biomass and diversity (Siebert et al., 1996; Vishniac, 1993). This was due to methodological limitations using

microscopy and culture-dependent approaches. With the advent of molecular tools, next-generation sequencing for example (Shendure & Ji, 2008), the ability to characterize microbial communities has greatly revolutionized.

Pointing et al (2009) reported the community structures in different niches substrates (open soil, hypolith, chasmoendolith and cryptoendolith) collected from the high inland McKelvey Valley by constructing clone libraries. Cyanobacteria were the major component of hypolithic and endolithic communities, however, hypoliths were dominated by filamentous *Oscillatoriales* phylotypes while endoliths supported *Chroococcidiopsis* phylotypes. The soil communities were dominated by Actinobacteria and Acidobacteria, but no cyanobacterial signatures were recovered in the open soil. The overall bacterial diversity spanned 16 phyla in McKelvey Valley.

Chasmoendolithic communities in granite rocks have been previously surveyed in maritime Miers Valley (Yung et al., 2014). Endoliths collected from moraine, south-facing slopes and north-facing slopes displayed distinct microbial assemblages. Filamentous *Oscillatoriales* cyanobacteria dominated all assemblages in contrast to the McKelvey Valley endoliths. On the north-facing (warmer and wetter) slope, *Chroococcales* were relatively abundant, while *Chroococcidiopsis* occurred only on the colder south-facing slope due to their adaptive advantages (Billi, Friedmann, Hofer, Caiola, & Ocampo-Friedmann, 2000). Small portions of *Acaryochloris* cyanobacteria were also discovered. These cyanobacteria possess chlorophyll *d* which is favourable to conduct photosynthesis under the low light regime beneath the rock surfaces (Mohr et al., 2010). Acidobacteria and Actinobacteria were the major components of heterotrophic bacteria at all sites. Interestingly, higher abundance of Deinococcus occurred at the colder and drier slope and this may reflect the desiccation stress response of this extremophile (Billi & Potts, 2002). Similar to hypoliths from McKelvey Valley, the hypoliths in Miers Valleys were dominated by filamentous *Oscillatoriales*

phylotypes but together with large portion of Proteobacteria (Khan et al., 2011).

Soil community assemblages crossing four different sites of the Dry Valleys were further investigated by next-generation sequencing (454 pyro-sequencing) (Lee et al., 2012). Actinobacteria was the dominant component in all soil communities, followed by Bacteroidetes; however, these soil communities were structurally and phylogenetically distinct at four different valleys. Moreover, the linkage between diversity and soil geochemical properties was established, indicating that conductivity, altitude and elemental copper abundance were significantly correlated. These correlations suggested that the soil bacteria might be at low level of inter-valley redistribution or shaped greatly by local geochemical features.

Archaea and fungi are the least understood group in Dry Valleys. In Miers Valley, the uncultured *Crenarcheota* was the only archaeal signature recovered in the hypolithic community (Khan et al., 2011) whereas in chasmoendoliths archaeal assemblages were majorly *Halobacteriales*, *Desulfurococcales*, and *Thermoplasmatales* (Yung et al., 2014). The presence of *Halobacteriales* seemed to reflect the high salinity environments in the Dry Valleys due to their osmotic stress response. The distribution of thermophilic *Desulfurococcales* and *Thermoplasmatales* might be related to the geothermal activity in the Ross Sea region by airborne inputs (Bottos, Woo, Zawar-Reza, Pointing, & Cary, 2014). Archaeal communities in soil were interrogated by 454 pyro-sequencing among three sites of the Dry Valleys (Richter et al., 2014). More than 80% of recovered signatures were *Thaumarchaeota*-affiliated and this group played a key role in nitrification in nitrogen cycling. Low abundance of phylum *Euryarchaeota* was also detected, however, the ecological role and the presence of this group remained unclear.

Fungal Ascomycota commonly occurred in the edaphic and niche substrates in Dry Valleys. Studies have indicated that fungi may be more tolerant of the freeze-thaw cycle in soil than prokaryotes (Sharma, Szele, Schilling, Munch, & Schloter, 2006; Yergeau

& Kowalchuk, 2008). In endoliths, free-living Ascomycetes yeast (Yung et al., 2014), Dothideomycetes and Sordariomycetes were discovered (Pointing et al., 2009). In soils, free-living *Debaryomyces* yeast were the most abundant, and filamentous *Zalerion* and *Helicodendron* were of relatively low frequency (Rao et al., 2011). Interestingly, some of these fungal phylotypes display low similarity to any known fungus. This suggested that these undiscovered fungi played a key role in carbon turnover within cold desert soils.

In hyporheic zone, soil communities have been shown to be distinct from other soil communities in Dry Valleys (Neiderberger et al., 2015). In austral summer the influences of wind erosion are less overwhelming and temperatures are warm enough to melt the surface of glaciers, resulting in the formation of streams, flowing down to the Valleys and/ or forming ephemeral lakes. This causes an interesting geographical phenomenon, the hyporheic zone (D. M. McKnight et al., 1999). The term "hyporheic zone" refers to the area adjacent to and underneath the stream, where the melting glacier passes through the substrate in the streambed and exchanges with water in the main channel (McKnight et al., 1999). These streams are very unique from any others on Earth. Firstly, they are fed by glacial metlwater, containing ferric, silicate and nutrient sources, and they exist only six to ten weeks each year (McKnight et al., 1999; McKnight, Runkel, Tate, Duff, & Moorhead, 2004; Takacs-Vesbach, Zeglin, Barrett, Gooseff, & Priscu, 2010). Secondly, the length and depth of the streams vary. For example, the length of Onyx River in Wright Valley is from 1 km up to 30 km. As for the stream depth, it is controlled by how deep do the permafrost layers thaw in summer (McKnight et al., 1999). Thirdly, due to the absence of precipitation, lateral inflows from groundwater and terrestrial runoff have very limited hydrological interaction between the glacial streams and the barren surroundings (McKnight et al., 1999). Thus, the hyporheic zone likely plays an important role in biogeography and constitutes an

unique ecosystem within the Dry Valleys.

Striking differences in the bacterial structure of hyporiheic zone between wet and dry soil communities were the higher abundance of cyanobacteria in the former communities, and dominant Acidobacteria and Actinobacteria in the latter. Deinococcus-Thermus generally occurred in dry soil, together with Bacteriodetes, Firmicutes, Gemmatimonadetes, Nitrospira and Planctomycetes. The eukaroytic community was dominated by fungi (Neiderberger et al., 2015).

1.4.4 Microbial functional roles in the McMurdo Dry Valleys

The microbial diversity and community structures have been widely surveyed, but only few studies have reported their ecological roles, and these have been largely restricted to indirect estimates of carbon and nitrogen turnover and cycling. Isotopic evidence has indicated that organic carbon and nitrogen in valley floor soils were commonly derived from endoliths (Hopkins et al., 2009). Enzymatic activities (denitrogenase, phosphatase and beta-glucosidase) in soil communities have also indirectly indicated occurrence of carbon and nitrogen cycling in the Dry Valleys (Hopkins et al., 2008). An *in situ* respirometry study indicated the primary productivity from cryptoendoliths was 1.2 g/m^2 per year with net photosynthetic gain 0.6 g/m^2 per year. The net productivity to the Dry Valleys ecosystem was significantly lower, with long-term estimation 3 mg/m^2 per year (Friedmann, Kappen, Meyer, & Nienow, 1993). Genetic nitrogen fixation capacity was identified and nitrogen input was also estimated. Hypoliths harbouring proteobacterial and cyanobacterial (*Nostocales*) nitrogenase genes (*nifH*) were considered as an important site for nitrogen input in the Dry Valleys ecosystem, and the minimum annual nitrogen input from hypoliths to soils of three valleys (Miers, Marshall and Garwood Valleys) were about 0.38 kg (Cowan, Sohm, et al., 2011). In hyporheic soil communities, cyanobacterial *nifH* genes were largely identified whilst

proteobacterial *nifH* were relatively dominant in arid soils. Moreover, *in situ* nitrogen fixation assay has shown that hyporheic soils displayed highest nitrogen fixation rates (0.04 to 5.8 nmol/cm^3) whist dry soils presented lowest nitrogen fixation rates. These results indicated that soil moistures played an important role influencing microbial functioning (Niederberger et al., 2012).

1.5 Research gaps in the microbial ecology of the McMurdo Dry Valleys

1.5.1 Microbial carbon and nitrogen cycling pathways

All living organisms rely on essential elements (hydrogen, oxygen, carbon, nitrogen, sulphur and phosphorus) and recycling of these elements is fundamental to prevent depletion. Microorganisms are critical in process of decomposing organic materials and transforming elements so that these nutrient can support other organisms (Gougoulias, Clark, & Shaw, 2014; Schimel & Schaeffer, 2012). The carbon cycling in terrestrial ecosystems is balanced between carbon fixation (fixing CO_2 to carbon) and respiration (releasing CO_2 into atmosphere). Carbon fixation is mainly performed by photosynthetic plants, and also photo- and chemo-autotrophic microorganisms (Gougoulias et al., 2014). Nitrogen is essential to biological macromolecules (DNA and amino acids, for example) but is often limited in ecosystems; however, atmospheric dinitrogen (N_2) is not a usable form to living organisms. Nitrogen needs to be fixed into ammonium (NH_4) for biological consumption (Fowler et al., 2013).

Many terrestrial surfaces, including soils, rocks and plants, are covered by photoautotrophic communities, and these communities are known as cryptogamic covers, composed of cyanobacteria, algae, fungi, lichens and bryophytes. It has been suggested that these cryptogamic covers are major players in global carbon and nitrogen cycling (Elbert et al., 2012). Globally CO_2 uptake and nitrogen input has been

estimated, approximately 3.9 Pg (Elbert et al., 2012) and 413 Tg per year (Fowler et al., 2013), respectively. In deserts, CO_2 uptake is low, about 7% of total uptake of global forest ecosystems; however, the amount of nitrogen fixed by cryptogamic cover in deserts is largest (~0.76 g/m^{-2} per year) than other ground covers (Elbert et al., 2012).

Microbial carbon and nitrogen cycling is complex, and different types of microorganisms and pathways are involved (Nelson et al., 2016; Schimel & Schaeffer, 2012). Carbon cycling pathways are mainly associated with broad ranges of carbon fixation and decomposition (Jones, 2008; Schimel & Schaeffer, 2012). Phototrophic microorganisms, such as cyanobacteria and proteobacteria, use Ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) to fix CO_2 in the Calvin-Benson-Bassham cycle (Badger & Bek, 2008; Calvin & Benson, 1948; Espie & Kimber, 2011; Stanier & Cohen-Bazire, 1977). Alternatively, archaea and some of proteobacteria are considered as chemoautotrophs that oxidize inorganic substrates (H_2, H_2S, S or NH^3, for example) to fix carbon with different enzymes involved (Berg et al., 2010; Jones, 2008). These pathways are aerobic 3-hydroxypropionate bicycle and hydroxypropionate–hydroxybutyrate cycle, and anaerobic reductive citric acid cycle, and reductive acetyl-coenzyme A pathway (Berg et al., 2010). Methanogenesis and methane oxidization pathways, also called one-carbon metabolism pathways, are conducted by anaerobic methanogens (archaea exclusively) and aerobic methanotrophs (Verrucomicrobia and Proteobacteria, for example), respectively. Key enzymes associated with methanogenesis and methane oxidization are methyl-coenzyme M (Blaut, 1994; Murrell & Jetten, 2009; Shima, Warkentin, Thauer, & Ermler, 2002) and methane monooxygenase (Crombie & Murrell, 2014; Murrell & Jetten, 2009), respectively.

Microorganisms also play important roles in contribution to nitrogen acquisitions and cycling, in further to shape microbial communities and productivities on large scale of ecosystems (Ward & Jensen, 2014). Nitrogen cycling are broadly referred as nitrogen

input and nitrogen loss (Jetten, 2008; Ward & Jensen, 2014). Nitrogen input is mainly associated with microbial nitorgenase (NifH) to fix dinitrogen to ammonium (NH_4^+) (Jetten, 2008; Ward & Jensen, 2014). Nitrification is a two-step aerobic process conducted ammonia-oxidizing bacteria (mainly Proteobacteria) and archaea (mainly Crenarchaea) (Jetten, 2008). Ammonium is firstly oxidized to hydroxylamine (NH_2OH) by ammonia monooxygenase, then is subsequently oxidized to nitrite (NO_2^-) by hydroxylamine oxidoreductase; secondly, nitrite oxidoreductase oxidizes nitrite to nitrate (NO_3^-) (Stahl & de la Torre, 2012). At this stage, nitrite is reduced by nitrite reductase (Nir) to dinitrogen via denitrification, a process being considered as a major source of nitrogen gas emitted to atmosphere (Jetten, 2008). Assimilatory nitrite reduction and dissimlilatory nitrate reduction are pathways reducing nitrite to ammonium and oxidizing ammonium to nitrate, respectively (Jetten, 2008; Nelson, Martiny, & Martiny, 2016). Recently, a new pathway involved in nitrogen releasing to atmosphere, anaerobic ammonia oxidation has been identified. This pathway is exclusively conducted by Planctomycetes (Jetten et al., 2009) and key enzyme is hydrazine dehydrogenase (Maalcke et al., 2016). Previous studies only used respiration, enzymatic and genetic assays to estimate carbon and nitrogen cycling activities in the Dry Valleys, but little is known about the wider functional roles in community level.

1.5.2 Stress response related to the Dry Valleys environments

Another knowledge gap is stress response of communities against multiple environmental stresses in Dry Valleys. Microbial stress responses at cellular and genetic level are well studied in several model organisms (Boor, 2006; Chung, Bang, & Drake, 2006; Potts, 1994; Sinha Häder D.P. & Photochem Photobiol, 2002; Slade & Radman, 2011; Yura, Kanemori, & Morita, 2011). The capability of bacteria responding to diverse environments and sudden environmental changes has been widely explained

using *Escherichia coli* as a model (Boor, 2006; Chung et al., 2006). Once a bacterial cell faces environmental threats, expression of stress response genes are activated by sigma factors σ54 and σ70 families, in further to synthesize proteins to protect integrity of cell membrane, DNA or cytosol (Boor, 2006).

In Dry Valleys, thermal extreme is the first challenge. The potential responses would primarily be increasing the fluidity of cellular membranes by unsaturated fatty acid synthesis enzyme and desaturase, and reducing freezing point of cytoplasm by antifreeze proteins (Casanueva, Tuffin, Cary, & Cowan, 2010). Desiccation and osmotic tolerance is mediated by compatible solutes, trehalose or sucrose for example (Oren, 2008; Potts, 1994), and trans-membrane pumps (Oren, 2008). Furthermore, desiccation stress, same as UV radiation stress, would lead the severe double-strand DNA breakage. Interestingly in certain bacteria genus, (*Deinococcus* and *Chroococcidiopsis* for example) the DNA repair pathways induced by these two stresses are the same: *recA*-dependent and *ddr*-dependent repair pathways (Cox & Battista, 2005). Yet, the stress tolerance studies that are elicited from SRSCs organisms remind unclear. Two studies have indicated that desert-derived *Chroococcidiopsis* could survive under the treatment of ionizing radiation (Billi et al., 2000), UV radiation, frequent freeze-thaw cycles and Mars-like environment (Billi et al., 2011).

Cyanobacteria are the dominant specie in dryland ecosystems and form an upper protective layer for lower-layer microorganisms. This is due to the EPS shielding the outmost layer of cyanobacterial cell structure (De los Ríos et al., 2014; De Los Ríos et al., 2014). The putative roles of EPS are involved in biomineralization, adhesion to solid substrates, nutrient capture, and most importantly, moisture retention (Pereira et al., 2009). It is also noticed that the UV-absorbing pigments were found in EPS shield (Pereira et al., 2009) and their synthesis were induced by UV radiation (Wright et al., 2005). Based on the above rationales, understanding the regulation and variation of

cyanobacterial stress response genes is important.

1.6 Application of novel methodology: the GeoChip DNA functional array

DNA Microarrays are a powerful tool for large-scale quantitative study in gene loci detections (Stoughton, 2005). Owing to the vast microbial diversity and uncultivated status of microorganisms (Amann, Ludwig, & Schleifer, 1995), the investigation of microorganisms in natural environments, and the relation between the microbial diversity and functionality in genetic level is very challenging, especially on a large-scale and parallel survey. To overcome this obstacle, GeoChip, a DNA functional gene array (FGA) has been developed and used for investigating functional genes involved in biogeochemical, ecological and environmental processes (He et al., 2007). To date, the GeoChip 4 contained approximate 82,000 50bp oligonucleotide probes covering 142,000 coding sequences from 410 functional gene families related to carbon, nitrogen, sulphur and phosphorus cycling, energy metabolism, antibiotic resistance, metal resistance, organic remediation, stress responses, bacteriophage and virulence (Tu et al., 2014). The high coverage of functional genes is able to reveal functional difference in microbial communities, and in further to predict ecological roles associated with different microbial taxa (He et al., 2010). For example, microorganisms play important roles in carbon cycling and key enzymes (Rubisco, carbon monoxide dehydrogenase, ATP cotrate lyase and acetyl-CoA carboxylase) involved in carbon dioxide fixation were designed as probes in GeoChip (He et al., 2010). The detection of these genes in communities would interpret the putative carbon fixation process associated with specific microorganisms.

Moreover, a phylogenetic marker, DNA gyrase unit B gene (*gyrB*) was also targeted for diversity investigation (He et al., 2010). Gene *gyrB* encodes a type II DNA topoisomerase, which plays an essential role in DNA replication and is ubiquitous

among bacterial species (Drlica & Zhao, 1997). Different from 16S rRNA gene, *gyrB* can achieve higher evolutional resolutions at species-strain level than 16S rRNA gene at genus-family level (Peeters & Willems, 2011; L.-T. Wang, Lee, Tai, & Kasai, 2007; Yamamoto & Harayama, 1998). This is due to the rate of molecular evolution inferred from *gyrB* gene sequences is faster than 16S rRNA gene sequences (Yamamoto & Harayama, 1995). Several bacterial strains among *Pseudomonans* (Yamamoto & Harayama, 1998), *Acinetobacte*r (Yamamoto & Harayama, 1996), *Mycobacteria* (Kasai, Ezaki, & Harayama, 2000), *Bacillus* (L.-T. Wang et al., 2007) and *Cronobacter* (Huang, Chang, & Huang, 2013) have been phylogentically examined by using *gyrB* sequences; the results have shown that the phylogeny based on *gyrB* resulted in higher resolution than that based on 16S rRNA sequences. Overall, functional genes derived from 173 archaeal, 4138 bacterial, 404 eukaryotic (fungal) and 252 viral strains were targeted in GeoChip 4, providing a platform to link the microbial diversity to ecosystem process and functions (Tu et al., 2014).

1.7 Research objectives

Previous GeoChip study has made a first attempt to interrogate the functional diversity for entire soil, hypolithic and endolithic communities and established the geobiological pathways in high altitude inland McKelvey Valley (Chan, Van Nostrand, Zhou, Pointing, & Farrell, 2013). The putative primary metabolic pathways, such as carbon and nitrogen cycling, were identified by presence of key enzyme genes from different microbial taxa. In this oligotrophic environment carbon sequestration is a major challenge, and hence autotrophic and heterotrophic pathways are important. Different forms of RubisCo enzymes in the Calvin-Benson-Bassham cycle were detected, indicating that cyanobacteria as the potential photoautotrophs and Archaea, Actinobacteria and Proteobacteria as chemoautotrophs. Other carbon transformation

pathways such as acetogenesis, methanogenesis and methane oxidation were also discovered in specific phyla. Importantly, the ability to transform recalcitrant aromatics carbon explained the main difference between rock and soil communities. The nitrogen fixation was largely contributed by Actinobacteria, Cyanobacteria, and Alpha- and Epsilon-proteobacteria. Mineralization ability, also known as ammonification, was mainly conducted by fungal basidiomycetes in rock substrates. ANNAMOX (anaerobic ammonium oxidation) was very active in the niches communities yet was only specific to Planctomycetes. These results suggested the significant plasticity in autotrophic, heterotrophic and diazotrophic strategies supporting the communities. The key functional genes involved in thermal, radiation, desiccation, osmotic stress, and nutrient limitation responses were also identified by GeoChip, indicating that all these communities have been capable of tolerating extreme stressors (Chan et al., 2013).

This GeoChip study provides an insight on the potential diversity of primary metabolic and stress response pathways of microbial communities in inland McKelvey Valley (Chan et al., 2013), but a knowledge gap remains for the lowland maritime Dry Valley ecosystem, Miers Valley (Fig 1.6). The overall goal of this thesis is to interrogate the functional diversity of hypolithic, chasmoendolithic and soil communities in Miers Valley. Two metagenomic approaches: GeoChip microarray and Miseq next-generation sequencing platform were applied. Specific objectives are guided by research questions as follows. First, among different cryptic (hypoliths and chasmoendoliths) and open soil communities, which microbial taxa are the main contributors to each specific carbon and nitrogen transformations? Second, are these communities able to respond to environmental stress since the Dry Valleys exhibit abiotic stress factors? Third, compared to the inland McKelvey Valley is the functional gene diversity of communities in the Miers Valley different? Furthermore, apart from those intense abiotic stressors in the Dry Valleys, are there any potential biotic factors

also involved in shaping community structure and biomass? On the other hand, Cyanobacteria are the dominant species in dryland ecosystem and this leads to a interesting question: are there any specific stress response mechanisms in desert cyanobacteria? Do they harbour different stress response genes than cyanobacteria derived from any other environments?

In chapter two, carbon and nitrogen cycling, together with stress response pathways of chasmoendolithic communities (granite substrates) in Miers Valley was identified using GeoChip. Community structure was estimated using the phylogenetic marker (*gyrB*) within GeoChip. This was the first report of functional ecology for granite chasmoendoliths in the Dry Valley ecosystems. In chapter three, taxonomic and functional diversity of hypolithic and soil communities in Miers Valley was interrogated using 454 pyrosequencing and GeoChip. The combined data provided an insight of spatial distribution of genes in productivity and metabolism, and establish a relationship between community structure and functional diversity. In chapter four, specific functional probes related to genes for antibiotic resistance and bacterial phage in GeoChip were selected to identify a potential biotic control on microbial biomass among different sites of the Dry Valleys. In chapter five, cyanobacteria isolated from different deserts were subjected to metagenome sequencing using the Miseq for identifying potential novel stress response genes. In chapter six, a synthesis of the findings, conclusion and opportunities for future work were identified, plus limitations of the two metagenomic approaches used were also discussed.

Table 1.3 The features of hyper-arid deserts: Atacama desert and McMurdo Dry Valleys

	Atacama desert	McMurdo Dry Valleys
Mean annual T °C	17.1 (McKay et al 2003) 16.4-17.6 (Crits-Christoph et al 2013)	-27.4 to -17.7 (Doran et al 2002) -23 to -10 (Cary et al 2010)
Max mean T °C	37.9 (McKay et al 2003)	-19.1 to -30.0 (Doran et al 2002)
Min mean T °C	-5.7 (McKay et al 2003)	-25.4 to -14.8 (Doran et al 2002)
Soil pH	6.6-7.3 (Lester et al 2007) 6.25-8.26 (Barros et al 2008) 7.4-9.2 (Crits-Christoph et al 2013)	6.5 - 8.3 (Cowan and Tow 2004) 8.1- 9.4 (Pointing et al 2009) 6.9-8.6 (Lee et al 2012) 8.5-9.1 (Richter et al 2014)
Total organic carbon g/kg	0.56-0.77 (Lester et al 2007) 5.2 (Orlando et al 2012) <10 (Crits-Christoph et al 2013)	0.2-3.8 (Cowan and Tow 2004) 0.13-0.96 (Hopkins et al 2009) 0.4-1.3 (Pointing et al 2009) 0.02 - 3.04 (Cary et al 2010) 0.5-1.2 (Richter et al 2014)
Total nitrogen %	0.01-0.15 (Barros et al 2008)	0.005-0.016 (Hopkins et al 2009) 0.04-0.12 (Pointing et al 2009) 0.05- 0.13 (Lee et al 2012) <0.05 (Richter et al 2014)
Mean annual relative humidity %	40-82 (Cáceres et al 2007) 36-70 (Crits-Christoph et al 2013)	55-74 (Doran et al 2002)
Mean annual precipitation	0.1-2.3 mm (McKay et al 2003) <20 mm (Orlando et al 2012) 1 -4.7 mm (Crits-Christoph et al 2013)	0-100 mm (Cary et al 2010) 3-50 mm (Fountain et al 2010)
Water content %	3 (Lester et al 2007)	0.14- 4.3 (Cowan and Tow 2004) 0 -2 (Pointing et al 2009) 0.5-2.3 (Lee et al 2012)
Fluoride µg/g	19-28 (Lester et al 2007)	
Chloride µg/g	8.5-36 (Lester et al 2007)	5-835 (Cowan and Tow 2004)
Sodium µg/g	1944-3498 (Crits-Christoph et al 2013)	1400-6100(Pointing et al 2009) 1400-3884 (Lee et al 2012)
Calcium µg/g	6013-34186 (Crits-Christoph et al 2013)	6700-14000 (Pointing et al 2009) 2747-12337 (Lee et al 2012)
Sulfate µg/g	18000-22000 (Lester et al 2007) 9015-25218 (Crits-Christoph et al 2013)	3-15 (Cowan and Tow 2004) 2.4 (Bao and Marchant 2006)
Nitrate µg/g	44-130 (Lester et al 2007) 17 (Orlando et al 2012)	2-7 (Cowan and Tow 2004) 0.62 -1.86 (Barrett et al 2007)

Table 1.4 The major bacterial assemblage of communities at different sites of Dry Valleys

	Hypolith	Cryptoendolith	Chasmoendolith	Dry soil	Wetted soil
McKelvey Valley	Cyanobacteria (Pointing et al., 2009)	Cyanobacteria Bacteriodetes (Pointing et al., 2009)	Cyanobacteria (Pointing et al., 2009)	Actinobacteria, Acidobacteria, Gemmatimonadetes (Pointing et al., 2009)	
Battleship Promontory				Actinobacteria, Bacterodetes, Unclassified bacteria (Lee, Barbier, Bottos, McDonald, & Cary, 2012)	
Upper Wright Valley				Actinobacteria, Bacterodetes, Firmicules (Lee et al., 2012)	
Beacon Valley				Actinobacteria, Bacterodetes, Firmicules (Lee et al., 2012)	
Miers Valley	Cyanobacteria (Khan et al., 2011; Makhalanyane et al., 2013), Proteobacteria (Khan et al., 2011)		North-facing slope: Cyanobacteria, Actinobacteria, Acidobacteria, Bacteroidetes, Planctomyete, γ-proteobacteria South-facing slope: Cyanoabcteria, Actinobacteria, Acidobacteria, Deinococcus Morain: Cyanoabcteria, Actinobacteria, Acidobacteria, Bacteroidetes, Deinococcus, α-proteobacteria (Yung et al., 2014)	Actinobacteria, Proteobacteria (Lee et al., 2012; Makhalanyane et al., 2013)	Nostoc Pond: Cyanobacteria, α-proteobacteria, β-proteobacteria, Unclassified bacteria Miers Stream: Cyanobacteria, α-proteobacteria, β-proteobacteria, Unclassified bacteria Miers Lake: Cyanobacteria, β-proteobacteria, Unclassified bacteria Buddha Lake: Bacteroidetes, α-proteobacteria, β-proteobacteria, Unclassified bacteria (Neiderberger et al., 2015)

Fig 1.2 Different types of SRSCs community. **a** A pinnacled biological soil crust; **b** epilithic colonization on rocks; **c** hypolithic community of a quartz substrate; **d** endolithic colonization has been exposed within a fractured rock (Chan et al., 2012; Pointing & Belnap, 2012).

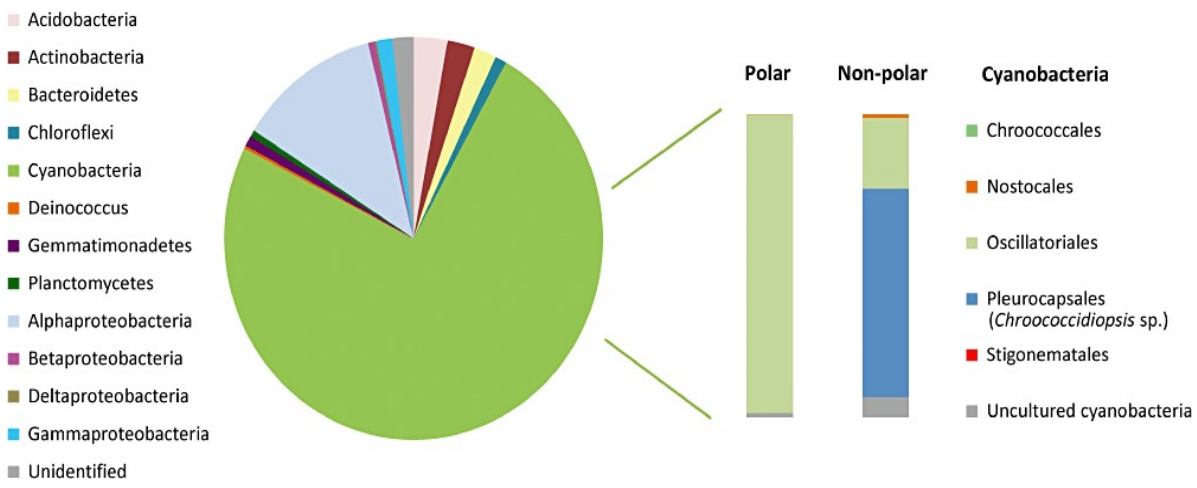


Fig 1.3 Relative abundance of recoverable bacterial phyla in hypolithic communities from deserts worldwide. Hypoliths from polar deserts greatly supported filamentous Oscillatoriales (Chan et al., 2012).

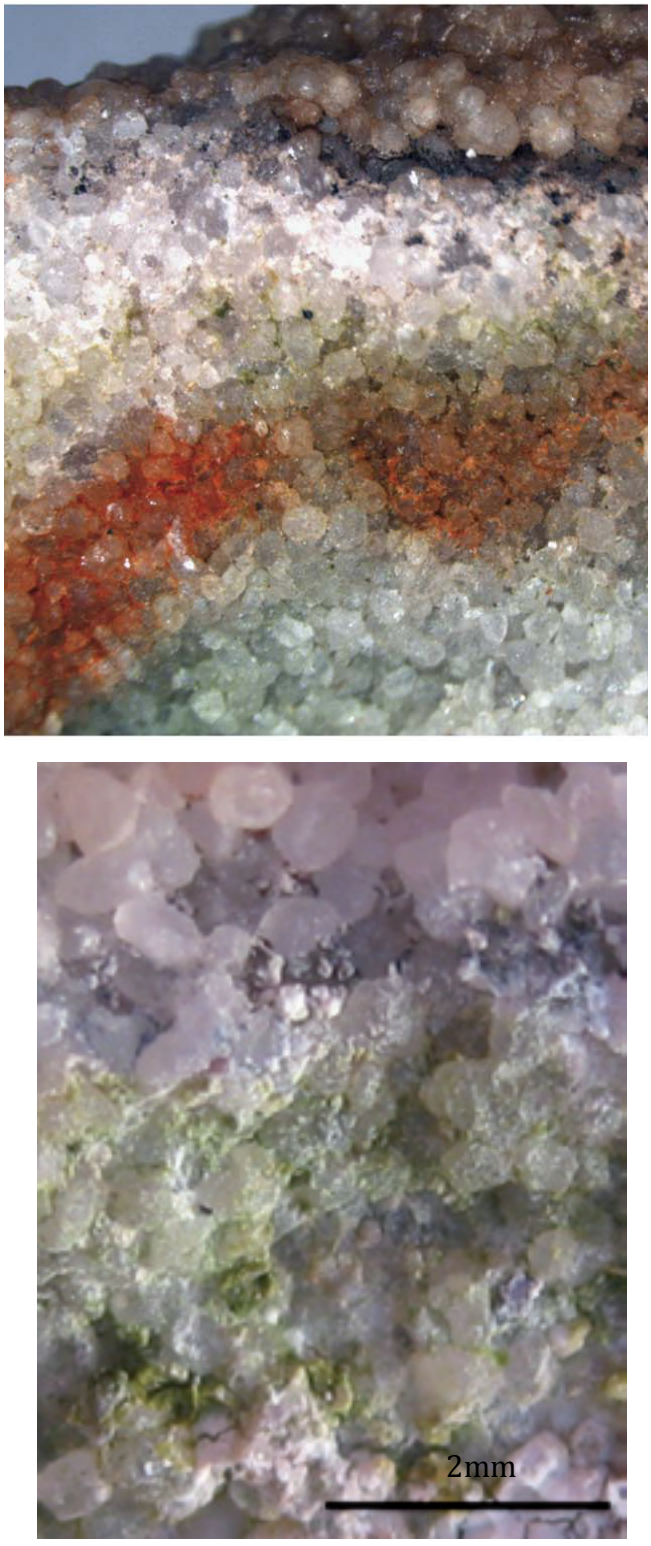


Fig 1.4 Coloured bands of endolithic communities in a fractured sandstone (De Los Ríos et al., 2014; Pointing & Belnap, 2012).

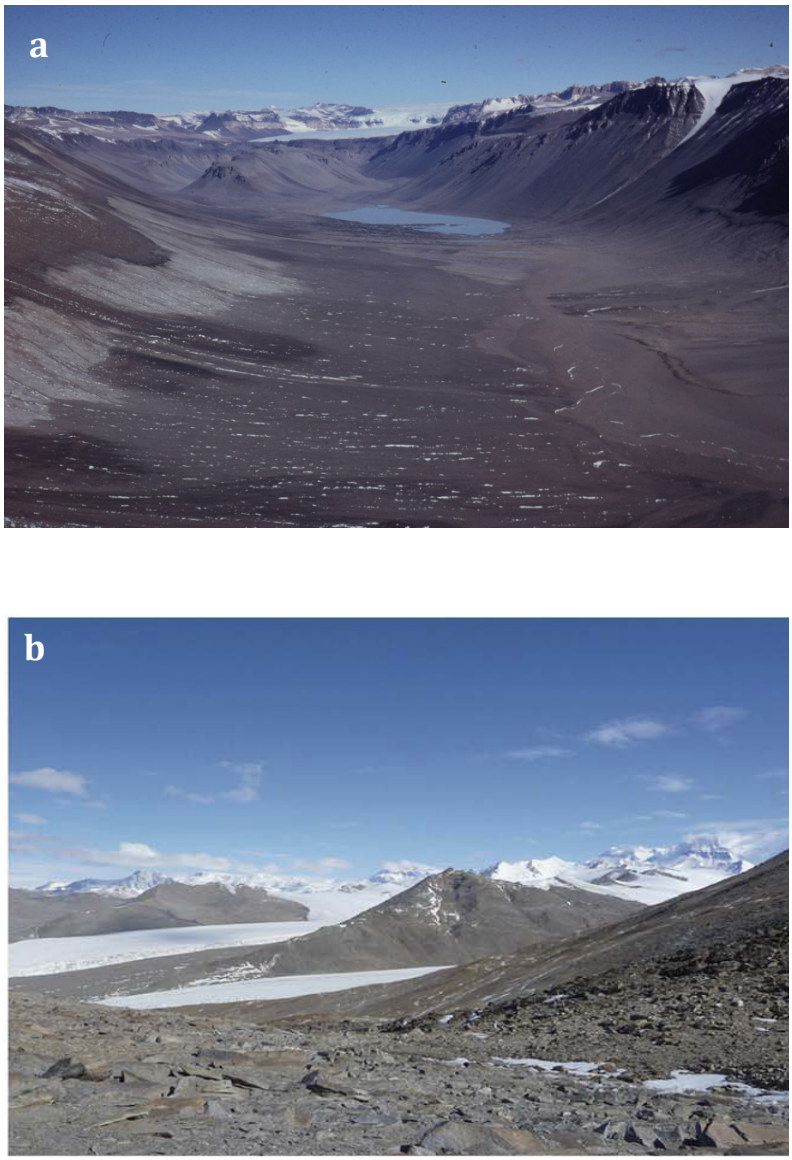


Fig 1.5 The landscape of McMurdo Dry Valleys. **a** the Only River in the Wright Valley; **b** the general landscape of Miers Valley (Cary et al., 2010; De Los Ríos et al., 2014)

Fig 1.6 The map of McMurdo Dry Valley. The lowland maritime Miers Valley is highlighted in red; the inland McKelvey is highlighted in white; the Taylor Valley is highlighted in orange; the Victoria Valley with hyporheic sites is highlighted in blue.

Chapter 2 Diversity of metabolic and stress-tolerance pathways in chasmoendolithic and soil communities of the McMurdo Dry Valleys

2.1 Introduction

Chasmoendolithic communities are microorganisms colonizing the cracks or fissures of weathered rocks (De Los Ríos et al., 2014). Freezing and fracturing processes facilitate the occurrence of chasmoendolithic colonization (Cowan & Tow, 2004), which been reported in granite, marble, silicified, sandstone, gypsum crusts and anorthosite (De Los Ríos et al., 2014). In contrast, cryptoendoliths have been discovered commonly in sandstone due to the porous nature of rocks (de los Ríos et al., 2007; De Los Ríos et al., 2014). Compared with hypolithic and cryptoendolithic communities, chasmoendolith diversity estimated by 16S rRNA genes has been less well studied in the Dry Valleys ecosystem (Cary et al., 2010). Two recent studies have reported the diversity of chasmoendoliths using 16S clone libraries in the McKelvey and Miers Valley. The McKelvey Valley is a high inland valley where the floor is dominated by quartz and sandstone. Chasmoendolithic colonization was discovered in sandstone and colonizing frequency of stones was 1% (Pointing et al., 2009). The community assemblage was dominated by cyanobacteria (*Chroococcidiopsis*-like phylotype) associated with a minor Proteobacterial component (Pointing et al., 2009). In Miers Valley, chasmoendolithic colonization has been widely discovered (up to 30%) in weathered granite substrates (Yung et al., 2014). The community was dominated by *Leptolyngbya*-like cyanobacteria (45% of total bacteria abundance), and *Synechococcus*- and *Chroococcidiopsis*-like phylotype in north-facing slope and south-facing slope, respectively. Heterotrophic bacteria were dominated by Acidobacteria and Actinobacteria with no slope-related differences. In contrast to McKelvey Valley, archaeal (*Halobacteriales*, *Desulfurococcale*s and *Thermoplasmatales*) and fungal

(Ascomycetes) signatures were recovered in Miers Valley (Yung et al., 2014).

Understanding of the functional roles of chasmoendoliths remains limited, only speculating from microbial diversity of chasmoendoliths. Yung (2014) suggested that chasmoendoliths might play a significant role in primary production due to the dominance of cyanobacteria. Nitrogen fixation might not be significant due to the low frequency of Alpha-proteobacteria (Yung et al., 2014). The presence of black yeast-like *Aureobasidium* (Yung et al., 2014) indicated bio-weathering processes in granites because such yeast produced several organic acids and extracellular substances (Hirsch, Eckhardt, & Palmer Jr., 1995). Apart from this, the ecological roles of other bacteria, archaea and fungi remain unclear. Therefore, a functional gene-targeted approach is warranted for resolving this issue, and for further elucidating basic ecology of chasmoendoliths in Dry Valleys.

To date, a single study using GeoChip has addressed the functional ecology of the McMurdo Dry Valleys systems (Chan et al., 2013). This revealed that all rock and soil inhabiting communities (included sandstone-colonized chasmoendoliths) in McKelvey Valley harboured functional capacities for carbon and nitrogen transformations, and also a great diversity in stress response pathways. However, microbial communities in the Miers Valley appear to harbour different suites of these functional pathways due to the different valley locations, homogeneity of substrates and community assemblage. In this study, the functional diversity of chasmoendolithic communities in weathered granite, a major lithic substrate for microbial colonization in lowland maritime Miers Valley, was interrogated using the GeoChip microarray. The putative metabolic (carbon and nitrogen cycling) and stress response pathways of chasmoendoliths were then compared to those from surrounding soil communities in order to elucidate the relative contributions to geobiological transformations, and also a range of stress tolerance pathways that may underpin the adaptive basis for community assembly in these two

niches.

2.2 Materials and methods

Miers Valley is a granite-dominated valley occupying a maritime location within the McMurdo Dry Valleys Antarctic Special Managed Area. It is a long-term ecological study site for the New Zealand Terrestrial Antarctic Biodiversity Survey (NZTABS, http://nztabs.ictar.aq). This valley lies between the latitudes 78°060 S and 78°070 S, and longitudes 163°440 E and 164°120 E. A landscape ecological survey of biodiversity was conducted in 2009 (Yung et al., 2014) and 2011. Triplicate independent weathered white granite boulders were obtained 200 m apart from each other in the north-facing slope in 2009 (Yung et al., 2014). Triplicate soil samples were collected at valley floor characterized by hypolith terrain and moraine deposition in 2011.

DNA from soil samples (n=3) in this study were extracted using PowerSoil® DNA isolation kit (Mo Bio Laboratories Inc., Carlsbad, California, USA) according to the manufacturer's protocol. DNA used for chasmoendoliths (n=3) biodiversity survey from previous study (Yung et al., 2014) was also used for GeoChip hybridization in this study. GeoChip hybridization and raw data output was carried out as previously described (Zhou, Kang, Schadt, & Garten, 2008). Microarray interrogation was conducted for triplicate samples from soil and granite (n=6). The GeoChip contains probes for detecting functional key genes involved in several metabolic and biogeochemical pathways. Three major functional categories were targeted in this study: carbon cycling, nitrogen cycling and stress response (He et al., 2007; Tu et al., 2014). The normalized hybridization output data of each single gene was then grouped based upon the total signal intensity of each functional category. Visualization of different phylum-level and/or class-level contributions to each metabolic pathway was achieved using spider dendrograms, where each arm of the plot was specific to a given

phylum/class. Table S2.1 provides two-character codes for the microbial taxa for this chapter.

Alpha diversity indices [Species Richness (S), Diversity indices; Margalef's Diversity (Da), Shannon's Diversity (H') and Simpsons Diversity Index (D)], and equality of species abundance (Pielou's Evenness (J')] were calculated from *gyrB* diversity data. Visualization of different phylum-level and/or class-level contributions to each metabolic pathway was achieved using spider dendrograms, where each arm of the plot was specific to a given phylum and signal intensity was used as a proxy for relative abundance (Chan et al., 2013). Statistical tests using analysis of similarity (ANOSIM) and analysis of variance (ANOVA) were performed to indicate confidence in similarities and differences observed, respectively. All analyses were performed using PRIMER-E v6 (Clarke, 1993).

Note: DNA from chasmoendoliths intended for high-throughput sequencing (as conducted in Chapter 3 for the hypolith-soil comparison) was destroyed during the relocation of this research group from The University of Hong Kong to Auckland University of Technology. That is why this chapter describes GeoChip work only, whilst Chapter 3 describes GeoChip complemented with high throughput sequencing.

2.3 Results

2.3.1 GeoChip analysis and functional diversity

The main focus of this study was to elucidate the microbial diversity that underpinned metabolic potential in carbon cycling, nitrogen cycling, and stress tolerance pathways. Output from the array data were grouped into functional categories related to major metabolic processes. Carbon fixation, acetogenesis, methane oxidation, methanogenesis and carbon degradation were grouped into carbon cycling; nitrogen fixation,

nitrification, denitrification, ammonification, assimilatory nitrate reduction, dissimilatory nitrate reduction (DNAR) and anaerobic ammonium oxidation (ANAMMOX) were grouped into nitrogen cycling; environmental stress (osmotic stress, radiation stress, heat shock, cold shock) and nutrient stress were grouped into stress response. Among the 34,473 probes returning positive signals, 5,888 were derived from genes involved in carbon cycling, 2,820 nitrogen cycling, and 7,230 from stress responses. Hybridization of DNA from chasmoendolith and soil samples was achieved with an average of 64.0 % of the probes, covering 93.2 % of the targeted genes of interest on GeoChip 4. The overall difference in metabolic and stress response pathways indicated for these two communities was significant (ANOSIM: global R = 0.761, P = 0.0027). A SIMPER analysis was performed to identify which categories of pathway-specific genes was primarily responsible for the observed difference between these two niches. Overall, the most striking difference (24 % of variation) between chasmoendoliths and soils was the catabolic pathways for aromatic compounds.

2.3.2 Biodiversity

For determining prokaryotic diversity GeoChip utilized *gyrB* as a phylogentic marker. In total, 658 probes were given the positive signal intensities, and these output were normalized and calculated for Alpha diversity indices (Table 2.1). Both communities displayed similarly high levels of evenness, indicating the distribution of microbial species was similar although soil communities were significantly more diverse than chasmoendoliths. Proteobacteria were prevalent in both communities; the relatively abundant phyla in chasmoendoliths were archaeal Halobacteria, Acidobacteria, Cyanobacteria and Thermoplasmata, whereas in soil Acidobacteria, Chlorobi, Halobacteria and Thermoplasmata were more dominant. The Deinococci and Archaeoglobi appeared specific to soil only, whilst Nitrospirae were markedly more

abundant in chasmoendliths (Fig 2.1).

2.3.3 Carbon utilization

Potential carbon fixation (photoautotrophy and chemoautotrophy) was indicated among 25 bacterial and archaeal phyla (Fig 2.2a & 2.2b). The overall autotrophic signatures in chasmoendoliths were slightly higher than in soils, and some phylum-specific differences were apparent. Notably the strong photoautotrophic and chemoautotrophic signatures were contributed by Zeta-proteobacteria and Thermotogae in soil only, respectively. The capability of acetogenesis was also identified (Fig. 2.2c) and dominant phyla for this pathway were different in chasmoendoliths (Actinobacteria, Alpha-proteobacteria and Firmicutes) and soils (Bacteriodetes, Chloroflexi and Gamma-Proteobacteria). In C1 metabolism the Methanococci dominated methanogenesis (Fig. 2.2d), whereas methane oxidation was indicated by seven phyla and dominated by Proteobacteria in both communities (Fig. 2.2e). The ability to utilize simple and complex carbohydrates was present in 47 archaeal, bacterial and fungal phyla in both substrates. The strongest signals were detected from Exobasidiomycetes and Dacrymycetes fungi, and Fibrobacteres (Fig. 2.2f). Catabolism of complex aromatic compounds was a less widespread trait and dominated by archaea and fungi (Fig. 2.2d). The overall signals were more abundant among soil communities.

2.3.4 Nitrogen utilization

The ability to fix nitrogen was indicated by a large number of phyla (Fig. 2.3a). Archaeal signals were more abundant due to Halobacteria and Methanopyri whereas bacterial signals were dominated by Alpha-proteobacteria, Fusobacteria, Spirochaetes and Cyanobacteria. Most abundant nitrifiers (oxidation of ammonia or ammonium into nitrite, and oxidation of the nitrite into nitrate) were Alpha-and Delta-proteobacteria,

Cyanobacteria and *Deinococcus*-Thermus (Fig. 2.3b). The potential for soil nitrate removal via denitrification was also indicated for 15 phyla, and differences in the dominant phyla between chasmoendoliths and soils were Halobacteria and Aquificaed (Fig. 2.3c). Ammonification (introduction of nitrate via decomposition) was indicated and strongest signals in chasmoendoliths were fungal (Sordariomycetes and Agaricomycetes), whereas in soils bacterial signals were relatively greater (Fig. 2.3d). Anoxic nitrate assimilation was largely indicated for Beta-proteobacteria, Firmicutes and Verrucomicrobia (Fig. 2.3e). Strongest DNRA signals were from Bacteriodetes, Delta-proteobacteria and Firmicutes (Fig. 2.3f). ANAMMOX genes were detected only in Planctomycetes.

2.3.5 Stress response

In stress response pathways, environmental stresses (osmotic stress, heat shock, cold shock and radiation stress) and nutrient stress (nitrogen and phosphate limitation) were targeted. Twelve phyla displayed pathways for osmotic stress tolerance, accounted for largely by Thermococci, Actinobacteria, Bacteriodetes and Delta-proteobacteria (Fig. 2.4a). Radiation stress responses were widespread, with stronger signals in Deinococci, Lentisphaerae, Nitrospirae and Proteobacteria (Fig. 4b). Heat shock pathways were very widespread among archaeal, bacterial and fungal phyla (Fig. 4c). Among chamsoendoliths strongest signals of heat shock genes were from Acidobacteria, Proteobacteria and Basidiomycete fungi. Cold shock pathways were indicated by relatively few phyla, namely Archaeoglobi, Firmicutes, and Proteobacteria (Fig. 4d). Pathways for nitrogen and phosphate limitation stress were widespread among almost all phyla, with relatively higher signal abundance in chasmoendoliths due to the fungal contribution.

2.4 Discussion

This GeoChip-based study links biodiversity to functional traits in the two major terrestrial microbial substrates (granites and soils) within the Miers Valley, and thus provides an inventory of putative geobiological contributions by Miers Valley microorganisms. The diversity estimates using the *gyrB* marker yielded taxonomic resolution at the species-strain level and thus generally provided higher biodiversity estimates than 16S rRNA gene-based estimates of this valley system: up to 18 bacterial and four archaeal phyla were discovered in soils and chasmoendoliths. This expands significantly on a previous diversity survey using 454 pyro-sequencing that revealed the Miers Valley soil community comprised only seven bacterial phyla with dominant Actinobactera (Lee et al., 2012). Chasmoendolithic community surveyed by constructing clone libraries has indicated that Cyanobacteria, and *Halobacteriales*, *Desulfurococcales* and *Thermoplasmatales* were the major groups encountered (Yung et al., 2014). The estimates using the *gyrB* phylogenetic marker in this study resulted in high abundance of all these phyla. However, a direct comparison between sequence- and microarray-based approaches might not be suitable due to the biases associated with primers (Suzuki & Giovannoni, 1996) and the specificity of oligonucleotide probes (He et al., 2007). Additionally, a microarray approach such as GeoChip is limited by the number of probes (2,347 specific probes for *gyrB*) (Tu et al., 2014), thus, the coverage of taxonomy might be specific to certain bacterial strains and result in different diversity estimation.

An interesting finding was that the *gyrB* estimates implied a relatively high abundance of Halobacteria in soil and granite, as this archaeal phylum was overlooked in previous studies focused on bacteria only. Halobacteria are a well-known halophilic archaeal phylum requiring 10 to 15% salt in environments for optimal growth and

cellular stability (Oren, 2008). The presence of this halophile might reflect that osmotic stress exerts a major force in shaping the microbial assemblages in Dry Valleys. Several studies have indicated that salt content (conductivity) presented strong correlation with diversity and structure in soil community (Lee et al., 2012; C. Magalhães et al., 2012; Pointing et al., 2009). Although there is no phylogenetic marker for fungi in GeoChip, other functional genes involved in degradation of polymers and aromatic compounds reveal the potentially important functional role for fungi. Only a few fungal phylotypes (filamentous Ascomycetes, Dothideomycetes, Sordariomycetes and Cystobasidiomycetes) have been recorded in previous studies (Arenz & Blanchette, 2011; Arenz, Held, Jurgens, Farrell, & Blanchette, 2006; Pointing et al., 2009; Rao et al., 2011; Yung et al., 2014) but here several functional gene signatures have been detected largely from different group of Ascomycota, Basidiomycota, Neocallimastigomycota fungi, implying that fungi may be an emerging force in Dry Valleys ecology.

Chasmoendoliths supported slightly higher abundance of Ribulose-1, 5-bisphosphate carboxylase/oxygenase (*rubisco*) genes than soil in this study, thus granites communities were envisaged as the major carbon fixation sites to this ecosystem along with soil communities. This speculation is supported by the dominant cyanobacterial assemblage in chasmoendolithic community (Yung et al., 2014) and the laboratory- and computer-based estimations of photosynthetic productivity for cryptoendolithic communities in Dry Valleys (E. Imre Friedmann et al., 1993). The recovered signatures of chemoautotrophy has indicated that non-photosynthetic inputs, such as novel archaeal 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) and dicarboxylate/4-hydroxybutyrate cycle (DC/4-HB) cycles (Berg et al., 2010), and proteobacterial ion-reducing pathways (Shively, van Keulen, & Meijer, 1998) may also occur in this ecosystem. The abundance of signals for aromatic compound degradation in soil

relative to chasmoendolith suggests that soil communities have evolved to utilize recalcitrant "legacy" carbon reservoirs (Chan et al., 2013). Since primary producers, carbon turnover rate and external carbon inputs in Dry Valleys are extremely low, it has been proposed that this subsidizing carbon source in soils might originate from ancient organic matters between 10,000 to 23,000 years ago (Hopkins et al., 2006).

Nitrogen fixation in both chasmoendoliths and soils was largely contributed by the known diazotrophic cyanobacteria (Nostocales and Oscillatoriales) and Alpha-proteobacteria. This is consistent with signatures (*nifH*) for diazotrophy obtained in hyporheic (Niederberger et al., 2012) and extreme inland Antarctic soil (Chan et al., 2013). Soils supported a higher overall abundance of nitrogen-fixing proteobacteria compared to chasmoendoliths. This raises an interesting paradigm as to whether biological carbon and nitrogen inputs to this ecosystem are spatially discreet, with greater carbon fixation in granite and greater nitrogen fixation in soils. This basis may result from substrate geochemistry, where sufficient nitrogen for slow-growing endolithic colonization was produced by atmospheric electric discharges, conveyed to the rock via atmospheric precipitation (E. Imre Friedmann & Kibler, 1980); whereas the soils in Dry Valleys were generally oligotrophic but displayed higher organic matters in lowland maritime Miers Valley (Lee et al., 2012). The higher organic matter in Miers Valley might be due to mummified seal carcasses provide exogenous organic matter and support microbial communities at point sources (Cary et al., 2010). This geochemical property of soil also affected the abundance of nitrification gene (*amoA*) and the distribution of ammonia-oxidizing bacteria (AOB) and archaea (AOA) in Dry Valleys. In this study most of the nitrifiers were bacteria phyla, and this was supported by a recent study that has suggested the soil communities in Miers Valley possessed more capacity of AOB *amoA* than AOA *amoA*, resulting from the low conductivity and high C/N ratio (C. M. Magalhães, Machado, Frank-Fahle, Lee, & Cary, 2014). The

potential for a complete pathway for nitrogen transformation was evident, and so under conditions favorable to ANAMMOX (a novel nitrogen cycling pathway only conducted by Planctomycetes) (Jetten et al., 2009) and denitrification pathways, microbially mediated loss of nitrogen from the system may also occur. It is not clear how cyanobacteria regulate nitrogen fixation under the 24-hour daylight of the austral summer, since it is a reaction under light-dark growth cycles inhibited by aerobic photosynthesis as nitrogenase is extremely oxygen-sensitive (Herrero, Muro-Pastor, & Flores, 2001). However, this may be due to the cyanobacterial circadian control, an adaptive advantage that allows cells to adjust their primary metabolism to the light cycle (Ng, Pointing, & Dvornyk, 2013). The overall findings support the notion of SRSCs-forming critical zones of activity, on or in the top few centimeters of rock and soil surface (Pointing & Belnap, 2012). Such critical zones may be relevant on a global scale, and they have been estimated to contribute a significant part of carbon and nitrogen fixation globally (Elbert et al., 2012).

The potential pathways of environmental and nutrient stress response were identified. The overall signal abundance of chasmoendoliths and soils were similar, in contrast to another GeoChip-based metagenomic study made at a colder drier inland site (McKelvey Valley) where the above-ground exposed niches (hypoliths and chasmoendoliths) supported greater and diverse stress response pathways than soil (Chan et al., 2013). This likely reflects that the substrate in inland valleys experiences more dramatic differences in microclimate than in less extreme lowland valleys. Conversely the number of phyla possessing osmotic, radiation-desiccation, heat shock and nutrient limitation stress response pathways was greater at the inland site (Chan et al., 2013). Overall, the less extreme Miers Valley supported relatively greater diversity in all pathways than McKelvey Valley. This may be due to relatively lower selective pressures resulting in multiple redundancies.

Since all phyla displayed multiple stress response pathways, the conclusion of this study is that environmental stress adaptation has been a major driver of microbial diversity in the Dry Valleys. Indeed, Cyanobacteria, Proteobacteria, Firmicutes, Actinobacteria and Bacteroidetes were prevalent in different Dry Valley communities (Lee et al., 2012; Pointing et al., 2009; Yung et al., 2014), and these organisms are known as salt-tolerant species (Oren, 2008) which display the capability to respond to osmotic stress. A further consideration is that this may also indicate a more profound difference in the way community stress tolerance is mediated: with individual taxa eliciting their own responses under more extreme stress, but in less extreme environments the community-wide benefits accruing from responses among a few taxa. A possible mechanism for this regulation is via quorum sensing communication using small signaling molecules, for example, *N*-acyl homoserine (Steindler & Venturi, 2007; Williams, Winzer, Chan, & Cámara, 2007). It has been demonstrated that in complex microbial communities with various functional capabilities, many bacteria use quorum sensing to mediate the community development, composition (d'Angelo-Picard, Faure, Penot, & Dessaux, 2005; Tan et al., 2015; Valle, Bailey, Whiteley, & Manefield, 2004) and function (Valle et al., 2004). In conclusion, the granite substrates (chasmoendoliths) were likely an important site for key geo-biochemical transformations, and fulfill complimentary ecosystem services to the surrounding soils.

Table 2.1. Diversity metrics for chasmoendolith and soil communities using DNA gyrase subunit B gene (*gyr*B) as a phylogenetic marker. A significant difference was observed for species richness (ANOVA: F=10.33, P=0.032, n= 6), Margalef's diversity (ANOVA: F=10.61, P=0.025, n=6) and Shannon's diversity index (ANOVA: F=12.72, P=0.031, n=6) between the two communities. No significant difference in the Simpson index and Pielou's evenness was observed.

	Chasmoendolith		Soil	
	Mean	Std. Dev.	Mean	Std. Dev.
Species Richness (S)	347	18.61	440	46.13
Shannon's Diversity Index (H')	5.834	0.049	6.071	0.106
Simpson Diversity Index (D)	0.986	0.003	0.997	0.002
Margalef's Diversity (Da)	59.12	2.68	72.26	6.34
Pielou's Evenness (J)	0.997	0.001	0.998	0.001

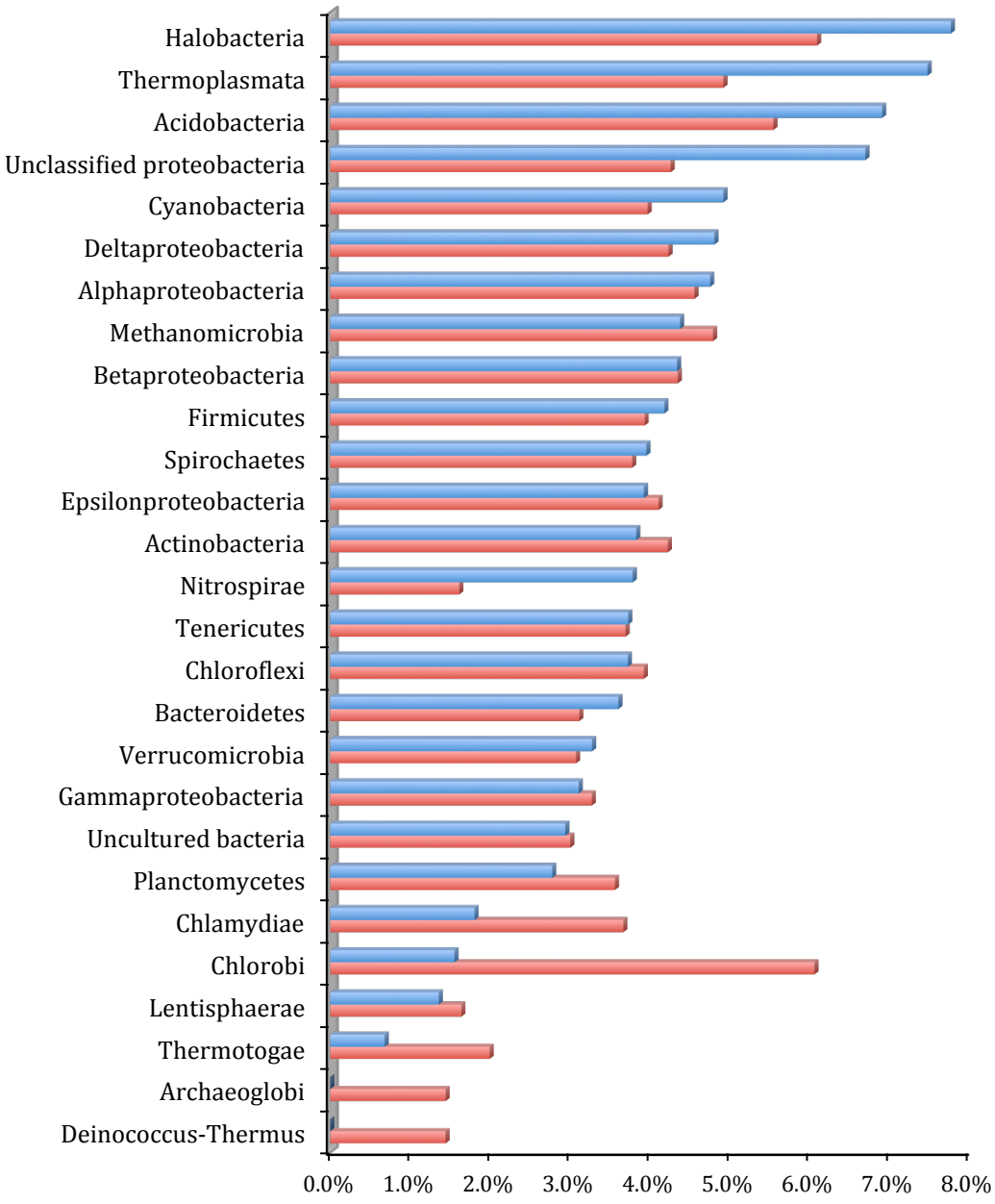


Fig. 2.1 Biodiversity estimates for chasmoendolith and soil communities using DNA gyrase subunit B gene (*gyrB*) as a phylogenetic marker. Blue bars denote chasmoendolith, red bars denote soil. Estimates were obtained by screening environmental DNA using the Geochip microarray.

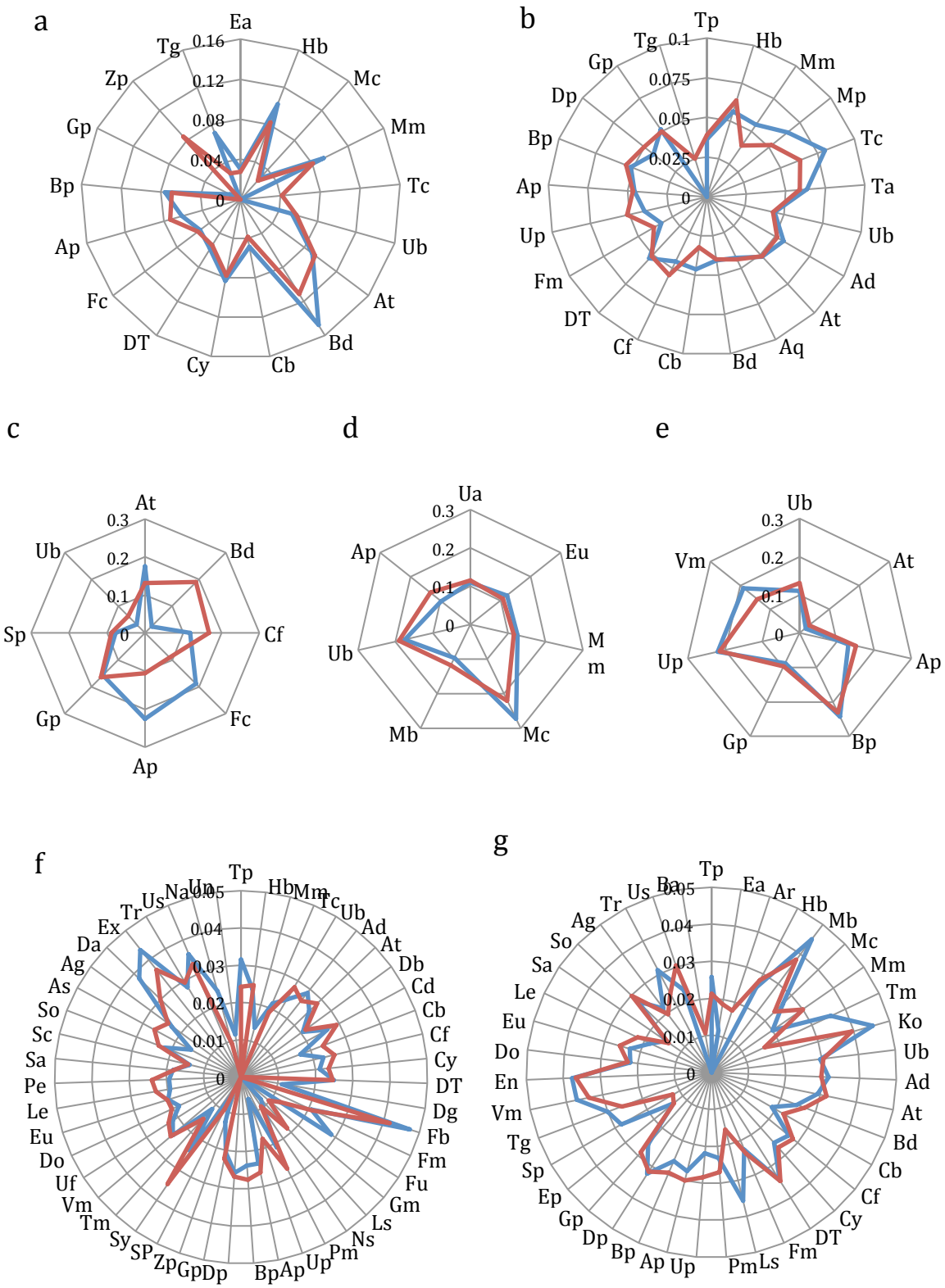


Fig 2.2 Taxa-function relationships for carbon cycling genes. The relative signal intensity was normalized for the number of probes per taxon. Mean values of samples from respective niches (chasmoendoliths [blue] and soils [red]) were plotted for the carbon cycling genes involved in **a** photoautotrophy, **b** chemoautotrophy, **c** acetogenesis, **d** methanogenesis, **e** methane oxidation, **f** carbon carbohydrates, **g** carbon aromatics and others.

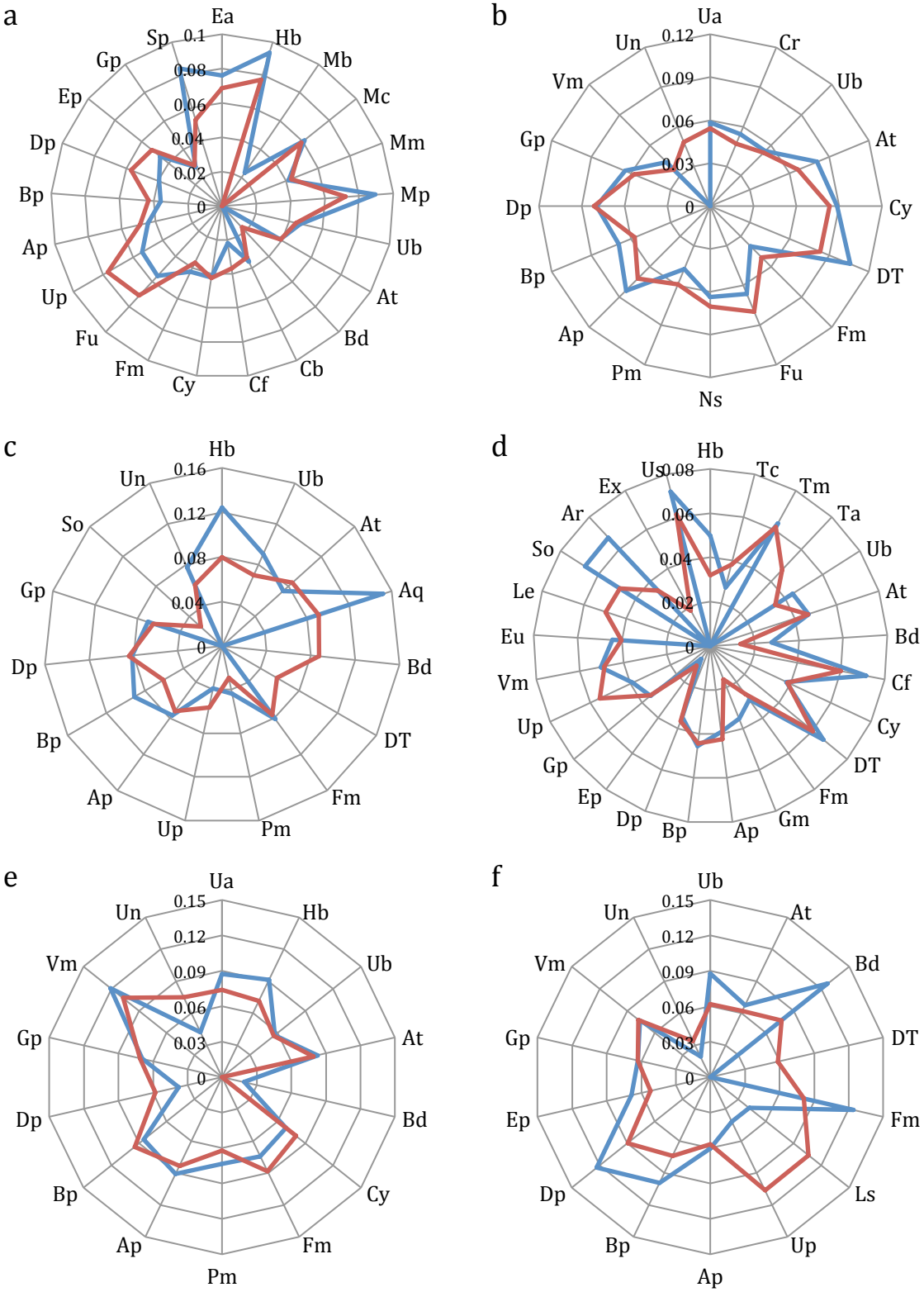


Fig. 2.3 Taxa-function relationships for nitrogen cycling genes. The relative signal intensity was normalized for the number of probes per taxon. Mean values of samples from respective niches (chasmoendoliths [blue] and soils [red]) were plotted for the nitrogen cycling genes involved in **a** nitrogen fixation, **b** nitrification, **c** denitrification, **d** ammonification, **e** assimilatory nitrate reduction, **f** assimilatory nitrate reduction.

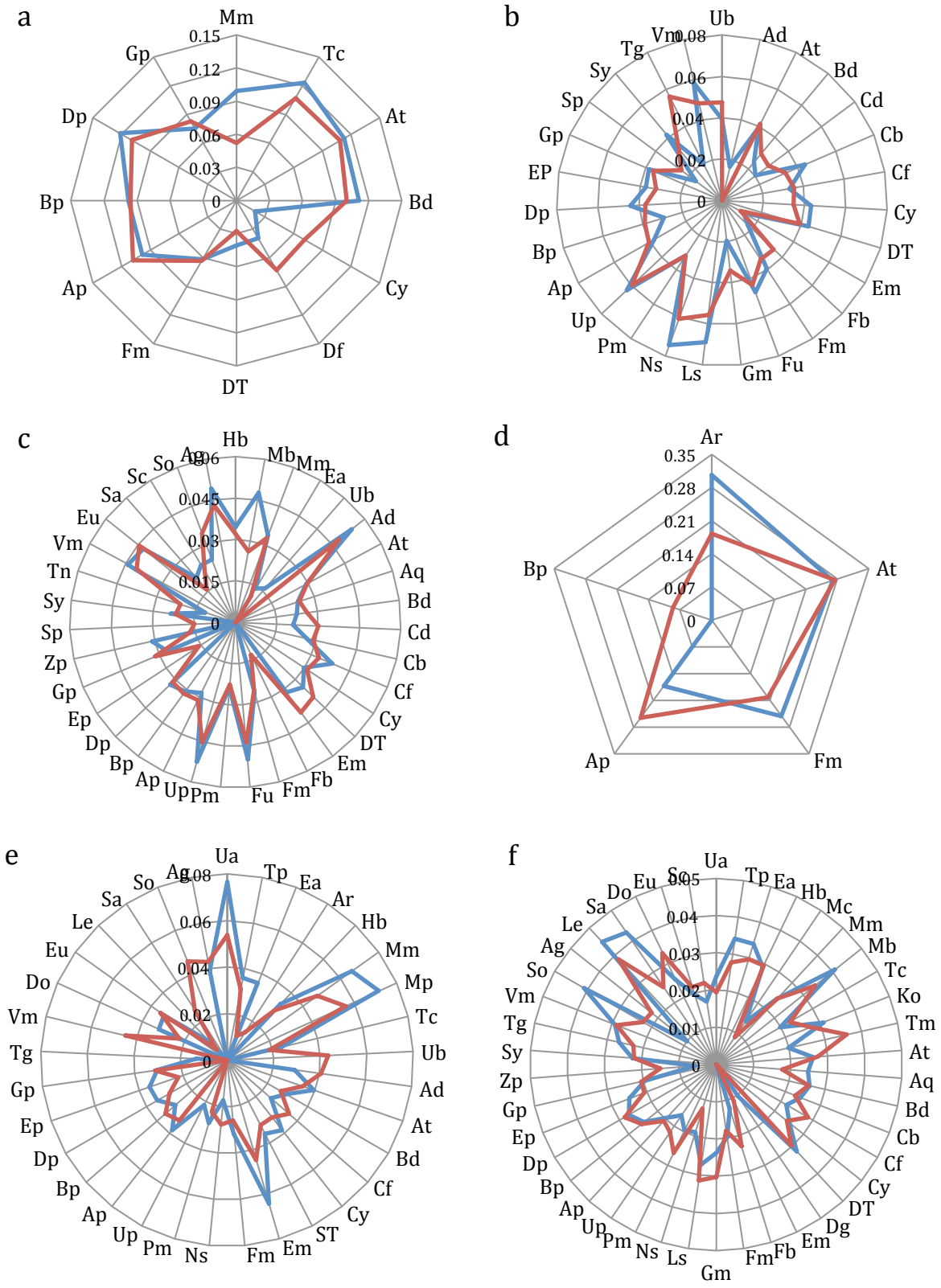


Fig. 2.4 Taxa-function relationships for stress response genes. The relative signal intensity was normalized for the number of probes per taxon. Mean values of samples from respective niches (chasmoendoliths [blue] and soils [red]) were plotted for the nitrogen cycling genes involved in **a** osmotic stress, **b** radiation stress, **c** heat shock, **d** cold shock, **e** nitrogen limitation, **f** phosphate limitation.

Table S2.1 Two-letter codes used to denote taxa in functional plots

Code	Taxon	Code	Taxon
Ad	Acidobacteria	Hb	Halobacteria
Ag	Agaricomycetes	Ko	Korarchaeota
Ap	Alphaproteobacteria	Le	Leotiomycetes
Aq	Aquificae	Ls	Lentisphaerae
Ar	Archaeoglobi	Mb	Methanobacteria
As	Ascomycota	Mc	Methanococci
At	Actinobacteria	Mm	Methanomicrobia
Ba	Basidiomycota	Mp	Methanopyri
Bd	Bacteroidetes	My	Mycetozoa
Bp	Betaproteobacteria	Na	Neocallimastigomycota
Cb	Chlorobi	Ns	Nitrospirae
Cd	Chlamydiae	Pe	Pezizomycetes
Cf	Chloroflexi	Pm	Planctomycetes
Ck	Candidatus Korarchaeum	Sa	Saccharomycetes
Cr	Crenarchaeota	Sc	Schizosaccharomycetes
Cy	Cyanobacteria	So	Sordariomycetes
Da	Dacrymycetes	Sp	Spirochaetes
Df	Deferribacteres	Sy	Synergistetes
Dg	Dictyoglomi	Ta	Thaumarchaeota
Do	Dothideomycetes	Tb	Thermobaculum
Dp	Deltaproteobacteria	Tc	Thermococci
DT	Deinococcus-Thermus	Tg	Thermotogae
Ea	euryarchaeote	Tm	Thermoplasmata
En	Entomophthoromycota	Tn	Tenericutes
Em	Elusimicrobia	Tp	Thermoprotei
Ep	Epsilonproteobacteria	Tr	Tremellomycetes
Eu	Eurotiomycetes	Ua	Uncultured archaea
Ex	Exobasidiomycetes	Ub	Uncultured bacteria
Ey	Euryarchaea	Uf	Uncultured fungi
Fb	Fibrobacteres	Un	Unidentified
Fc	Firmicutes	Up	Unclassifiedd proteobacteria
Fu	Fusobacteria	Us	Ustilaginomycetes
Gm	Gemmatimonadetes	Vm	Verrucomicrobia
Gp	Gammaproteobacteria	Zp	Zetaproteobacteria

Chapter 3 Diversity of metabolic and stress-tolerance pathways in hypolithic and soil communities of the McMurdo Dry Valleys

3.1 Introduction

The McMurdo Dry Valley terrain is covered with quartz, marble and other translucent stones embedded in the soil surface (Wynn-Williams, 1990), often support hypolithic colonization since the harsh abiotic stressors are filtered out by these substrates (Chan et al., 2012; Cowan, Pointing, et al., 2011). Several studies have revealed that hypoliths support greater biomass than soils (Khan et al., 2011; Pointing et al., 2009) and are dominated by cyanobacteria (Chan et al., 2012). The diversity of soil communities was also widely surveyed across several sites in the Dry Valleys with dominance of Acidobacteria and Actinobacteria (Lee et al., 2012; Pointing et al., 2009). Archaea and fungi comprised low biomass and diversity in soils (Pointing et al., 2009), associated primarily with *Thaumarchaeota* and *Debaryomyces* yeast, respectively (Rao et al., 2011; Richter et al., 2014).

Hypolithic and soil communities comprise a critical zone where most biological transformations (carbon and nitrogen fixation) occur in dryland ecosystem (Chan et al., 2012; Pointing & Belnap, 2012) since biodiversity and metabolic activity decrease largely in subsurface (15-20cm depth from surface) layers of soils (Stomeo et al., 2012). Limited individual *in situ* respirometry studies have implicated microbial activity in carbon and nitrogen transformations for the hypolith and soil system in the Dry Valleys (Cockell & Stokes, 2004; Cowan, Sohm, et al., 2011; Hopkins et al., 2006, 2009; Niederberger et al., 2012). Hypoliths were identified as important nitrogen sources to valley soils due to the presence of proteobacterial and cyanobacterial nitrogen fixation genes (*nifH*) (Cowan, Sohm, et al., 2011). There has been a general lack of cohesion between molecular genetic and substrate utilization studies, and so major questions

remain concerning this pathway.

Yergeau (2007) first used the GeoChip to investigate functional genes involved in several carbon and nitrogen cycling pathways in different soil communities crossing sub-Antarctica, but this study mainly focused on elucidating environmental factors driving the abundance and distribution of these genes among different soil communities, indicating that carbon fixation genes were more abundant in vegetation-lacking areas, nitrogen fixation genes were more prevalent in lichen-vegetated areas and denitrification genes were more dominant in higher temperature soils. Another GeoChip study identified the functional roles and stress response pathways of entire hypolithic and soil communities in the McKelvey Valley across bacteria, archaea and fungi domain (Chan et al., 2013).

Previous studies have addressed the taxonimic diversity and functional diversity of microbial communities separately. In this chapter, hypoliths and soils from maritime Miers Valley were interrogated for functional diversity of carbon and nitrogen transformation pathways and stress response pathways using GeoChip. High throughput 454 pyro-sequencing was also conducted for estimating bacterial diversity in these two niches. The combined analyses are important not only to identify the spatial distribution of genes in productivity and metabolism in these critical active zones, but also establish the relationships between taxonomic and functional diversity of soil and hypolithic communities in maritime Miers Valley.

3.2 Materials and Methods

Hypolith and soil samples were collected during a field survey conducted in January 2011 at a hypoliths dominated location in the central part of the Miers Valley (78°05.486’ S, 163°48.539’ E). Triplicate independent biomass was scratched from hypoliths. Hypolithics biomass (n=3) and triplicate soil samples (n=3) then transferred

to PowerSoil® DNA isolation kit (Mo Bio Laboratories Inc., Carlsbad, California, USA) for DNA extraction according to the manufacturer's protocol. DNA used for 454 pyrosequencing in the diversity survey (sequencing conducted by Dr. Subramanya Rao, at University of Hong Kong) was also used for GeoChip microarray hybridization (n= 6) in this study. GeoChip hybridization was carried out as previously described in section 2.2. In total 34,464 probes derived from genes involved in carbon, nitrogen cycling, and stress responses were used for this hybridization, which achieved of an average 50.6% of the total probes, covering 88.6% of the targeted genes of interest on GeoChip 4. The GeoChip dataset is publicly available at http://ieg.ou.edu/4download. Three major functional categories were targeted in this study: carbon cycling, nitrogen cycling and stress response (Tu et al., 2014). The normalized hybridization output data of each single gene was then grouped based upon the total signal intensity of each functional category. Visualization of different phylum-level and/or class-level contributions to each metabolic pathway was tabulated using heat maps, where each color block [dark green (above 99% percentile) to light yellow (below 70% percentile)] of the map was specific to a given phylum/class. Differences stated as significant were tested using one-way analysis of variance (ANOVA) on signal abundance of functional genes between two niches.

3.3 Results

3.3.1 Carbon transformation pathways

Carbon fixation was indicated by photoautotrophic and chemoautotrophic pathways for both hypolith and soil communities (Fig 3.1). Gene variants that are known to encode enzymes critical to both carbon fixation pathways were included. Of interest was the taxonomic breadth, with 12 bacterial and 8 archaeal phyla were involved. Acetogenesis

and methane oxidation appeared to be exclusively bacterial transformations: Proteobacteria, Actinobacteria, Bacteroidetes, Firmicutes and Chloroflexi; Proteobacteria and Verrucomicrobia respectively, whilst methanogenesis was predominantly an archaeal pathway (Methanococci and Methanobacteria in major). The ability to transform carbohydrate substrates was widespread among all archaeal, bacterial and fungal phyla, and surprising so was the ability to transform recalcitrant aromatic compounds. Significant differences were observed in certain carbon transformation pathways (Fig 3.2). Hypoliths supported significantly greater signal for cyanobacterial photoautotrophy (ANOVA: $F= 19.461$, $P< 0.01$) after excluding non-cyanobacterial *rubisco* genes, although the soils supported an overall greater carbon fixation signal than hypoliths. For acetogenesis pathways hypoliths displayed significantly greater signal intensity, whilst the soils supported larger signal for aromatic compound transformation (Fig 3.2).

3.3.2 Nitrogen transformation pathways

The key observation for nitrogen transformation was that the complete nitrogen cycle pathway was present in both hypolith and soils (Fig. 3.3). Three phyla (Actinobacteria, Firmicutes, Proteobacteria) possessed all pathways for the nitrogen cycle except ANNAMOX. Fixation of atmospheric nitrogen was an archaeal and bacterial pathway, with strongest signals from Halobacteria, Methanopyri, and Proteobacteria. Bacterial phyla that possessed nitrogenase genes for nitrogen fixation generally also possessed pathways for nitrification, although no nitrogenase genes were detected in *Deinococcus*-Thermus, Nitrospirae and Plantomycetes. Denitrification, assimilatory nitrate reduction and dissimilatory nitrate reduction (DNAR) were mainly bacterial pathways, but together with relatively higher signal from archaeal Halobacteria. The fungi only displayed pathways related to ammonification. There was no significant difference

between nitrogen fixation, nitrification and ammonification pathways between hypolith and soil communities. However, other significant differences between these two communities were observed: soils supported significantly greater signal for denitrification, assimilatory nitrate reduction and DNAR. The ANNAMOX pathway was only present in Planctomycetes (Fig 3.3) and was significantly greater in hypoliths (Fig 3.4).

3.3.3 Stress response pathways

Stress response pathways were the most variable among all functional pathways evaluated in the GeoChip analyses (Fig 3.5). Stress response pathways are specifically related to Antarctic environments, including osmotic, radiation (desiccation), heat and cold shock were identified, together with pathways of nutrient (nitrogen, phosphate and glucose) limitations, oxidative stress, and protein overexpression. The most prevalent functional genes were related to nutrient limitation and oxygen stress, and these were present in archaea, bacteria and fungi. Other pathways related to protein stress and glucose limitation were only recovered for bacteria. Strongest signals for osmotic stress tolerance were recovered for Actinobacteria, Bacteroidetes, Cyanobacteria, Firmicutes and Proteobacteria. Dessiccation-tolerance as evidenced by radiation stress response pathways was common to most bacterial phyla encountered in both niches. Heat shock was nearly ubiquitous whilst cold shock pathways were detected only for Actinobacteria, Firmicutes and Proteobacteria, plus the archaeal Archaeoglobi. There were significant differences between hypolith and soil communities. Hypolith communities displayed greater signal for glucose and phosphate limitation, whilst in soils abiotic stress pathways including radiation (desiccation) stress, heat shock and oxygen stress were more abundant (Fig 3.6).

3.4 Discussion

The above analysis of functional traits obtained from GeoChip for soil and hypolith communities revealed that the most abundant phyla namely Actinobacteria, Cyanobacteria, Proteobacteria and Firmicutes (Fig S3.1), also displayed greatest functional diversity. This metabolic plasticity may explain a central role for these abundant phyla as keystone taxa that perform the major primary transformations and facilitate the development of entire community richness (Chan et al., 2013). In addition, most of the functionalities were shared across many phyla in both communities, and this indicated some degrees of functional redundancy within the community that suggested the communities were highly resilient. The GeoChip analyses revealed that the bacteria displayed the highest functional diversity in all the targeted pathways for soil and hypoliths (Fig 3.1, 3.3 & 3.5). Combined with previous findings that bacteria were the most dominant microorganisms in hypoliths and soils (Khan et al., 2011; Pointing et al., 2009), this suggests that bacteria perform the majority of trophic transformations and are highly resilient to environmental stress.

Hypoliths clearly represent a concentration in the landscape for primary metabolic pathways, in particular for photosynthetic carbon fixation and carbohydrate utilization. This is consistent with the notion that hypoliths are considered islands of productivity among the generally depauperate desert soils (Pointing & Belnap, 2012). The major contributors to photoautotrophy were filamentous Oscillatoriales cyanobacteria and unicellular *Synechococcus*, matching the majority of cyanobacterial assemblage (*Oscillatoria* sp. and *Synechococcus* sp.) in the pyrosequencing data (Fig S3.1). Oscillatorian cyanobacteria are common to hypolithic communities in cold and polar deserts (Chan et al., 2012; Pointing et al., 2009; Wood, Rueckert, Cowan, & Cary, 2008), but have lower occurrence in hot desert hypoliths where the *Chroococcidiopsis*

were more prevalent (Bahl et al., 2011; Pointing et al., 2007; Stomeo et al., 2013; Warren-Rhodes et al., 2007). Interestingly, a relatively high level of *rubisco* signatures was also recovered for the genera *Synechococcus* and *Prochlorococcus*, the major primary producers in marine ecosystem and aquatic environments globally (Flombaum et al., 2013; Partensky, Hess, & Vaulot, 1999). Large areas of the valley floor hypolithic habitat in Miers Valley could become moisture sufficient and often submerged during the brief Antarctic summer, and thus display characteristics of hyporheic environments with aquatic taxa present (McKnight et al., 1999; Neiderberger et al., 2015).

Nitrogen transformation potential appeared to be widely distributed in both communities and this might reflect the nitrogen limiting nature of oligotrophic hypolithic and soil habitats in the Dry Valleys compared to nitrogen-sufficient endoliths (Friedmann & Kibler, 1980). High levels of cyanobacterial and proteobacterial nitrogenase were detected in this study, and this is consistent with a previous *in situ* study of acetylene reduction assay, demonstrating that Antarctic hypolithic communities were a potentially important source of nitrogen input to the Dry Valleys ecosystem (Cowan, Sohm, et al., 2011). In contrast to nitrogen input, nitrogen losses mediated by Planctomycetes (ANAMMOX pathway) were also identified in hypoliths and soils. However, the abundance of Planctomycetes in hypoliths and soils were low (Fig S3.1), indicating that the actual rates of ANAMMOX might be quite low. An *in situ* study of nitrogen cycling in desert biological soil crusts across the southwestern U.S.A has demonstrated that the ANAMMOX rate compared to nitrogen fixation was extremely low, perhaps due to the insufficient ammonium and nitrate in the soil crust (Strauss, Day, & Garcia-Pichel, 2011). Planctomycetes occur in both aquatic and soil environments. The dominant species in terrestrial habitats are *Candidatus Brocadia*, *Ca. Scalindua* and *Ca. Anammoxoglobus* (Sonthiphand, Hall, & Neufeld, 2014), supporting the recovered signatures for GeoChip and pyrosequencing analyses in this study.

In terms of stress response pathways, soil communities were more adapted to abiotic stressors than hypolithic communities. This may relate to the refuge status of the hypolithic niche (Pointing & Belnap, 2012; Wierzchos et al., 2012). It has been shown that the quartz substrate fro hypoliths, for example, effectively filtered out the most harmful UV irradiation (Cowan, Pointing, et al., 2011) and this may explain why radiation (desiccation) tolerant phylum *Deinococcus*-Thermus in hypolithic habitat displayed lower signature abundance for radiation stress. In addition, *Deinococcus* sp. and *Rubrobacter* sp. were relatively under-represented in this study (Fig S3.1) compared to other studies of cold desert hypoliths (Wong et al., 2010). This suggests that desiccation may not be the principal driver of diversity in the Miers Valley. However, Cyanobacteria, Deinococcu*s* and other bacteria displayed greater abundance of osmotic stress response genes in soils, reflecting that the bacterial and cyanobacterial community structure was significantly related to soil conductivity in Miers Valley (Lee et al., 2012; C. M. Magalhães et al., 2014). Hypoliths displayed more pronounced diversity of nutrient stress responses, suggesting that this more productive community was also adapted to nutrient fluctuation since cyanobacteria are periodically active; photosynthesis occurs under high relative humidity conditions (Palmer & Friedmann, 1990) and nitrogen fixation only occurs during darkness (Herrero et al., 2001).

In this study the functional ecology data obtained using the GeoChip microarray broadly matched the taxonomic diversity from sequence analysis, and this provided a strong 'eco-triangulation' for this dataset in that taxonomic data aligned well with predicted functionality for each phylum. In conclusion, Actinobacteria, Cyanobacteria, Proteobacteria and Firmicutes are the dominant phyla and also played important roles in carbon and nitrogen transformation in these two niches. The major differences in functionality were that soil communities displayed greater diversity in stress tolerance and recalcitrant carbon utilization pathways, whilst hypolithic communities supported

greater diversity of photoautotrophic carbon fixation and nutrient limitation adaptation pathways.

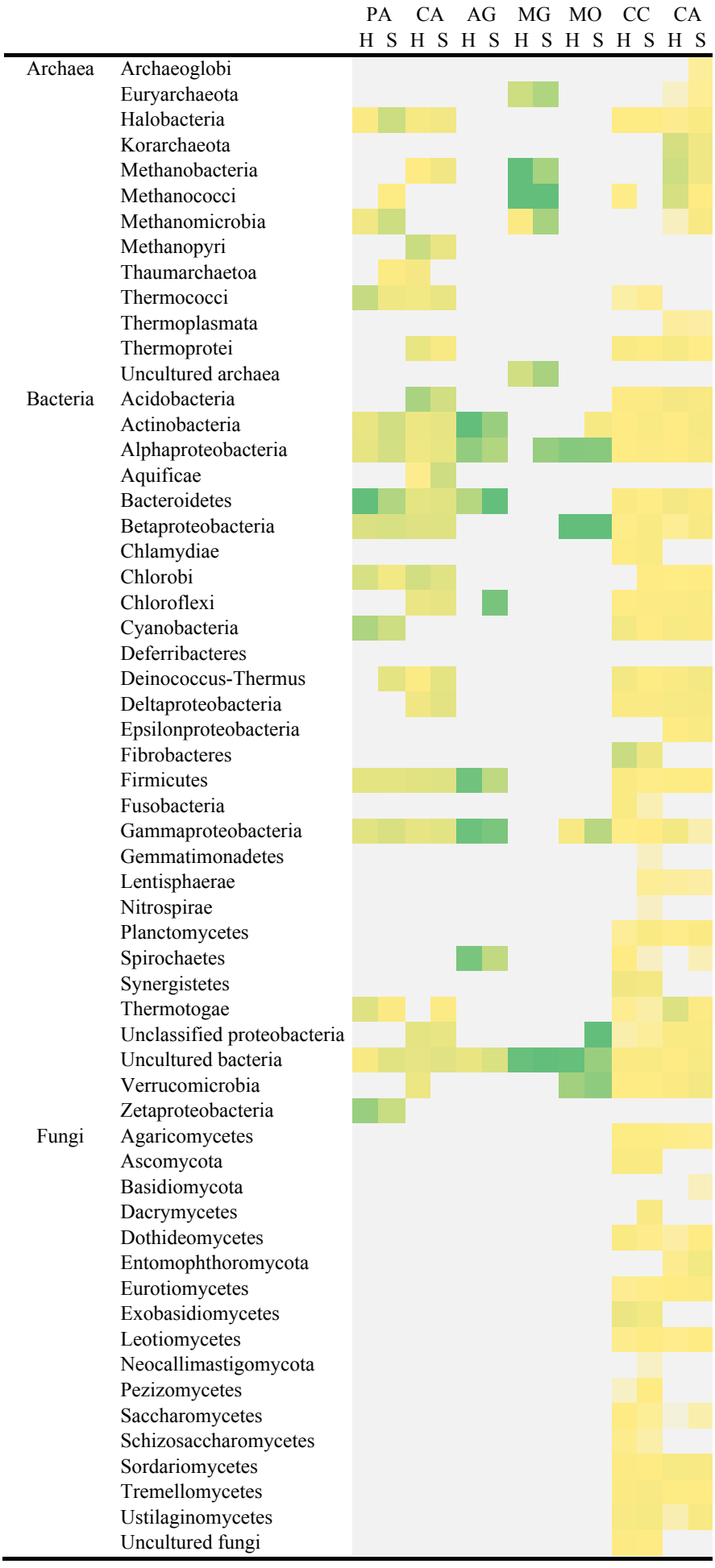


Fig 3.1 Distribution of pathways for carbon transformation.

H hypolith, S soil,
PA photoautotrophy,
CA chemoautotrophy,
AG acetogenesis,
MG methanogenesis,
MO methane oxidation,
CC carbon carbohydrate,
CA carbon aromatics.

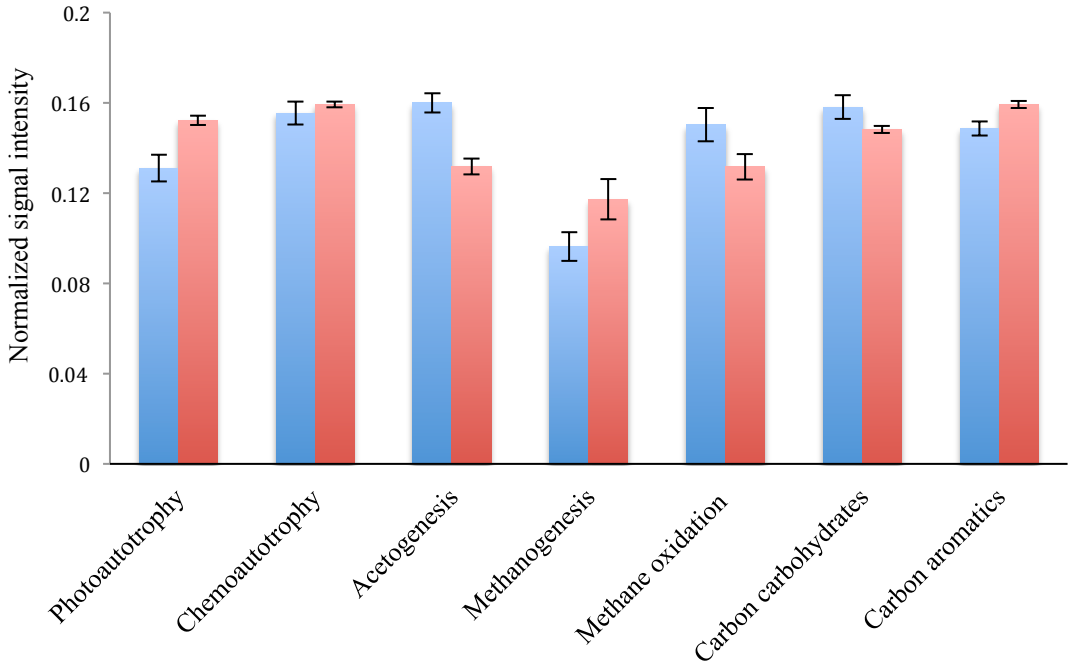


Fig 3.2 Overall abundance of different groups of carbon cycling genes. One-way analysis of variance (ANOVA) was tested to indicate confidence in difference. Blue denotes hypoliths, red denoted soils. There were significant difference in photoautotrophy ($F= 11.49$, $P< 0.05$), acetogenesis ($F= 26.11$, $P< 0.001$) and carbon aromatics ($F= 8.55$, $P< 0.05$). No significance difference was observed in chemoautotrophy, methanogenesis, methane oxidation and carbon crbohydrates.

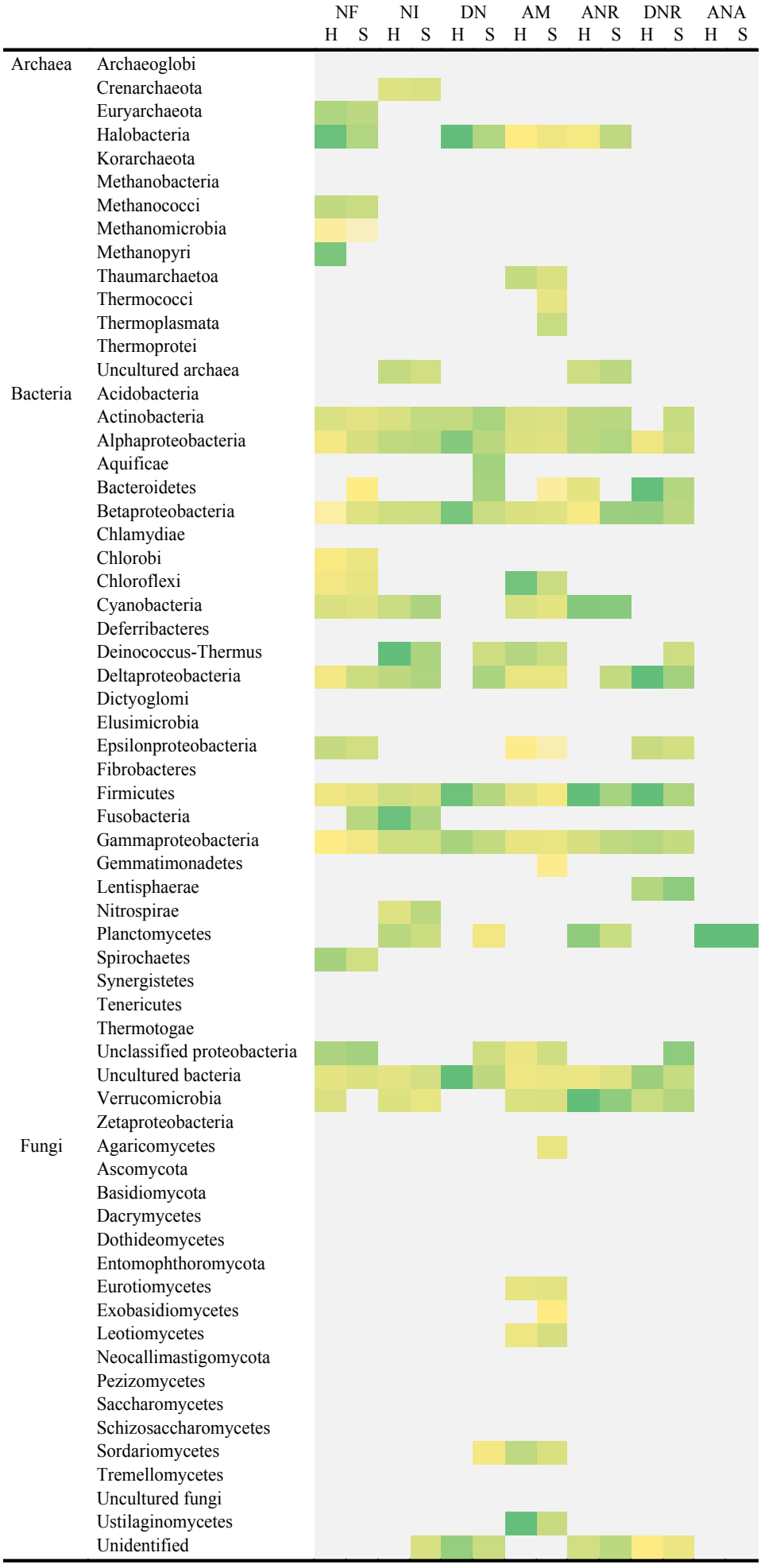


Fig 3.3 Distribution of pathways for nitrogen transformation.

H hypolith, S soil,
NF nitrogen fixation,
NI nitrification,
DN denitrification,
AM ammonification,
ANR assimilatory nitrate reduction,
DNR dissimilatory nitrate reduction,
ANA ANAMMOX

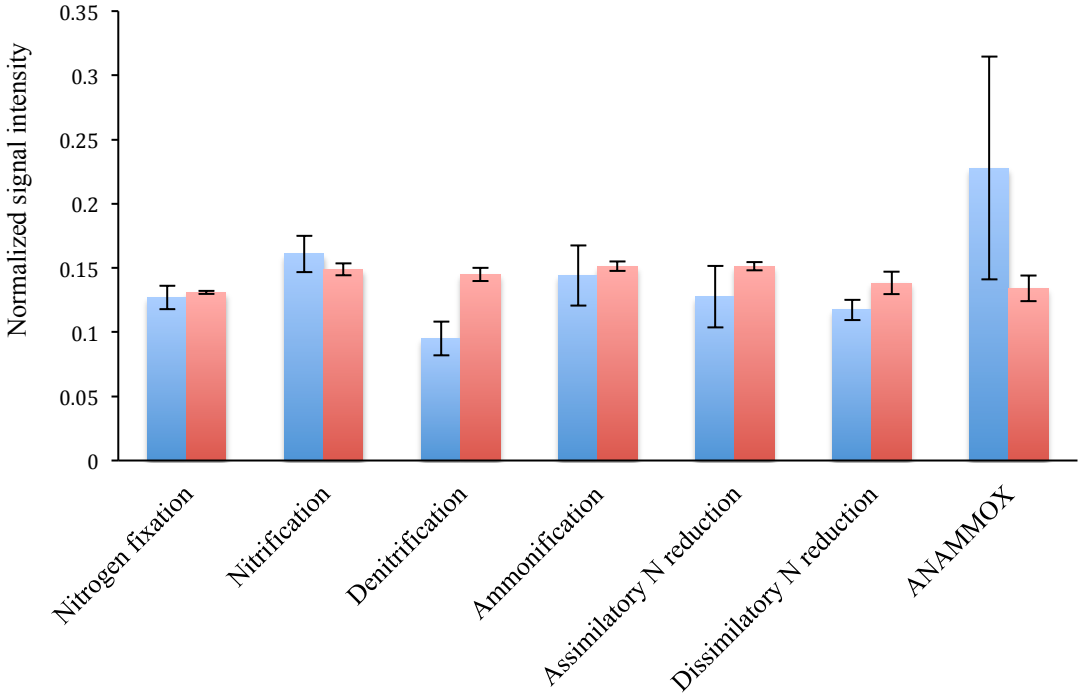


Fig 3.4 Overall abundance of different groups of nitrogen cycling genes. One-way analysis of variance (ANOVA) was tested to indicate confidence in difference. Blue denotes hypoliths, red denoted soils. Significant difference was observed in denitrification ($F= 111.63$, $P< 0.01$), assimilatory nitrate reduction ($F= 8.76$, $P< 0.05$), dissimilatory nitrate reduction ($F= 28.23$, $P< 0.01$) and ANAMMOX ($F= 10.37$, $P< 0.05$). No significant difference in nitrogen fixation, nitrification and ammonification.

Fig 3.5 Distribution of pathways for stress response. H hypoliths, S soils, OS osmotic stress, RA radiation stress, HS heat shock, CS cold shock, NL nitrogen limitation, Phosphate limitation, PS protein stress, SF sigma factor, GL limitation, OL oxygen limitation, OS oxygen stress.

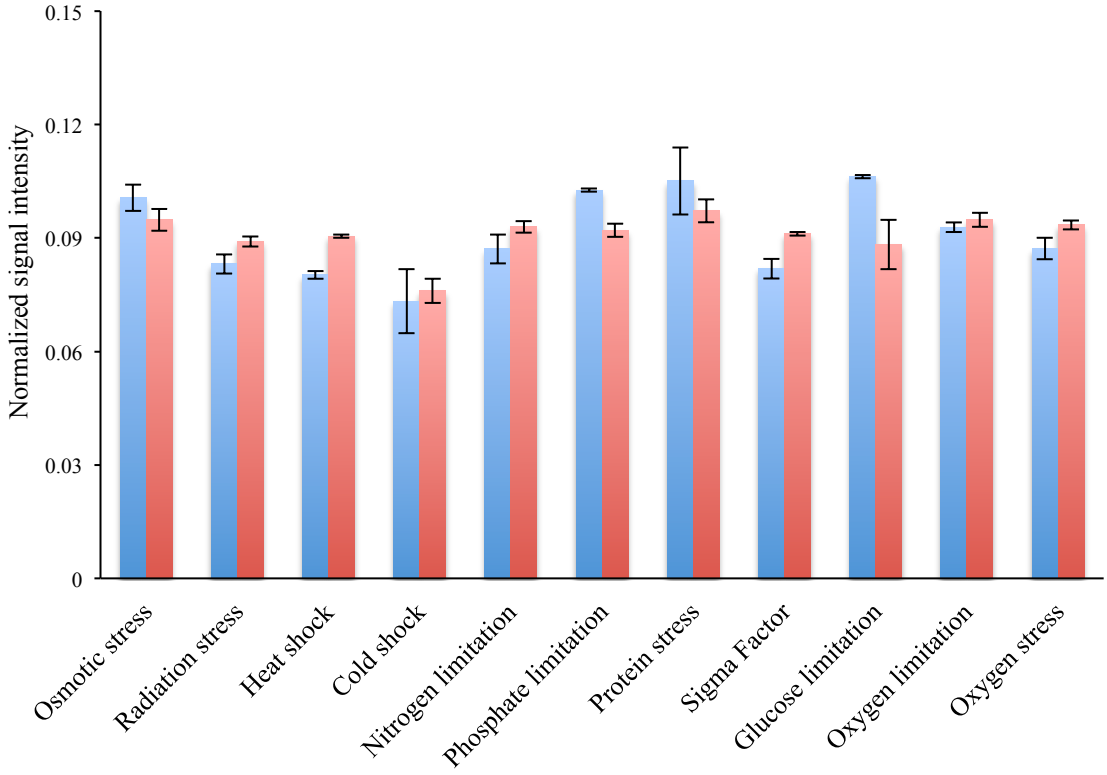


Fig 3.6 Overall abundance of different groups of stress response genes. One-way analysis of variance (ANOVA) was tested to indicate confidence in difference. Blue denotes hypoliths, red denoted soils. Significant difference was observed in radiation stress ($F= 12.83$, $P< 0.05$), heat shock ($F= 254.23$, $P< 0.01$), phosphate limitation ($F= 109.95$, $P< 0.01$), sigma factor ($F= 36.16$, $P< 0.01$), glucose limitation ($F= 22.37$, $P< 0.01$) and oxygen stress ($F= 12.89$, $P< 0.05$). No significant difference in osmotic stress, cold shock, nitrogen limitation, protein stress and oxygen limitation.

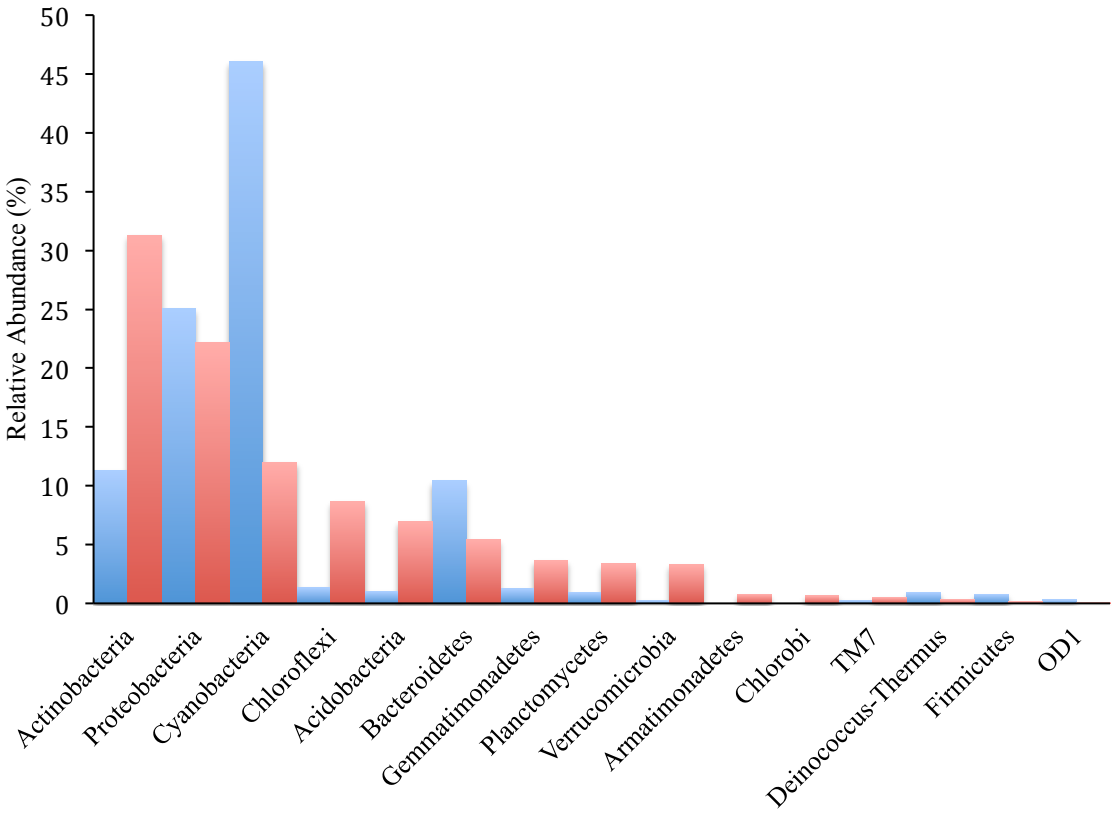


Fig S 3.1 Relative abundance of bacterial phyla in soil (red) and hypolith (blue) communities as determined by pyrosequencing of 16S rRNA genes. Other phyla present in soils at low abundance and not shown in the bar chart are: Elusimicrobia, Nitrospirae, OP3, WS3, WS5 and three unidentified phyla.

Chapter 4 Potential biotic controls on endolithic, hypolithic and soil communities in the McMurdo Dry Valleys

4.1 Introduction

Under the extreme abiotic stress of the McMurdo Dry Valleys, vascular plants and complex metazoans are unable to survive (Boyd et al., 1970; Cowan & Tow, 2004; Horowitz et al., 1972; Pointing et al., 2014), whilst lichens and mosses are visible as patchily distributed surface biota (De Los Ríos et al., 2014; Friedmann, 1982; Pointing et al., 2015). However, microbial communities in this extreme habitat comprise the most abundant standing biomass in cryptic niches: hypolithic, endolithic and soil communities (Pointing et al., 2015, 2014). The structures of these lithic communities have been widely identified and are dominated by cyanobacteria and lichens, but also support a diverse range of bacteria, archaea, chlorophytes and fungi (Chan et al., 2012; De Los Ríos et al., 2014; Friedmann, 1982; Pointing et al., 2009; Rao et al., 2011; Yung et al., 2014).

Soil nematodes and protozoa, primitive metazoan life, are ubiquitous and abundant in all ecosystems including deserts (Fierer et al., 2009). These two organisms help to regulate and maintain microbial diversity and biomass by grazing on and dispersing soil microbes; however, in the Dry Valleys the occurrence of nematodes and protozoa appears to be sparse, and so direct evidence of regulation of microbial abundance and diversity is limited (Wall & Virginia, 1999). This raises the question as to how persistent cryptic microbial communities that are protected from extreme physical disturbance and are not apparently grazed are regulated in terms of biomass turnover.

An important feature of microbial community dynamics under limited nutritional resources is interspecific interaction, an interaction between limited nutrient sources and

different bacterial population growth (Hibbing, Fuqua, Parsek, & Peterson, 2010). This leads to cooperative existence or domination among different bacterial species through competitive strategies. Nevertheless, an additional mechanism, non-transitive competition network, can contribute to the maintenance of diversity in environments. This is mediated by antimicrobial compounds (antibiotics), secreted by many bacterial phyla that inhibit the growth and activity of competing taxa (Hibbing et al., 2010). For example, under threatening environmental conditions, *Escherichia coli* strains that either produce antibiotics, are sensitive to or resistant to, but do not produce toxin that kills the other (Kerr, Riley, Feldman, & Bohannan, 2002). A study of psychrophilic bacteria isolated from the Dry Valley cryptoendoliths have shown their extracellular enzymatic activity and susceptibility to antibiotics *in vitro*, suggesting that this functionality allowed coexistence and stabilization of taxa within the community (Siebert et al., 1996). Nevertheless, nothing is known about the widespread ability to defend against antibiotics among the entire lithic and soil communities in the Dry Valleys.

Another factor that is also potentially important in regulating microbial communities is phage lysis (Pace & Cole, 1994). This mechanism, where viral infection causes cell lysis and, thus, exerts a control on microbial biomass, would represent a bottom-up control on microbial communities since predators (nematodes and protozoa) in Dry Valleys are sparse or absent. Phage populations have been interrogated in extreme environments, including hot springs (Breitbart, Wegley, Leeds, Schoenfeld, & Rohwer, 2004), Antarctic lakes (López-Bueno et al., 2009) and hypoliths from hyper-arid Namib desert (Adriaenssens et al., 2014). More recently, the phages of soil and hypolithic community in Miers Valley have been characterized, presenting prokaryotic and eukaryotic virus diversity together with potential novel species (Zablocki et al., 2014). Investigations of phage–host interaction, including cell lysis, have yet to be addressed in

Antarctica (Laybourn-Parry, 2009).

Apart from the abiotic factors (photosynthetically active radiation, soluble salts, soil water content, sulphur and nitrate) shaping microbial community structure in the Dry Valleys (Cowan, Pointing, et al., 2011; Lee et al., 2012; T. P. Makhalanyane, Valverde, Birkeland, et al., 2013; Pointing et al., 2009), little is known about the biotic factors controlling microbial communities in the Dry Valleys. In this chapter, the potential microbial regulation through interspecific competition and phage lysis was interrogated using GeoChip containing probes for antibiotic resistance genes and genes involved in phage structures and replication. These insights would indicate potential biotic factors and a bottom-up control in structuring microbial communities and abundance.

4.2 Material and methods

Four locations in three separate valleys were selected for this study. McKelvey Valley (77°26'S, 161°33'E, elevation approx. 800 m) is a high altitude inland valley that has no lake system, Victoria Valley (77°23'S, 162°70'E, elevation approx. 400 m) was used as a representative lake-bearing inland valley and Miers Valley (78°45'S, 163°15'E, elevation approx. 200 m) as a maritime lake-bearing valley. An additional hyporheic site (Kite Stream) in Victoria Valley (77°23'S, 162°E, elevation approx. 400 m) was also assessed, and this site experienced moisture sufficiency from glacial meltwater for a minimum one-month period during the 2012–2013 austral summer prior to sampling. Moisture content for all substrates was <2 % w/w with the exception of hyporheic samples where moisture content was >100 % w/w for all samples as measured gravimetrically in the field. DNA from the substrates (single sample of dry soil and cryptoendolith) collected from Victory Valley and its hyporheic site (single sample of wet soil and cryptoendolith) were extracted using PowerSoil® DNA isolation kit (Mo Bio Laboratories Inc., Carlsbad, California, USA) according to the manufacturer's

protocol.

DNA of open soil, hypoliths, chasmoendoliths, cryptoendoliths from McKelvey and Miers Valley recovered in previous studies were triplicates (Pointing et al., 2009; Yung et al., 2014). Microarray interrogation using GeoChip was conducted for all samples plus negative controls by Dr. Joy D. Van Nostrand and Dr. Jizhong Zhou at Department of Microbiology and Plant Biology, Institute for Environmental Genomics, University of Oklahoma, USA (Chan et al., 2013; Zhou et al., 2008). The GeoChip (He et al., 2010; Tu et al., 2014) contained probes for 11 antibiotic resistance coding loci of beta-lactamase, tetracycline, vancomycin and several transporter genes (Table 4.1) and 39 loci involved in host recognition, phage replication, phage structure and host cell lysis (Table 4.2).

4.3 Results

4.3.1 Antibiotic resistance genes

Antibiotic resistance genes were identified by hybridizing with different categories of probes. These were beta-lactamases, tetracycline and vancomycin resistance, together with a range of transporter genes associated with antibiotic resistance (Table 4.1). Overall, most samples returned signals from the microarray for all categories of antibiotic resistance genes. The exception was vancomycin gene absent in hyporheic wet soil (Fig 4.1). Signals for antibiotic resistance closely tracked observed bacterial and archaeal diversity, such that the stronger antibiotic resistance signals were associated with Halobacteria, Methanobacteria, Actinobacteria, Proteobacteria, Deinococcus-Thermus and photosynthetic bacteria (Chlorobi, Chloroflexi and Cyanobacteria) in all samples (Fig 4.2).

The patterns of antibiotic genes distribution were slightly discernible in these three

valleys. The stronger signals recovered from inland McKelvey Valley were mainly derived from Methanobacteria, uncultured bacteria, Actinobacteria, *Deinococcus*-Thermus, Beta- and Delta-proteobacteria (Fig 4.2a). In Miers Valley, antibiotic genes were nearly-ubiquitous among all taxa (Fig 4.2b). A striking difference between dry and wet substrates (soils and cryptoendoliths) from Victory Valley was apparent: the wetter substrates supporting greater relative abundance of antibiotic resistance genes from Halobacteria, Thermoprotei, photosynthetic bacteria (Chlorobi, Chloroflexi and Cyanobacteria) and Spirochaetes (Fig 4.2c & d). Notably, *Deinococcus*-Thermus displayed the strongest signal intensity from all substrates in McKelvey Valley (Fig 4.2a).

4.3.2 Bacteriophage diversity

A range of target loci was used to identify phages, including single-stranded (ss) and double-stranded (ds) DNA phage and ssRNA phage. It is interesting to note that ssDNA viruses were not present in all locations (Fig 4.3); however, these analyses revealed phage spanning eight families in diverse substrates and locations (Fig 4.4). The *Leviviridae*, *Myoviridae*, *Podoviridae* and *Siphoviridiae* families were ubiquitous in all substrates and all locations; an unidentified group within the *Caudovirales* was detected only at the dryer inland site whereas the *Corticoviridae* (dsDNA) only occurred in hyporheic soil (Fig 4.4b); the *Microvirida*e (ssDNA) occurred only in maritime Miers Valley and hyporheic soils but not the inland McKelvey Valley.

4.4 Discussion

The interrogation of antibiotic resistance genes using GeoChip in soil and rock microbial community among different sites of Dry Valleys revealed that these genes were widespread, indicating potential for interspecific competition in these microbial

communities. The distribution and abundance of antibiotic genes in this study basically reflected the microbial community diversity. It is interesting that wetter and dry substrates in Victory Valley (Fig 4.2 c & d) supported antibiotic genes recovered from different taxa, and this indicates that composition of microbial communities may be influenced by moisture status of the substrates. It has been suggested that community structure in maritime Antarctic soils were significantly influenced by soil water content which was enhanced by vegetation cover (Yergeau & Kowalchuk, 2008). Also, soil community structure of four geographically distinct Dry Valley has shown the correlation with water content (Lee et al., 2012). More recently, the microbial diversity of several hyporheic sites has been surveyed and indicated that wet soils supported higher abundance of chlorophyll *a* and cyanobacteria than dry soil (Neiderberger et al., 2015). In McKelvey Valley Deinococcus-Thermus displayed the strongest signal intensities. This is congruent with previous studies suggesting that Deinococcus-Thermus were only specific to drier (Neiderberger et al., 2015) and low-productivity site (Niederberger et al., 2008). This GeoChip data implied that antibiotic resistance genes might also result in changes in microbial interaction and structure; however, a metagenomic analysis of soils worldwide including Dry Valley soil has demonstrated that antibiotic resistance genes are widespread on a global scale, but are relatively less abundant in desert soils than non-desert soils, suggesting that abiotic conditions, not competitive interactions, were more important in shaping the desert bacterial communities (Fierer et al., 2012).

The majority of antibiotic resistance signals identified in GeoChip analysis were of bacterial origin (He et al., 2007; Tu et al., 2014); however, one fungal signature (Eurotiomycetes) was detected in this study, even though antibiotics are produced by both bacteria (Hibbing et al., 2010) and fungi (Shearer, 1995) during interspecific competition. It is noted that previous studies have reported substantial discrepancies in

the diversity and abundance of fungal phylotypes (Arenz & Blanchette, 2011; Rao et al., 2011), and so it is interesting to speculate whether inter-domain level competition would be a factor limiting fungi in the Dry Valley niches. These GeoChip analyses, however, were unable to identify any quorum sensing mechanism, a bacterial communication system of competitive or cooperative signals conveyed between groups of bacteria or between bacteria and the host (Sifri, 2008; Williams et al., 2007). It is also a major regulator of antibiotic production and release in bacteria cells (Hibbing et al., 2010). It would be interesting to study interspecific competition in the relatively low diversity and abundance community in Dry Valley, as antibiotics only being triggered once the numbers of threatening populations are achieved (Hibbing et al., 2010).

The phage diversity was identified using GeoChip with genes involved in phage structure, replication and host recognition. The most common phage families were *Leviviridae*, *Myoviridae*, *Podoviridae* and *Siphoviridiae* among the Dry Valleys (Fig 4.4). This is consistent with a recent study that identified great abundance of *Siphoviridae* (53 %), *Myoviridae* (21 %), *Podovirdae* (9%) in Dry Valley hypoliths and soils (Zablocki et al., 2014). The *Podoviridae* have a diverse host range that includes several cyanobacterial genera (Hulo et al., 2011). The *Levriviridae*, the only ssRNA virus signature recovered in this study, are found worldwide, and the isolated strains infect Gram-negative bacteria (Hulo et al., 2011). This study also revealed some interesting patterns regard to phage distribution. The *Corticoviridae* signature was only recovered in wet hyporheic soil, suggesting that they infect microbial hosts that are active only during the moisture sufficiency. A possible explanation is that most *Cortivirus*, a genus of viruses in the family *Corticoviridae*, are of marine origin (Hulo et al., 2011). Also, the lack of *Microviridae* in McKelvey Valley indicates a host range that spans taxa occurring in only wetter or ephemerally wet habitats. Indeed, the *Microviridae* viruses infect intracellular parasitic bacteria such as Bacteroidetes (Hulo

et al., 2011), which comprised lower abundance in inland valley (Pointing et al., 2009) than maritime valley and hyporheic sites (Lee et al., 2012; Neiderberger et al., 2015). This is supported by the notion that the abundance of phage groups in a given environment reflects the abundance of community members (Breitbart & Rohwer, 2005). Nevertheless, the cyanophage (under the family *Podoviridae*) were expected to be dominant in all rocky substrates and hyporheic soils due to the greatest abundance of cyanobacteria in these bacterial communities (Neiderberger et al., 2015; Pointing et al., 2009; Yung et al., 2014). A possible explanation for low abundance of the cyanophage was the low presence of soil cyanobacterial phage in public databases which are generally dominated by aquatic phage (Padan & Shilo, 1973; Sullivan, Waterbury, & Chisholm, 2003; Xia, Li, Deng, & Hu, 2013). Unlike the classification for prokaryotes and eukaryotes based on the phylogenetic markers (16S rRNA genes and 18S rRNA genes, respectively), virus classification is based on types of viral genome, viral structural protein, host range and their short genomic sequences (Buchen-Osmond, 2003; Lwoff, Horne, & Tournier, 1962), thus, classification of phage might be incomplete. Another reason might be due to the putative role of cyanobacterial extracellular polymer substance as a defensive mechanism to prevent phage adhesion (Looijesteijn, Trapet, de Vries, Abee, & Hugenholtz, 2001; Pereira et al., 2009; Xu, Khudyakov, & Wolk, 1997).

Given the paucity of information available on viral diversity in Antarctic environments, this study represents an important initial step in revealing the extent of phage diversity in Antarctic edaphic and lithic systems, and also indicates a possible bottom-up trophic regulation on the Dry Valleys edaphic communities. The recovery of viral signature in this extreme environment is consistent with the suggestion that phage distribution is ubiquitous on Earth (Breitbart & Rohwer, 2005). A recent study of the Dry Valleys metavirome indicated that, in addition to virulence, disease and defence

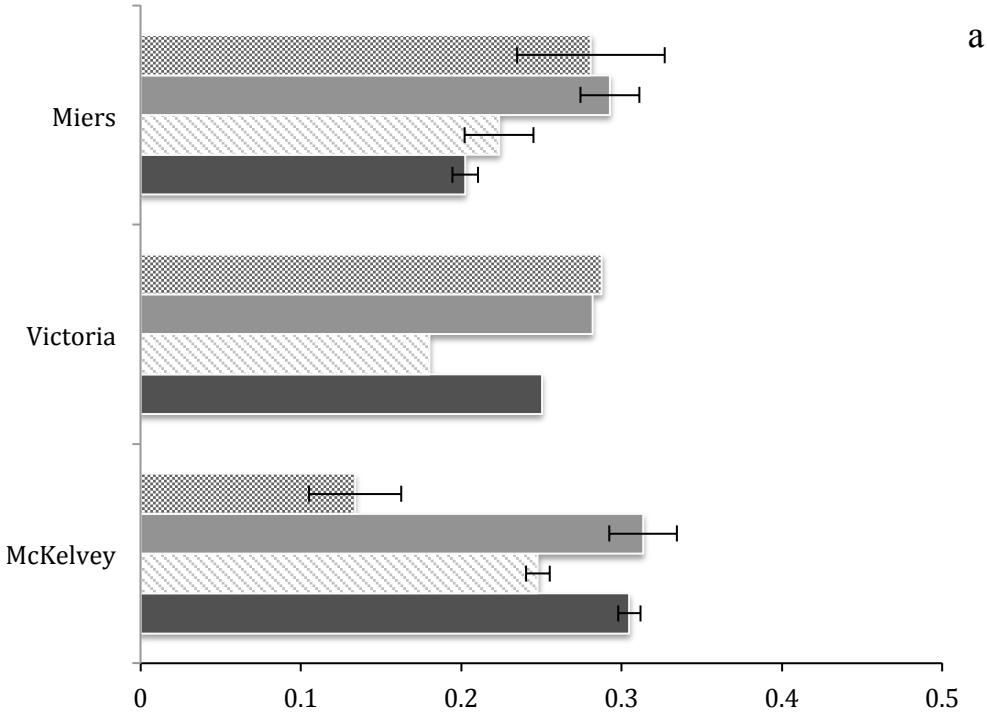
function genes, a range of genes involved in microbial physiological pathways may also be encoded within phage genomes (Zablocki et al., 2014). Thus, the high viral loads in Dry Valley microbial communities may also reflect the acquisition of favourable auxiliary metabolism genes (Roossinck, 2011) and this may increase the bacterial diversity by selective resistance traits (Koskella & Brockhurst, 2014).

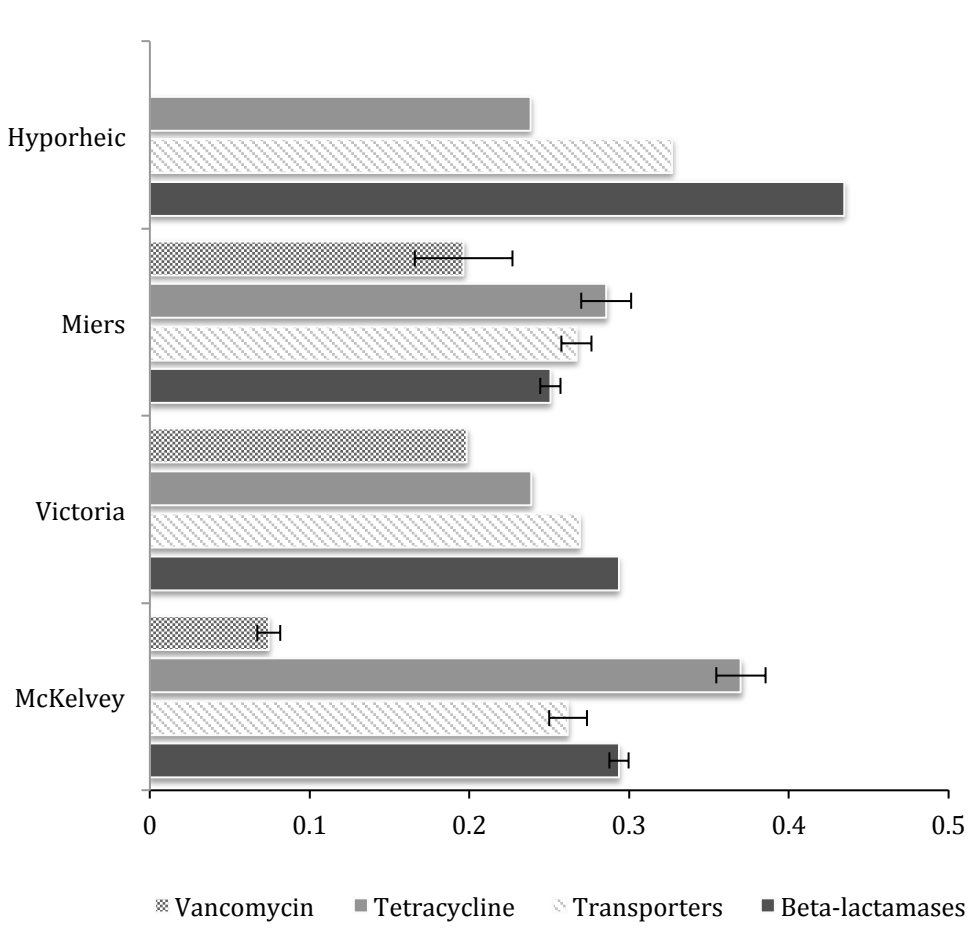
Table 4.1 Functional loci targeting antibiotic resistance in the GeoChip 4 microarray.

Antibiotic resistance	**Transporters**
β lactamase A	ATP-binding cassette transporter
β lactamase B	Multidrug toxic compound extrusion transport
β lactamase C	Mex pump
Tetracycline^{R}	Major facilitator efflux pump
Vancomycin^{R}	Small multidrug resistance efflux pump

Table 4.2 Functional loci targeting phage in the GeoChip 4 microarray

Host recognition	Replication	Structure	Lysis
T2 Long-tail fiber protein P38	Dda-like Helicase family-1	Contractile central-tail tube	Endolysin glycosidase
T4 Long-tail fiber protein P37	DNA ligase	Contractile tail sheath protein	Endolysin transglycosylase
	DNA polymerase type I	Major capsid protein	Type 1 holin
	DnaB-like Helicase family-4	Non-contractile major tail protein	Type 2 holin
	Helicase family-4	Scaffold protein	Type 3 holin
	Large subunit terminase	Tape measure protein	Lysin
	Linear DNA phage terminal protein		Mycobacterium lipase LysB
	P4 alpha type helicase		
	Primase		
	RNA dependent RNA polymerase		
	Serine integrase		
	Single strand annealing protein		
	Single-stranded DNA-binding protein		
	Small subunit terminase		
	T4 Clamp loader ATPase		
	T4 portal protein		
	T4 recombination endonuclease		
	T4 RNase H		
	T4 sliding clamp		
	T4 transcription activator MotA		
	T4 UvsW		
	T5 genome internalization A1		
	Theta replication initiator protein		
	Tyrosine integrase		


a
Miers
Victoria
McKelvey
0
0.1
0.2
0.3
0.4
0.5


b
Hyporheic
Miers
Victoria
McKelvey
0
0.1
0.2
0.3
0.4
0.5
Vancomycin
Tetracycline
Transporters
Beta-lactamases

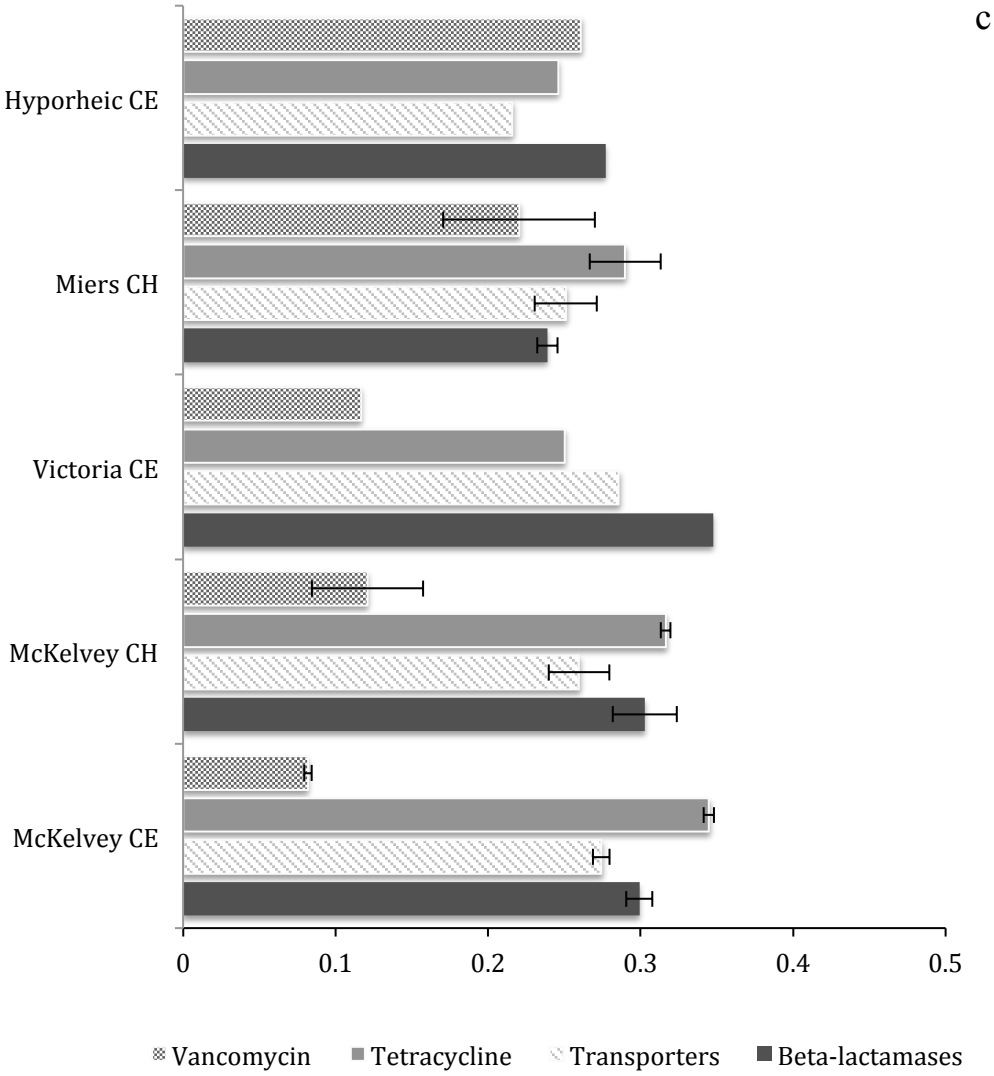


Fig 4.1 Categories of antibiotic resistance genes recovered from specific substrates at different sites of Dry Valleys. **a** hypolithic communities; **b** soil communities; **c** endolithic communities (CH chasmoendoliths, CE cryptoendoliths). Error bars represent standard deviation of the mean for three independent GeoChip replicates, where these are not visible standard deviation is less than the minimum scale increment.

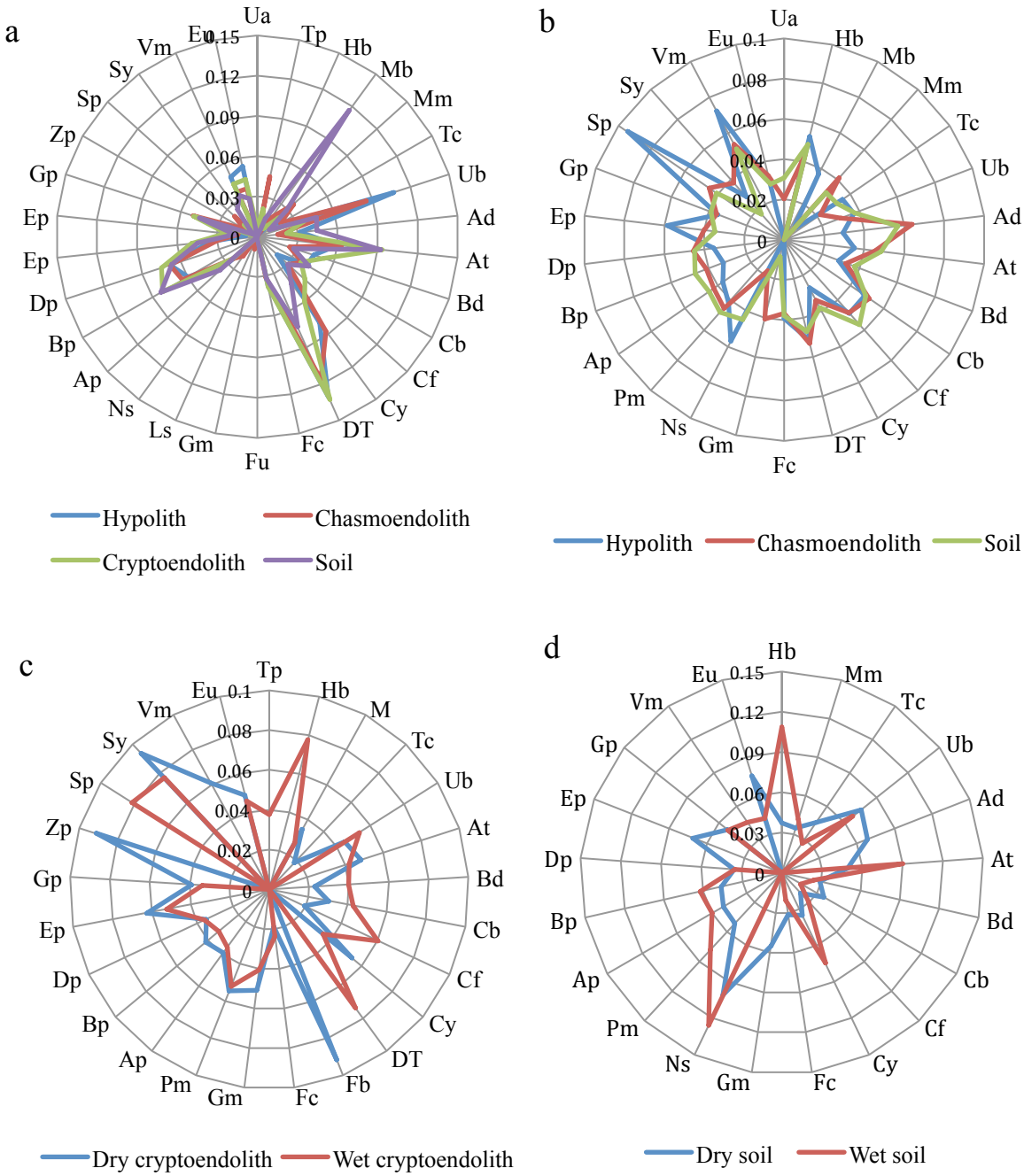


Fig 4.2 Relative occurrence of antibiotic resistance genes by phylum for bacteria, archaea and fungi from different specific substrates at different sites of Dry Valleys. **a** McKelvey Valley; **b** Miers Valley; **c** cryptoendoliths from Taylor Valley; **d** soils from Taylor Valley. Table S2.1 provides two-character codes for the microbial taxa for this figure.

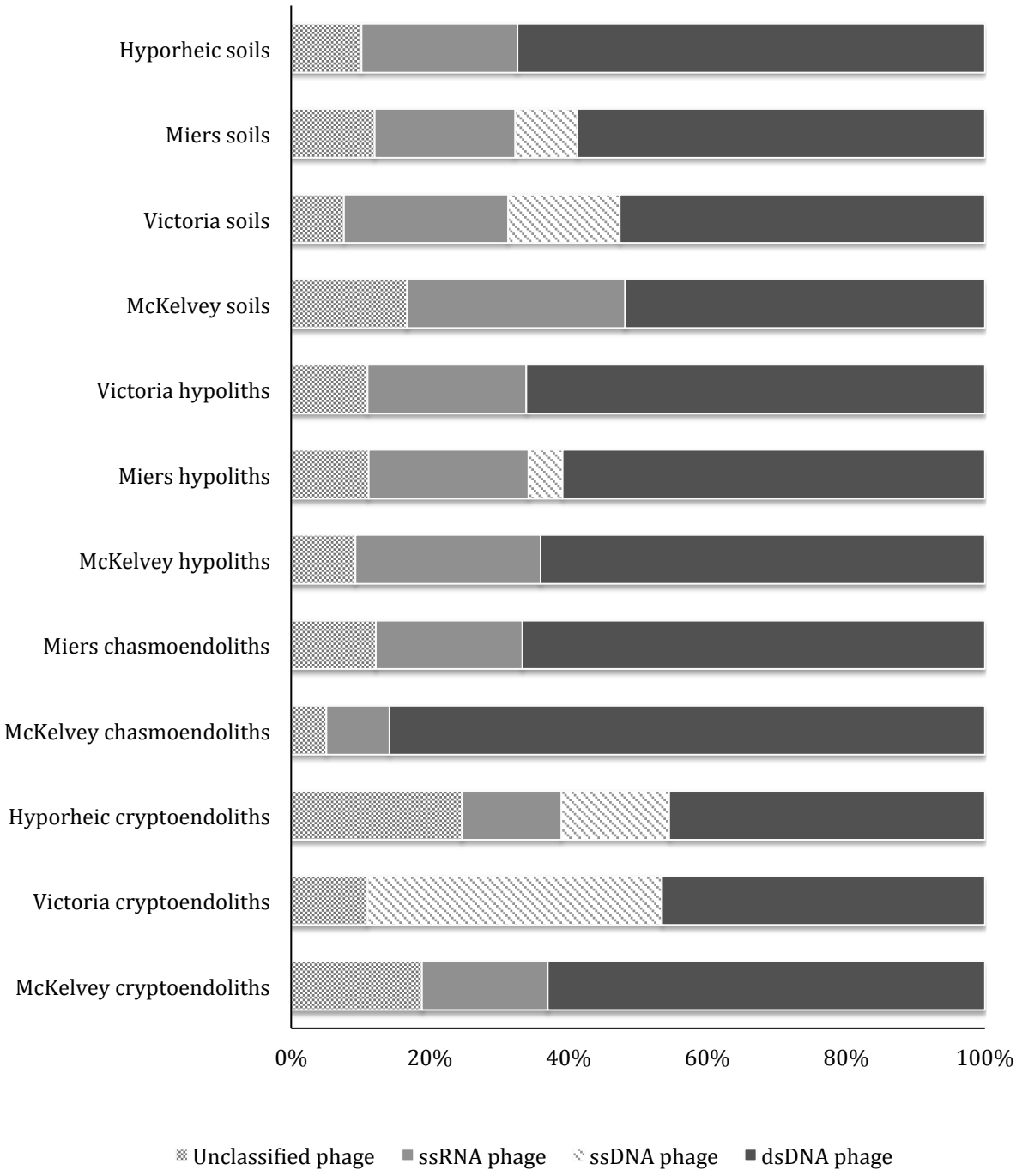
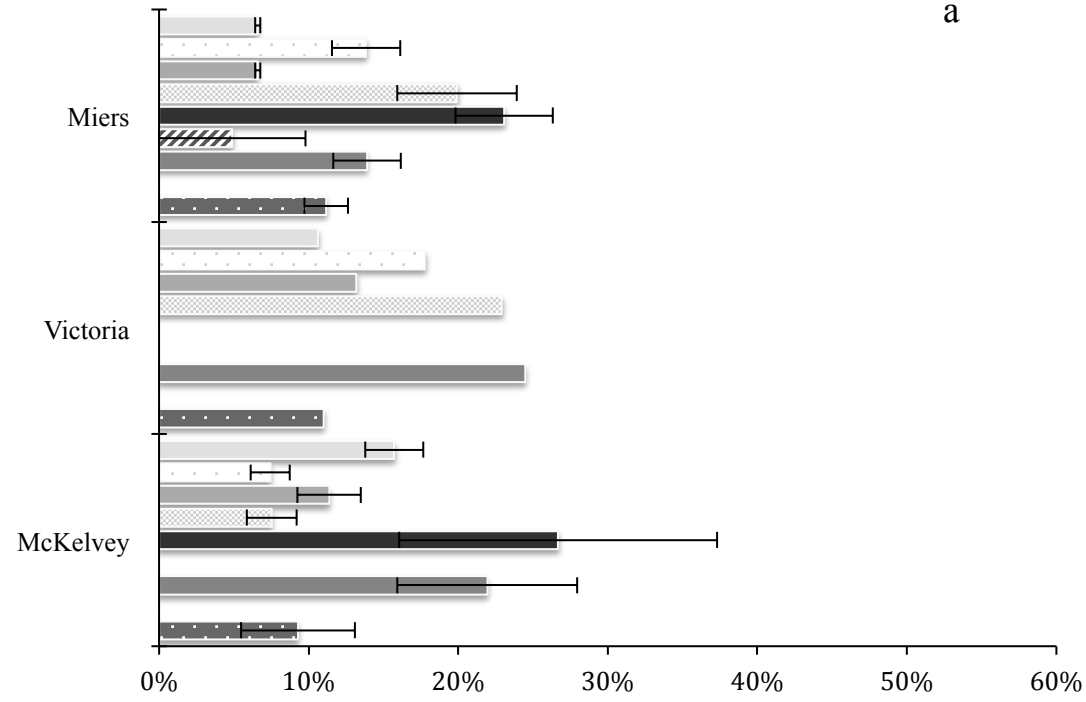
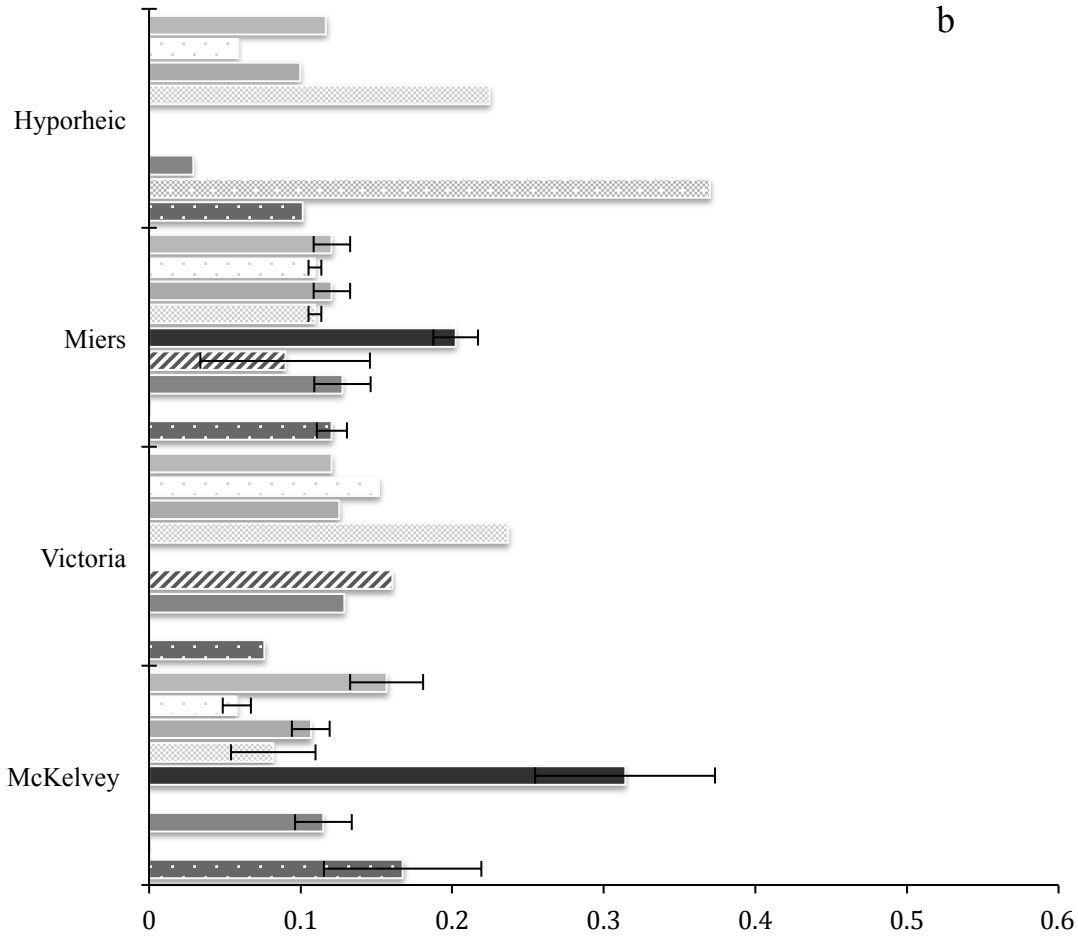


Fig 4.3 Categories of phage genomes recovered from Dry Valleys soil and lithic communities.


a
Miers
Victoria
McKelvey
0%
10%
20%
30%
40%
50%
60%


b
Hyporheic
Miers
Victoria
McKelvey
0
0.1
0.2
0.3
0.4
0.5
0.6

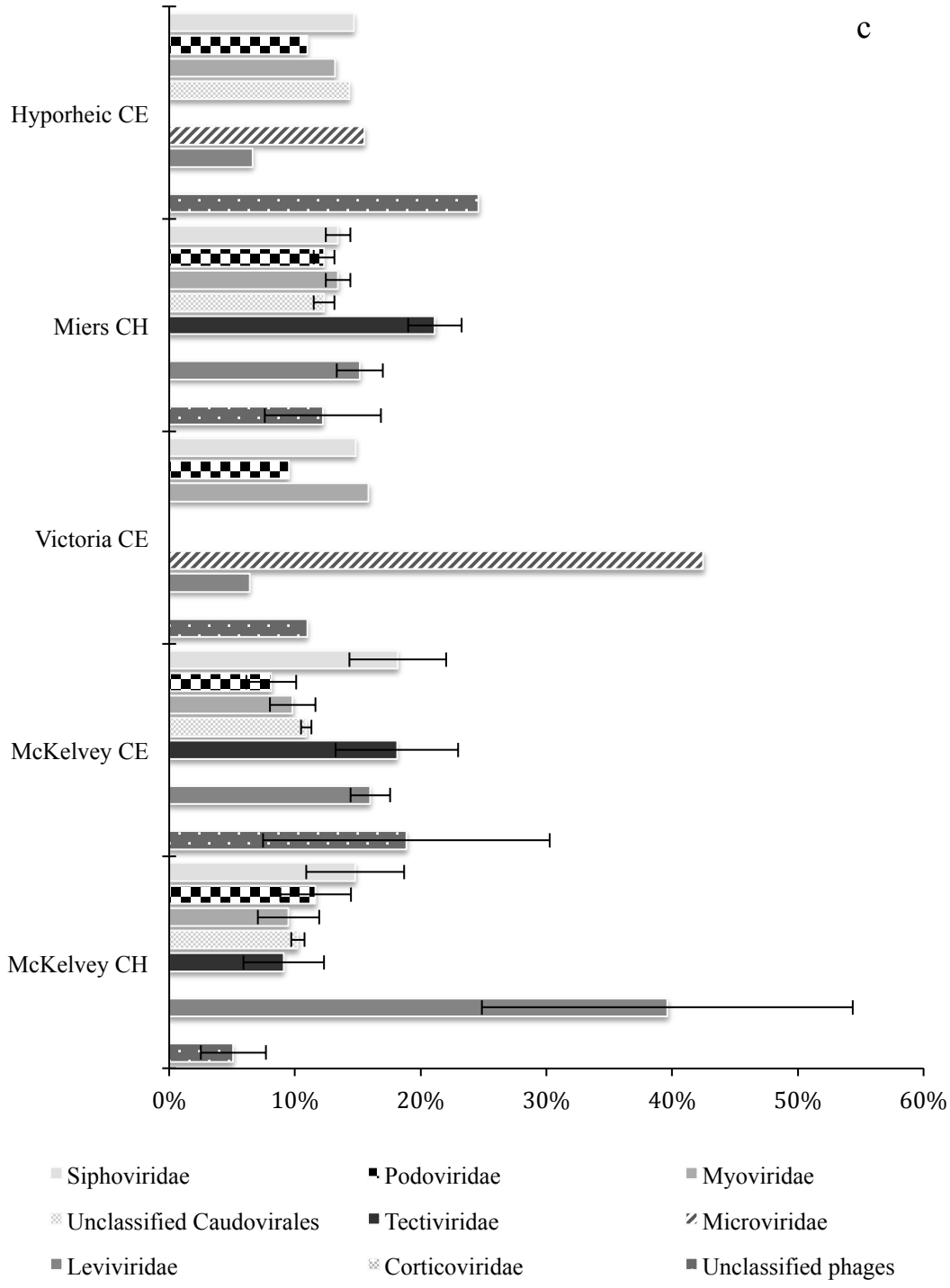


Fig 4.4 Phage families recovered from specific substrates at different dry valleys locations. **a** hypoliths; **b** soils; **c** endolithic communities (CH chasmoendoliths; CE cryptoendoliths). Error bars represent standard deviation of the mean for three independent GeoChip replicates, where these are not visible standard deviation is less than the minimum scale increment.

Chapter 5 Identification of molecular markers for genes encoding extracellular polymeric substance secretion pathways in desert cyanobacteria

5.1 Introduction

Cyanobacteria are the largest and most diversified group of prokaryotes (Whitton & Potts, 2012). They are divided into five subsections based on their great morphological diversity. Subsections I and II comprise unicellular species, whereas subsections III and VI are made up of filamentous cell forms without and with differentiated akinete and heterocyst cells, respectively. Subsections V represent the most complex cyanobacteria morphologically with filamentous branches (Roger Y Stanier, Deruelles, Rippka, Herdman, & Waterbury, 1979; Whitton & Potts, 2012). Because of the features of photoautotrophy (Calvin & Benson, 1948; Stanier & Cohen-Bazire, 1977) and diazotrophy (Stanier & Cohen-Bazire, 1977), and the low nutrient requirements, cyanobacteria are essential to carbon and nitrogen cycling in ecosystem, and are ubiquitous in different environments on the Earth (Seckbach & Oren, 2007; Whitton & Potts, 2012).

In drylands, microbial communities in soil crusts and lithic habitats are dominated by cyanobacteria (Bahl et al., 2011; Belnap et al., 2003; Chan et al., 2012; De Los Ríos et al., 2014; Pointing et al., 2015; Pointing & Belnap, 2012; Wierzchos et al., 2012). Filamentous *Microcoleus* sp. dominate biological soil crusts and coccoid *Chroococcidiopsis sp.* dominate hypoliths (Chan et al., 2012) and endoliths in warmer deserts (DiRuggiero et al., 2013; Dong et al., 2007; Wong et al., 2009). Apart from the importance of the carbon and nitrogen cycling, cyanobacteria also largely contribute to biogeological processes in dryland ecosystems. Desert landscapes are mostly covered by rocky substrates (Laity, 2008), and these minerals have been weathered by endolithic

cyanobacteria, resulting in the detachment of rock flakes from surfaces by an exfoliation process (Budel et al., 2004). It has been demonstrated that bioalkalization, a by-product of photosynthesis, increases the pH to above 9 by releasing OH^- to surrounding rocky grains facilitating the dissolution of silicate from endoliths. Furthermore, this exfoliation is enhanced by cyanobacterial extracellular polymer substances (EPS) which absorb water and hold those grains for longer exposure to alkaline (Budel et al., 2004; Mazor, Kidron, Vonshak, & Abeliovich, 1996). This bioweathering process is mainly involved in formation of desert landscapes, shaping the visual nature of this biome (Pointing & Belnap, 2012).

The cyanobacterial EPS are primarily hetero-polysaccharide associated with bacterial cell surface. According to their thickness and appearance, EPS are defined as sheath, capsule and slime. The sheath is a thin, electron-dense layer surrounding cell or cell groups; the capsule is generally a thicker and slimy layer closely associated with the cell surface whilst the slime refers to the mucilage material around the cells but not reflecting the cell shape. The EPS also can be secreted into environments, referred to as released polysaccharides (RPS). These water-soluble RPS are either derived from the external cell surface or from intracellular biosynthesis, and they can be easily recovered from environments or liquid medium for industrial application (De Philippis & Vincenzini, 1998). Cyanobacteria EPS are unique and more complex than other bacterial EPS, usually comprising different types of monosaccharides, acetyl groups, polypeptides, uronic acids and sulphate groups. Both uronic acids and sulphate groups contribute to the negative charge of this polymer, resulting in a strong affinity towards cations and metal ions. Moreover, the compositions of cyanobacterial EPS are distinct among different cellular types of cyanobacteria. Unicellular cyanobacteria primarily produce large amounts of uronic acids and deoxy-sugars, non-heterocystous filamentous cyanobacteria have no pyruvate and acetate in their EPS, and EPS produced by

heterocystous filamentous cyanobacteria contain only traces of sulphate (Pereira et al., 2009).

Apart from bio-weathering, EPS also play important roles in soil stability, hydrology and fertility. Dryland soil particles are poorly aggregated and highly erodible because of low moisture content. Sparse vegetation cover also means that soils are constantly exposed to wind and water erosion (Laity, 2008; Thomas, 2011). In soil crusts, the filaments of upper-layer-dwelling *Microcoleus* facilitate the surface soil aggregation (Belnap, 2006). During wet periods, *Microcoleus* forms the microbial mats on the surface to stabilize the soil (Belnap, 2006) while in drought periods these filaments penetrate deeper layers of the crust to entangle soil grains (Mager & Thomas, 2011). EPS produced either by filamentous or non-filamentous cyanobacteria act as network tissues for adhesion of soil particles (Belnap, 2006; Mager & Thomas, 2011). In hyper-arid deserts where the soils crusts are absent, cyanobacteria-dominated hypoliths may contribute to soil stability around these colonized rocks with similar roles (Pointing & Belnap, 2012). Soil moisture is also influenced by cyanobacterial EPS. Cyanobacteria can absorb up to 10 to 23 times their volume of water in EPS matrix (Belnap, 2006; Satoh et al., 2002). Therefore, EPS not only increase the water-holding capacity of soils (Mager & Thomas, 2011) but also decrease evaporation rates by blocking the soil pores (Pointing & Belnap, 2012). However, porosity of soils decreases as the cyanobacterial mats increase at the surface layers of soils, resulting in reduction of rainfall infiltration to lower layers of soils (Belnap, 2006). EPS also contributes to soil fertility due to its negative charge. EPS acts as a reservoir that not only facilitates the excess of growth-essential chelate elements from the cell surfaces but also concentrates nutrients by trapping nutrient-rich soil grains and by glycosidic bonds, thus, prevents the nutrient loss from soil systems (Mager & Thomas, 2011; Pereira et al., 2009; Pointing & Belnap, 2012).

In desert environments, freeze-thaw cycle, moisture deficiency, and UV stress are the main challenges for microorganisms; however, cyanobacteria appear to tolerate these stressors due to external EPS layers. EPS are highly hygroscopic (Nicolaus et al., 1999) and are able to retain moisture outside the cells, stabilize the cell membrane and slow down the drying process during desiccation (Pereira et al., 2009; Pointing & Belnap, 2012; Potts, 1994), and speed up the rehydration when moisture becomes available (Satoh et al., 2002). In *Nostoc commune*, the EPS matrix apparently absorbed water 20 times its initial dry weight in a few minutes during the rewetting process, facilitating the restoration of cellular pigment levels and allowing a rapid returning to photosynthesis activity (Satoh et al., 2002; Tamaru, Takani, Yoshida, & Sakamoto, 2005). Moreover, EPS also appeared to be responsible for freeze-thaw tolerance since EPS-depleted cells displayed lower photosynthetic activity after such treatment (Tamaru et al., 2005). On the other hand, secretion of EPS is enhanced under different stresses. For example, UV radiation may lead to direct DNA damage to microbial cells such as pyrimidine dimerization (Cox & Battista, 2005; Gao & Garcia-Pichel, 2011). Fortunately, cyanobacteria are able to synthesize UV-absorbing secondary metabolites that act as sunscreen compounds (scytonemin and mycosporines) which are secreted extracellularly and deposited in EPS layer to relieve UV stress (Gao & Garcia-Pichel, 2011). UV-B not only induced the synthesis of these sunscreen compounds but also enhanced the secretion of EPS in *Nostoc* (Ehling-Schulz, Bilger, & Scherer, 1997; Feng, Zhang, Feng, & Qiu, 2012).

The understanding of EPS synthesis pathways in cyanobacteria is very limited since most studies have been conducted with other Gram-positive and Gram-negative bacteria as models (Laws, Gu, & Marshall, 2001; Welman & Maddox, 2003; Whitfield, 2006). The EPS synthesis pathway is very complex, including genes that are directly involved in EPS synthesis as well as genes engaged in the formation of cell wall polysaccharides

and lipopolysaccharides (Pereira et al., 2009). In spite of this complexity, the synthesis pathway appears to be relatively conserved in bacteria (Pereira et al., 2009). The first step is the conversion of monosaccharides into sugar nucleotides by different sugar-synthesis enzymes in cytoplasm; however, these enzymes are not specific to the EPS synthesis pathway (Reeves et al., 1996). The second step is the transfer of sugar nucleotides from activated donor molecules to a lipid-linked carrier by specific glycosyltransferases, where the repeating unit of sugar nucleotides are assembled via glycosidic linkages in the plasma membrane (Roberts, 1996; Whitfield, 2006). The final step is the polymerization and export of EPS by the Wzy-dependant pathway, taking place between the plasma membrane and outer membrane in cyanobacteria (Fig 5.1) (Islam & Lam, 2014; Pereira et al., 2009). The newly synthesized lipid-linked saccharides are then translocated from the plasma membrane to periplasmatic space by integral membrane protein Wzx (flippase). The polymerization of these saccharides is then carried out by Wzy (polymerase) in the periplasmatic space. The polymerization step also requires the auxiliary protein Wzc for control of polymer chain length. The phosphorylation state of Wzc is regulated by kinase Wzb. The process of externalizing EPS is mediated by Wzc and another auxiliary protein Wza, which form a channel from the periplasmatic space to the outer membrane, exporting the polymeric substance to the cell surface (Islam & Lam, 2014; Pereira et al., 2009).

Analyses of cyanobacterial genomes (unicellular *Cyanothece*, and filamentous *Lyngbya* and *Nostoc*) have revealed the putative genes involved in EPS synthesis by the Wzy-dependant pathway (Pereira et al., 2009). Furthermore, some of the putative EPS synthesis genes in freshwater cyanobacteria *Synechocystis* sp. PCC 6803 were functionally characterized by constructing EPS-depleted mutants (Fisher, Allen, Luo, & Curtiss, 2013; Jittawuttipoka et al., 2013). It has been indicated that Wza channel proteins are essential for EPS production, and for cell survival under osmotic stress,

heavy metal stress and iron starvation (Jittawuttipoka et al., 2013). However, none of these studies have investigated EPS genes in any desert cyanobacteria.

Lowland Dry Valleys in Antarctica support mainly *Leptolyngbya*-like cyanobacteria (Yung et al., 2014), whilst high inland sites were dominated by *Chroococcidiopsis sp.* (Pointing et al., 2009). The images of low-temperature scanning electron microscopy showed a copious EPS matrix, which is secreted by hypolithic and endolithic cyanobacteria, entrapped other heterotrophic bacteria (De los Ríos et al., 2014; De Los Ríos et al., 2014). A "microbial cabana" has been proposed that these lithic cyanobacteria form an upper protective layer of biomass to filter the environmental stressors for entire communities since cyanobacteria are stress-tolerant (Pointing & Belnap, 2012). Indeed, an *in vitro* study has indicated that *Chroococcidiopsis* strains isolated from hyper-arid deserts (Atacama desert and Dry Valleys) displayed cell membrane integrity, cellular vitality, and genomic DNA integrity after being treated with desert-like environmental stressors: desiccation, freeze-thaw cycle and UV irradiation (Billi et al., 2011). These desert cyanobacteria appeared to adapt to extreme environments by developing unique EPS synthesis pathways; in contrast to marine-origin cyanobacteria (*Synechococcus* and *Prochlorococcus*), where species have lost of EPS synthesis genes during adaptation as their habitats rarely bear any environmental stress (Pereira, Mota, Vieira, Vieira, & Tamagnini, 2015). To test whether desert cyanobacteria have developed such adaptation to their environment, several cyanobacteria were isolated from deserts in China, Tibet and Antarctic Dry Valleys. *De novo* metagenome sequencing of each enriched culture was performed on Miseq, Illumina to identify each gene involved in EPS Wzy-dependent pathway by assembling each cyanobacterial genome.

Next generation sequencing (NGS) is known as massively sequencing technology capable of sequencing millions of small fragments of DNA in parallel (Shendure & Ji,

2008). One of the applications is whole genome sequencing, that is, bacterial genome can be obtained by assembling individual short sequences generated by NGS platform (Hernandez, François, Farinelli, Osterås, & Schrenzel, 2008). This process is essential to identify several functional gene-encoding regions in a given genome (MacLean, Jones, & Studholme, 2009). In this study, de novo sequencing was applied to identify EPS synthesis genes. These putative genes were then determined by searching for Pfam functional domains, and by BLAST searches using annotated EPS gene sequences of *Cyanothece* sp. CCY110 as the query. To establish the phylogenetic relationships of genes involved in Wzy-dependent pathway in cyanobacteria between different habitats and morphology, desert EPS genes identified in this study, together with other cyanobacterial EPS genes, were analyzed.

5.2 Material and methods

5.2.1 Cyanobacteria isolation and establishment of enriched cultures

Several cyanobacterial strains have been isolated from Dry Valley polar deserts (Table 5.1). Antarctic strains B5E3 and CBM were isolated from cryptoendoliths and chasmoendoliths, respectively, from McKelvey Valley. THF was isolated from cryptoendoliths from Taylor Valley. Biomass was scratched from the surface of endoliths and then placed into liquid BG-11 medium (Sigma-Aldrich, St. Louis, Missouri, United States) for enrichment. Enrichment cultures were streaked onto BG-11 agar medium for single colony formation after 10 and 100 times dilution in liquid BG-11 medium. Single colonies were picked for pure culture enrichment.

Other hypolithic cyanobacteria strains were isolated from China and Tibet deserts by Wong (Wong et al., 2009, 2010): SORK was from Sorkuli, Qaidam Basin, Qinghai, China; RQEc was from Taklimankan Desert, Xinjiang, China; LS2 and LS5 were from Gertse County, Tibet; and Q11A and Q11B were from hypoliths in Gyirong County,

Tibet. Cultures were maintained in BG-11 liquid medium at 10ºC for Antarctic strains, and 20ºC for China and Tibet strains under the 12hrs :12hrs light:dark cycle with low light level of 10 μmol/photons m^2 s^1. Morphology of each culture was observed under bright-field microscopy with total magnification of 1000X.

5.2.2 Purification of cyanobacteria cultures

Two strategies were attempted to obtain axenic cynoabcterial cultures: antibiotic selection and flow cytometery. Kanamycin (100 μg/ml) and cefotaxime (100 μg/ml) were used in an attempt to purify cyanobacteria cultures. By inhibiting heterotrophic and contamination bacteria, cultures were incubated in the dark for 72 hours (Ferris & Hirsch, 1991; Katoh, Furukawa, Tomita-Yokotani, & Nishi, 2012; Vázquez-Martínez, Rodriguez, Hernández-Hernández, & Ibarra, 2004).

Flow cytometery was also attempted for purification (operated by Mr. Stephen Edgar, Technical Officer in Medical Sciences, University of Auckland). A Becton Dickinson AriaII SORP cell sorter (BD Biosciences, San Jose, CA, USA) was used, fitted with a 70μm nozzle and operating at a sheath pressure of 70 psi. Because of the small size of the cyanobacteria and other bacterial cells, forward scatter (FSC) and side scatter (SSC) were displayed on a biexponential scale, and a SSC threshold of 200 was applied. With these settings, the cyanobacteria could be visually separated from bacteria and debris. A doublet discrimination gate of FSC-Area vs. SSC-Area was applied to the cyanobacteria population. The green fluorescence of phycobilin was excited by a 488nm laser and observed with a 505LP-530/30 filter set. Red fluorescence of chlorophyll *a* was excited with a 640nm laser and observed with a 670/30 filter (Cellamare, Rolland, & Jacquet, 2009). Cells that fell within the FSC-SSC cyanobacteria gate and also exhibited high fluorescence of both phycobilin and chlorophyll *a* were collected. Up to 5 million cells were collected from each culture. The efficiency of these processes was examined using 16S rRNA sequencing.

5.2.3 DNA extraction from cyanobacteria

The mixture of 200µl of cyanoabcterial cell and an equal volume of the phenol/chloroform/isoamyl alcohol solution (Sigma-Aldrich) were added to 1.7ml eppendorf, labeled as tube1. After inverting tube1 20 times, centrifugation was applied at 12,000g for 5 mins. The top aqueous solution was transferred into eppendorf, labeled as tube2 without picking up any of phenol/chloroform/isoamyl alcohol phase. Then additionl 200µl of elution buffer (10mM Tris-HCL, pH 8.5) (QIAGEN, Hilden, Germany) were added into tube1. After inverting tube1 20 times and centrifuging at 12,000g for 5 mins, transfer as much as possible of the top aqueous solution without picking up any of phenol/chloroform/isoamyl alcohol phase into tube2. Now the tube2 contained the DNA from two times extraction. For obtaining cleaner DNA, the re-extraction was necessary with the same procedure as above (Sambrook et al., 1989).

The re-extracted DNA were then precipitated by the following procedures: (1) add 7.5M of NH4OAc solution (Sigma-Aldrich) to tube2, making a final concentration to 0.75 M; (2) add 1µl of glycogen (20 mg/ml) (Sigma-Aldrich) and mix the solution by inverting tube2 20 times; (3) add 2.5 times the volume of 100% ethanol and mix well by inverting, then centrifuge at 14,000g for 20 mins at 4ºC; (4) decant the supernatant carefully without disturbing the pellet; (5) wash the pellet by adding 300µl of 80% EtOH and vortex shortly for 3 times; (6) centrifuge at 14,000g for 20 mins at 4ºC and then decant supernatant carefully without disturbing the pellet; (7) repeat the washing steps 5 and 6. The remaining ethanol was removed by p20 pipetteing and air dried until all ethanol was vapored. The DNA pellet was re-suspended by ultra pure distilled water (Invitrogen, Waltham, Massachusetts, USA) to 20µl (Sambrook et al., 1989).

5.2.4 Assessment of cyanobacteria culture composition and purity

5.2.4.1 Illumina 16S library preparation and sequencing

Each DNA sample was diluted to 5 ng/µl in ultra pure distilled water (Invitrogen) before library preparation according to the Illumina16S Metagenomic Sequencing Library Preparation kit (Part #15044223 Rev. B.) In brief, the first amplicon PCR was conducted with the primer set: PCR1 forward primer (1 µM, 5µl) 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3' and PCR1 reverse primer (1 µM, 5µl) 5'-GTCTCGTGGGCTCGGAGATGTG TATAAGAGACAGGACTACHVGGGTATCTAATCC-3', 12.5µl of KAPA HiFi Hotstart Readymix (Kapa Biosystems, Inc., Wilmington, MA) and 2.5µl of DNA with the following thermocycling parameters: (1) 95°C for 3 minutes; (2) 25 cycles of 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds, 72°C for 5 minutes and (3) holding the samples at 4°C. The quantity of PCR product was determined using a Qubit double-stranded DNA High Sensitivity reagent kit (Life Technologies, Foster City, CA, USA) with a Qubit fluorometer (Thermo, Fisher Scientific, Waltham, MA). The quality and the size of amplicon was assessed using a High Sensitivity DNA kit on a 2100 Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA, USA) with a DNA 1000 chip. Once the size was confirmed (~550 bp), the amiplicon was used as a template for the index PCR using Nextera XT v2 set C index kit (Illumina) according to the manufacturer's instructions. PCR cycling was carried out as follows: (1) 95°C for 3 minutes; (2) 8 cycles of 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds; and (4) 72°C for 5 minutes and (5) holding the samples at 4°C. The indexed amplicons were purified and size selected using AMPure XP beads (Beckman Coulter), followed by library size validation (~630 bp) as described above. All 16S libraries were diluted to 4nM and pooled together. After the denaturing step, the pooled library was diluted to 8pM, followed by sequencing on an Illumina Miseq using the 500 cycle V2 chemistry (250 bp paired-end reads). A 5% PhiX spike-in was used, as per the manufacturer's recommendation.

5.2.4.2 16S rRNA sequence analysis

The composition of the enriched cultures and the purification efficiency were examined by analysis of 16S rRNA gene sequences generated in section 5.2.4.1. USEARCH v8.0.1623 (Edgar, 2010) was used to quality control and process the raw sequencing reads, with the following workflow to remove anomalous sequences. First, the forward and reverse paired-end reads were merged, and the merged reads with lengths outside a 200-500 bp range or exceeding 6 homopolymers were removed using MOTHUR v1.36.1 (Schloss et al., 2009). Next, the sequences were subjected to Q score filtering to remove reads with a maximum expected error >1. Singleton reads were then removed. These sequences were assembled *de novo* (with a minimum identity of 97 %) into 182 operational taxonomic units (OTUs) among the eight enriched and three purified culture samples. The representative sequences of the OTUs were taxonomically assigned using the RDP classifier (Q. Wang, Garrity, Tiedje, & Cole, 2007) implemented in QIIME v1.9.1 (Caporaso et al., 2010). Greengenes release 13_8 (McDonald et al., 2012) was used as the reference taxonomic database. The original merged reads were mapped back to the representative sequences to establish abundance of each assigned taxonomy.

5.2.5 Cyanobacterial metagenome analysis

5.2.5.1 Metagenome library preparation and sequencing

Each DNA sample was diluted to 0.2 ng/µl in ultra pure distilled water (Invitrogen) before library preparation according to Illumina Nextera XT DNA Sample Preparation kit (Part #15044223 Rev. C). Tagmented (tagged and fragmented) DNA was prepared by incubating input DNA (5µl) at 55°C for 30 seconds with the Illumina Tagment DNA Buffer (10µl), and Amplicon Tagment Mix containing the Nextera XT transposome. The transposome simultaneously fragments the input DNA and adds adapter sequences to the fragment ends, facilitating subsequent PCR amplification. Once cooled to 10°C,

this reaction was then neutralized by incubation with the Illumina Neutralize Tagment Buffer (5µl) at room temperature for 5 minutes. The tagmented DNA was then amplified and indexed using Nextera PCR Master Mix (15µl) and Nextera XT v2 set A index primers (5µl for each) with the following thermocycling parameters: (1) 72°C for 3 minutes; (2) 95°C for 30 seconds; (3) 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds, 72°C for 30 seconds; and (4) 72°C for 5 minutes and (5) holding the samples at 10 °C. The indexed amplicons were purified and size selected using AMPure XP beads (Beckman Coulter) as specified by the Illumina protocol. Library size and quantity were determined as described in section 5.2.4.1. Libraries were normalized and pooled before sequencing on an Illumina Miseq with the 600 cycle V3 chemistry (300 bp paired-end reads). A 5% PhiX spike-in was used as per the manufacturer's recommendation.

5.2.5.2 Cyanobacterial genome assembly and annotation

Low quality sequencing reads (any nucleotide of each read with quality score of lower than 30) were removed from the raw Illumina data using FastX. Paired-end reads were assembled using two assemblers: MEGAHIT 1.0.2 (Li, Liu, Luo, Sadakane, & Lam, 2015) with "meta-sensitive" preset, and SPAdes 3.6.0 (Bankevich et al., 2012) with the setting of long Illumina pair-end read length 2x250 bp. The assembled reads were then subjected to QUAST (Gurevich, Saveliev, Vyahhi, & Tesler, 2013) for quality assessment. The assembled contigs of cultures LS2, RQEc and THF were annotated using PROKKA (Seemann, 2014) after removing all the unknown nucleotides (N) in assembled reads.

Note: The analyses in 5.2.4.2 and 5.2.4.3 were performed with assistance of Dr. Kevin C. Lee, School of Applied Science, Auckland University of Technology.

5.2.5.3 Identification of putative homologues of WZY-dependent pathway

Two strategies were used for identifying protein homologues of the Wzy-dependent EPS synthesis pathway in assembled contigs. First, *Cyanothec*e sp. CCY110 *wza*, *wzb*, *wzc*, *wzx* and *wzy* nucleotide sequences were used as queries (Pereira et al., 2009) for BLASTX searches against the predicted amino acid dataset from the assembled contigs (see section 5.2.5.2). Second, putative homologues were screened for specific Pfam functional domains (Finn et al., 2015) found in the Wzy-dependent pathway protein from *Cyanothece* sp. CCY0110, *Lyngbya aestuarii* PCC 8106 (Pereira et al., 2009) and *Synechocystis* sp. PCC 6803 (Pereira et al., 2015). Wza contains a single Poly_export domain (Pfam session no. PF02653), which is a member of the periplasmic protein families involved in polysaccharide biosynthesis and/or export. Wzb contains a single LMWPc domain (PF01451), which is a family of the protein tyrosine phosphatases. Wzy protein belongs to the O-anti_assemgly superfamily containing either Pfam domain Wzy_C (PF04932) or O-antigen_lig (PF13425). For Wzx screening, Pfam domain MatE (PF01554), polysacc_synt (PF01943) and polysacc_synt_3 (PF13440), which are members of the MATE-like superfamily, were searched individually. Wzc is a large trans-envelope membrane protein containing three Pfam domains: Wzz (PF02706) for chain length determination, GNVR (PF13807) for G-rich putative tyrosine kinase and AAA_31 (PF13614) for ATPase associated protein with diverse cellular activities (Table 5.2). Non-cyanobacterial EPS contigs were excluded using protein BLAST on NCBI.

5.2.3 Phylogenetic analysis of putative EPS homologues

The amino acid sequences of putative cyanobacterial EPS homologues (Wzy-dependent pathway) identified in this study, together with other EPS homologues of cyanobacteria with different morphologies and environmental sources, were all included in a

phylogeny analysis. These latter protein sequences were obtained by using protein BLAST searches of NCBI (Altschul, Gish, Miller, Myers, & Lipman, 1990; States & Gish, 1994) using assembled EPS contigs from this study as query sequences. All protein sequences were aligned by MUSCLE (Edgar, 2004) in MEGA6 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). The best substitution model was determined by Model Test before constructing a maximum likelihood tree in MEGA6 (Tamura et al., 2013). The best model for each tree is shown as follow: LG+G+I+F for Wza phylogeny, LG+G+I for WZB phylogeny, LG+G+I+F for Wzc phylogeny, LG+G+F for Wzx phylogeny and JTT+G+I+F for Wzy phylogeny. Branch support was determined by bootstrapping 100 times. Sequences of *Cyanothece* sp. CCY110 were assigned as outgroup in each tree. The source of each cyanobacteria (Table 5.3) was obtained from the database of NCBI, Integrated Microbial Genomes (Markowitz et al., 2014) and The Pasteur Culture Collection of Cyanobacteria (PCC: http://cyanobacteria.web.pasteur.fr).

5.3 Results

5.3.1 Cyanobacterial cultivation and purification

All cyanobacterial isolates were cultured in BG-11 liquid medium without agitation and maintained at 10 or 20°C according to their habitats (Table 5.1). Most cyanobacterial cultures were found to be low-light adapted (10 μmol/ photons m^2 s^1) although the culture B5E3 could tolerate light level up to 20 μmol/ photons m^2 s^1. Most of the isolates were unicellular cocci except for B5E3, which was unicellular rod-shaped, and THF, which was filamentous (Fig 5.1).

During cultivation, turbidity was observed occasionally in most cultures, indicating they were mixed with certain amount of other bacteria. The composition of enriched cultures was then examined by 16s rRNA sequencing. The sequencing results showed

that LS2, LS5, Q11A, RQEc and SORKI were *Chroococcidiopsis*-like phylotype and THF was *Leptolyngbya*-like phylotype; culture B5E3 and CBM were identified as uncultured cyanobacteria (Table 5.4). However, the results also revealed that these enriched cultures were not pure as expected. The majority of 16S OTUs in each culture were cyanobacterial, they represented 42 % to 89% of total 16S rRNA sequences (Table 5.4). The additional cyanobacteria OTU indicated different cyanobacteria species found in certain cultures (culture CBM: OTU 2 and OTU 123; LS2: OTU129 and OTU 1; RQEc OTU1 and OTU 129; SORK: OTU1 and OTU 129), and Actinobacteria, Proteobacteria and Deinococcus-Thermus OTUs revealed that each individual culture were not axenic. This surprising result was verified by also analysing 'axenic' cyanobacteria from German and US culture collections, which also proved to be mixed cultures similar to the above (data not shown).

16S rRNA sequences were also determined as a way of assessing of culture purification efficiency after the first sequencing. Two methods for purifying cyanobacterial cultures revealed opposite results. Culture CBM was applied to flow cytometery and the purity was improved after the cell sorting. From the flow cytometery analysis, the population of cyanobacteria increased from 73.7 % to 94.9% (Fig 5.2); however, the sequencing outcome showed that the numbers of cyanobacteria OTU only increased from 77% to 81%. All cultures were treated with antibiotics and only CBM and LS2 were applied to 16S rRNA sequencing. Antibiotic selection did not improve the culture purity. The numbers of cyanobacterial OTUs in these two cultures decreased after antibiotic treatment (Table 5.4), indicating that these cyanobacteria were sensitive to kanamycin and cefotaxime at the level used.

5.3.2 *De novo* metagenome assembly

Metagenome analysis was carried out on the enriched cultures, 4.3 millions pass-filter reads were generated. Genome assemblies using MEGAHIT and SPAdes pipelines gave similar results. The overall quality of assembly of all cyanobacteria cultures was poor; however, Antarctic THF, Tibet LS2 and Chinese RQEc gave the best assemblies according to the longer N_{50} length. The final assembly of LS2 and RQEc via SPAdes contained, respectively, 9345 and 8001 assembled contigs with the longest contig length of 85,805 and 88137 bp, N_{50} contig length of 7194 and 6064 bp, and total G+C contect of 59% and 57% (Table 5.5). The final assembly of THF by MEGAHIT contained 4528 assembled contigs with the longest contig length of 61,674 bp, N_{50} contig length of 6,180 bp and total G+C content of 54% (Table 5.6). The annotation process via PROKKA, and the identification of EPS homologues involved in the Wzy-dependent pathway were performed using the metagenomes of THF, LS2 and RQEc only since these three assemblies gave the better results.

5.3.3 Phylogeny analysis of putative WZY-dependent pathway homologues

The putative homologues of the proteins involved in the Wzy-dependent pathway were identified by detection of relevant Pfam functional domains. The assembled contigs with lower expectation value obtained using BLASTX searches were also identified as putative homologues of proteins in the Wzy-dependent pathway. Overall, homologues for each protein involved in the Wzy-dependent pathway were identified in THF, LS2 and RQEc. Three contigs (THF_05431, LS2_01563, RQEc_05876) were identified as putative Wzb homologues, while Wza homologues were contigs THF_07723, THF_02157, LS2_04994 and RQEc_02740. Contigs THF_08804, THF_06609, LS2_01044 and RQEc_04622 with domain Polysacc_synt_3 were determined as Wzx homologues, same as contigs LS2_00776, LS2_03996, LS2_07123,

RQEc_02592, RQEc_04585 and ROEc_06442 containing domain Polysacc_synt. However, no contigs with homology to domain MatE were recovered. Contigs THF_00290, LS2_08392 and RQEc_03312, and contigs LS2_00010 and RQEc_01924, contained the Wzy_C and O-anti_lig domains, respectively, suggesting they are Wzy homologues. Several contigs (THF_07722, LS2_00769, RQEc_00387 and RQEc_07349) containing multiple domains Wzz, GNVR and AAA_31 were identified as Wzc homologues. Contigs THF_08171, THF_06826, LS2_01047, LS2_05641, RQEc_03216 and RQEc_04624 containing only Wzz and AAA_31 domains were also identified as Wzc homologue. The sequences of the above contigs are shown in appendices.

Maximum likelihood phylogenetic trees were constructed for each protein involved in the Wzy-dependent pathway. General observations of these trees are that the sequences tend to group by cell morphology and habitat types (aquatic or terrestrial). For example, sequences within a clade would be largely from either aquatic or terrestrial species that are unicellular, non-heterocystous filamentous or heterocystous filamentous. Also, in general, lineages leading to sequences from unicellular and non-heterocystous filamentous species are older than those leading to sequences from heterocystous filamentous species. Furthermore, no clear correlation was observed between sequence relationships and habitats the species were derived from.

The phylogenetic tree of Wzb homologues displayed a relatively clear pattern according to the cyanobacteria morphology (Fig 5.4). Sequences isolated from aquatic unicellular cyanobacteria, together with clade F and G, were the most ancient lineage among all the Wzb homologues. Clade E represented non-heterocystous cyanobacteria from aquatic environments whereas clades C and D shared a common ancestor with clade E were from aquatic unicellular and filamentous cyanobacteria originated from different environments, respectively. Homologues from heterocystous filamentous

cyanobacteria (clade A) were all grouped together and divided into different clades largely based on habitat, although some sequences from terrestrial species *Nostoc*, *Calothrix*, *Scytonnema*, *Hassallia* and *Cylindrpspermum* were grouped within the aquatic clade. The putative Wzb proteins from this study were all grouped within clade B associated with others from terrestrial unicellular cyanobacteria except sequence from *Chroococcales*. This suggests they are orthologous genes.

Similar to Wzb phylogeny, the oldest Wza homologues were aquatic unicellular (Fig 5.5). However, the desert-derived Wza homologues were grouped into different clades. Contig THF_02157 isolated from non-heterocystous filamentous cyanobacteria in Dry Valleys was positioned with other sequences from non-heterocystous filamentous in clade E, especially with the sequences from the terrestrial *Nodosilinea nodulosa*. Contigs RQEc_02740 and LS2_04994 from *Chroococcidiopsis*-like isolates were orthologues, belonged to the unicellular lineage B, a clade of sequence from species such as *Chroococcidiopsis thermalis* PCC 7203. Also Wza homolog from THF is the oldest than other desert-derived Wza.

In the Wzx phylogenetic tree (Fig 5.6), clade H, together with other unicellular aquatic homologues, was the oldest lineage. Two contigs derived from THF were paralogoues and positioned within clade G with other non-heterocystous filamentous sequences; however, one homolog from LS2 was also positioned in clade G. Contig RQEc_04622 formed a clade F with other heterocystous and non-heterocystous sequences. Another set of homologues with domain polysacc_synt formed a separate lineage (clade A to D). These sequences appeared to be more variable than others since the branches are longer. Contig RQEc_04585 was positioned in clade A with homologues from various types of cyanobacteria, especially with unicellular ones. LS2_07123 and RQEc_02592 (orthologues) were grouped into non-heterocystous filamentous clade B. LS2_03996 and RQEc_0644 (orthologues) were positioned in

terrestrial clade C, especially with homolog from *C. Thermalis* PCC 7203. LS2_00776 was placed into heterocystous filamentous clade D.

Similar to the Wzx phylogeny, homologues with different Pfam domains were divided into two lineages distantly in the Wzy phylogenetic tree (Fig 5.7). Contigs LS2_08392 and RQEc_03112, and THF_00290 containing domain Wzy_c were positioned into the unicellular terrestrial clade D, and non-heterocystous filamentous clade B, respectively, with other homologues derived primarily from aquatic habitats. Contigs RQEc_01924 and LS2_00010 with domain O-antigen_lig were grouped into clade A with most homologues isolated from unicellular cyanobacteria. All the clades formed multiples lineages from common ancestors. Sequences in clades A and B seemed to be more variable than other clades. Orthologues (clade A and E) were identified from cultures LS2 and RQEc, and only one Wzy homolog was identified from culture THF.

The Wzc phylogeny included up to 285 sequences of Wzc homologues (Fig 5.8). RQEc_00387 and LS2_00769 were grouped into heterocystous filamentous aquatic clade A. Contigs RQEc_03216 and LS2_05641 (orthologues) formed a unicellular clade A with other Wzc homologues derived from terrestrial environments, whereas RQEc_04624 and LS2_01047 (orthologues) were places into clade E with Wzc homologues from aquatic habitats. THF_07722 was positioned into heterocystous filamentous clade B, whereas THF_08171 was placed into clade D with homologues from different types of cyanobacteria. However, THF_06826 and RQEc_07349 did not belong to any specific clades. It is also clear to see that all the clades formed multiples lineages from common ancestors in this phylogeny.

5.4 Discussion

This study attempted to assemble the whole genome of desert-derived cyanobacteria searching for stress response genes since cyanobacteria are the keystone taxa in desert

environments (Chan et al., 2012; Cowan et al., 2014; De Los Ríos et al., 2014; Makhalanyane et al., 2015; Pointing & Belnap, 2012; Pointing et al., 2015; Wierzchos et al., 2012) and display high resistance to environmental stress (Billi et al., 2011, 2000). However, the assemblies resulted in a large numbers of contigs with short N50 values; that is, many sequences were too short to be used for assembling the whole genome. Several studies of cyanobacteria genomes have presented that contig numbers after assembly as 4 to 61 with N_{50} ranges from 76,101 to 546,047 bp for small sized genomes (1.78M bp) of *Prochlorococcus*. For the large genome of *Leptolyngbya* sp. (8.08 Mbp) the number of contigs was only 119 with the N_{50} length of 103,122 bp (Paul et al., 2014). The failure here of assembling the whole genome was likely due to the complex community composition in cultures.

Culture analysis by flow cytometery has indicated the purity of CBM culture was improved; however, result from the 16S sequence analysis revealed that up to 20% of Proteobacteria and other bacteria were still present after sorting. Although BG-11 medium for cyanobacteria cultivation was mineral-based for selection of heterotrophic bacteria, Proteobacteria are facultative autotrophs, harboring different forms of Rubisco for carbon fixation (Badger & Bek, 2008). The presence of heterotrophic Actinobacteria and Deinococcus-Thermus can be explained by the carbohydrates produced via photosynthesis carried out by the cyanobacteria (Calvin & Benson, 1948; Stanier & Cohen-Bazire, 1977). Apart from the limitation of the cell sorter, another explanation may be that these other bacteria are intimately attached to cyanobacterial EPS layers (De los Ríos et al., 2014; de Los Ríos, Ascaso, Wierzchos, Fernández-Valiente, & Quesada, 2004; De Los Ríos et al., 2014). The cell sorter was capable of separating particles bearing cholorophyll *a* and phycobilin (Cellamare et al., 2009), whereas other bacteria associating with cyanobacteral EPS layers were indivisible and were all collected together with cyanobacterial cells. Some studies have reported using antibiotic

selection for obtaining axenic cyanobacterial cultures (Ferris & Hirsch, 1991; Katoh et al., 2012; Vázquez-Martínez et al., 2004). The significant decrease of cyanobacteria OTUs after antibiotics treatment has indicated that the desert-origin cyanobacteria were more sensitive to antibiotics than cyanobacteria from other ecosystems. This is supported by a previous metagenome study showing that antibiotic resistance genes were less abundant in desert microbial communities (Fierer et al., 2012).

Several studies have focused on identification of gene encoding proteins involved in the Wzy-dependent pathway (Pereira et al., 2009), leading to a misconception that the primary EPS synthesis pathway for cyanobacteria is the Wzy-dependent pathway. However, a recent study interrogating phylum-wide putative EPS synthesis genes within cyanobacteria hypothesized that two other pathways, such as ABC transporter and synthase-dependent pathways might also be prevalent in cyanobacteria (Pereira et al., 2015). The ABC transporter pathway is only responsible for export of EPS. The polysaccharides are completely polymerized at the inner leaflet of the plasma membrane (cytoplasm) before being translocated across the inner membrane by two auxiliary protein components of ABC transporter: KpsM and KpsT (Willis & Whitfield, 2013). KpsT functions as an ATPase, coupling ATP hydrolysis to translocate the polysaccharides across the membrane. KpsM functions as an integral inner membrane transporter, exporting the substrates to the periplasmatic space (Willis & Whitfield, 2013). The roles of these two proteins have been functionally characterized in cyanobacterium *Synechocystis* PCC 6803. By constructing KpsM- and KpsT-deleted strains, it has been demonstrated that the EPS compositions between mutants and wild type were the same, but the mutant strains displayed flocculent phenotypes and increased adherence to object surfaces. This might be due to the failure of exporting EPS and then altering the charge of the cell surface compared to the wild type (Fisher et al., 2013). In this study, four contigs (THF_00579,THF_08412, L2_04827 and RQEc_-

2598) were determined as putative KpsM homologues since they contained the Pfam domains ABC2_membrane (PF01061) and ABC2_membrane_3 (PF012698), indicating that the ABC transporter pathways were also present in desert-origin cyanobacteria (data not shown).

The synthase-dependent pathway is relatively complicated, involving different functional proteins for alginate and cellulose production, polymer modification, translocation and final exportation (Whitney & Howell, 2013). No putative homologues of this pathway were identified by Pfam domain search in this study, although several cyanobacteria have been proposed to harbour Alg8 and Alg44 inner membrane protein homologues of the synthase-dependent pathway (Pereira et al., 2015). Alg8 functions as alginate synthase and Alg 44 is responsible for polymerization of alginate (Whitney & Howell, 2013). The lack of Alg8 and Alg 44 homologues from this study might be due to the poor assembly quality. However, since the cultures were not purely cyanobacterial, a few contigs were identified as proteobacterial Alg44 homologues (Pfam PilZ domain, PF07328). Thus, further verification of genes involved in synthase-dependent pathway in desert cyanobacteria would be necessary.

In this study, identification of genes involved in the Wzy-dependent pathway was the main focus since this pathway is involved in synthesis/polymerization and export (Islam & Lam, 2014), and each protein homologue involved in this pathway were all identified. Phylogenetic trees were constructed to test whether these homologues had a putative role in the adaptation of cyanobacteria to desert environments. Phylogenies of Wzb and Wza showed similar patterns; that these homologues were generally correlated with morphology, although some homologues derived from unicellular species were more closely related to species from non-heterocystous filamentous cyanobacteria (Fig 5.4 & 5.5). Interestingly, the phylogenetic relationship of cyanobacteria inferred from 16S rRNA sequences presented a similar outcome. Non-differentiated (non-

heterocystous) filamentous cyanobacteria (subsection III *Oscillatoriales*) usually formed clades with unicellular species (subsection I *Chroococcales*, and subsection II *Pleurocapsales*), whereas heterocystous-cyanobacteria (subsection IV *Nostocales,* and subsection V *Stigonematales*) formed a monophyletic clade (Tomitani, Knoll, Cavanaugh, & Ohno, 2006). This indicates that *wza* and *wzb* genes are highly conserved intra-specifically among heterocystous filamentous, non-heterocystous filamentous and unicellular cyanobacteria and possibly can be used as cyanobacterial EPS markers since GeoChip (Tu et al., 2014) were unable to identify cyanobacteria EPS synthesis pathways.

Several homologues of Wzc, Wzx and Wzy were recovered from each culture in this study. The Wzx phylogeny (Fig 5.6) shows that homologues contained functional domain polysacc_synt were more variable than others with domain polysacc_synt_3, and this indicates different functional traits of Wzx flippase in certain cyanobacteria. In plant pathogens *Erwinia amylovora* and *Pantoea stewartii*, different Wzx displayed substrate specificity, recognizing sugar with identical main-chain repeating units but bearing different terminal side-branch modifications (Islam & Lam, 2014; X. Wang, Yang, & von Bodman, 2012). The Wzy phylogeny (Fig 5.7) also indicates the potential substrate specificity of cyanobacterial Wzy polymerases. In certain bacteria, Wzy polymerases were incapable of polymerizing tetra-saccharides once their side chains were modified (Islam & Lam, 2014). The actual functionality of Wzc proteins among different bacteria has yet to be established. It has been proposed that the Wzy-mediated repeat-unit polymerization is under regulation of Wzc protein. So far their substrate specificity remains unclear (Islam & Lam, 2014).

In addition, these gene duplicates were possibly considered as adaptive evolution to environments (Kondrashov, 2012). However, duplicated functional genes appear to have evolved in different ways, and might not be completely advantageous to the

organisms (Force et al., 1999). The function of original genes may retain the same, whereas newly duplicate genes may acquire a novel, possibly advantageous function (neo-functionalization). Duplicated paralogues may bear the original gene function but are less functionally efficient compared to the ancestor genes (sub-functionalization). In the case of sub-functionalization, all the genes are essential to exhibiting the full functionality. Non-functionalization means that one of these gene copies may totally lost function due to nucleotide mutation, deletion or replication error (Force et al., 1999). In this case, these disadvantageous functional genes may provide a extra buffer to prevent further DNA mutations to the original functional genes (Clark, 1994). A recent study has reported that aquatic unicellular cyanobacteria (*Synechococcus* and *Prochlorococcus*) have lost most of their EPS synthesis genes, which may be an adaptation strategy (Pereira et al., 2015). *Synechococcus* strains commonly inhabit high-nutrient marine environments, whereas *Prochlorococcus* genus are restricted to warmer oligotrophic deep oceans (Flombaum et al., 2013). In their environment, EPS no longer provides any advantages against desiccation, UV radiation, nutrient absorption or adherence to solid substrates, hence, it is very likely that many of the EPS synthesis genes have been lost (Pereira et al., 2015). This study has not addressed any functional characterization of these putative EPS proteins from desert cyanobacteria; however, several duplicate genes (especially in LS2 and RQEc) indicated the functional redundancy and essentials of EPS synthesis for desert environments. The frequency of homologues from Antarctic THF was relatively lower and this might be due to the poor genome assembly. It is very unlikely filamentous cyanobacteria have lost certain EPS synthesis genes since filamentous cyanobacteria usually harbour more copies of each EPS synthesis genes due to the large genome size (Pereira et al., 2015).

It is interesting that in each phylogenetic tree, some homologues from *Chroococcidiopsis*-like desert isolates (LS2 and RQEc), *C. Thermalis* PCC7203 and

heterocystous filametous *Scytonema millei* VB511283 were grouped in the terrestrial clades. *C. Thermalis* PCC7203 was isolated from a soil habitat in Germany with temperature at 22°C (Stanier et al., 1979) and *S. millei* VB51283 was isolated from monument stone from east India where the relative humidity was high (84-95%) and temperature ranged from 26°C to 35°C (Keshari & Adhikary, 2013). This result indicates that environmental factors seemed to be the adaptive drivers for these Wzy synthesis pathway proteins. However, phylogenetic analyses showed that most of the homologues derived from either aquatic or terrestrial cyanobacteria were grouped together based on their cell morphology. Moreover, a previous report has summarized that EPS produced by the same subsection of cyanobacteria share similar constitutes of EPS (Pereira et al., 2009). These findings indicate that EPS proteins involved in the Wzy synthesis pathway seemed to be more correlated to cell morphology or subsections than environments for adaptation. This leads to a hypothesis that these EPS synthesis genes have not had a role in the adaptation of cyanobacteria. Rather environmental adaptation may be the consequences of how these genes are regulated and induced under certain mechanisms and conditions to respond to environmental changes.

Table 5.1 The data of sample site and growth conditions of each cyanobacteria culture

	Sample site*					
	Location	**Substrate**	**GPS coordinate**	**Soil Temperature**	**Morphology**	**Growth temperature**
B5E3	McKelvey Valley, Antarctica	Endolith	S77º24.595', E161º11.747'	<0 ^{O}C	Rod	10 ^{O}C
CBM	McKelvey Valley, Antarctica	Chasmoendolith	S77º24.604', E161º11.702'	<0 ^{O}C	Cocci	10 ^{O}C
THF	Taylor Valley, Antarctica	Endolith	S77º43.011', E162º38.064'	<0^{o}C	Filamentous	10 ^{O}C
LS2	Gertse County, Tibet	Hypolith (limestone)	N32º11.546', E84º12.001'	5.8-28.4 ^{O}C	Cocci	20 ^{O}C
LS5	Gertse County, Tibet	Hypolith (limestone)	N32º11.546', E84º12.001'	5.8-28.4 ^{O}C	Cocci	20 ^{O}C
Q11A	Gyirong County, Tibet	Hypolith (quartz)	N29º07.943', E85º22.508'	5.8-28.4 ^{O}C	Cocci	20 ^{O}C
RQEc	Taklimankan desert, XingJiang Province, China	Hypolith (quartz)	N38º24.233' E88º53.806'	13±0.4^{O}C (annual)	Cocci	20 ^{O}C
SORK	Sorkuli, Qaidam Basin, QingHai Province, China	Hypolith (quartz)	N38º55.410' E92º16.703'	3.2 ±0.5^{O}C (annual)	Cocci	20 ^{O}C

*The sample site data was adapted from Priscu et al., 2008; Wynn-Williams et al., 2007; Wynn et al., 2010; except Taylor Valley

Table 5.2 Pfam domain involved in Wzy-dependent pathway selected for anal

Putative protein	Pfam designation	Pfam session
Wza	Poly_export	PF02563
Wzb	LMWPc	PF01451
Wzc	Wzz	PF02706
	GNVR	PF13807
	AAA_31	PF13614
Wzx	Polysacc_synt	PF01943
	Polysacc_synt_3	PF13440
	MatE	PF01554
Wzy	Wzy_C	PF04932
	O-antigen_lig	PF13425

Table 5.3a The sample source of unicellular cyanobacteria strains

Unicellular strains	Sample source
Acaryochloris marina MBIC11017	Marine
Acaryochloris sp. CCMEE 5410	Marine
Chamaesiphon minutus PCC 6605	Freshwater
Chroococcales cyanobacterium CENA595	Marine
Chroococcidiopsis thermalis PCC 7203	Soil
Crocosphaera watsonii WH 0003	Marine
Crocosphaera watsonii WH 0401	Marine
Crocosphaera watsonii WH 0402	Marine
Cyanobacterium aponinum PCC 10605	Freshwater
Cyanobacterium stanieri PCC 7202	Freshwater
Cyanothece sp. CCY110	Freshwater
Cyanothece sp. PCC 7424	Freshwater
Cyanothece sp. PCC 7425	Freshwater
Cyanothece sp. PCC 7822	Freshwater
Dactylococcopsis salina PCC 8305	Freshwater
Geminocystis herdmanii pcc 6380	Freshwater
Geminocystis sp. NIES-3709	Freshwater
Gloeocapsa sp. PCC 73106	Freshwater
Gloeocapsa sp. PCC 7428	Hot spring
Microcystis aeruginosa DIANCHI905	Freshwater
Microcystis aeruginosa NIES-2549	Freshwater
Microcystis aeruginosa NIES-88	Freshwater
Microcystis aeruginosa PCC 7941	Freshwater
Microcystis aeruginosa PCC 9432	Freshwater
Microcystis aeruginosa PCC 9443	Freshwater
Microcystis aeruginosa PCC 9701	Freshwater
Microcystis aeruginosa PCC 9808	Freshwater
Microcystis aeruginosa PCC 9809	Freshwater
Myxosarcina sp. GI1	Marine
Neosynechococcus sphagnicola sy1	Wetland
Pleurocapsa sp. PCC 7319	Marine
Pleurocapsa sp. PCC 7327	Hot spring
Stanieria cyanosphaera PCC 7437	Freshwater
Stanieria sp. NIES 3757	Freshwater
Synechococcus sp. PCC 6312	Freshwater
Synechococcus sp. PCC 7335	Marine
Synechococcus sp. PCC 7336	Marine
Synechococcus sp. PCC 7418	Freshwater
Synechococcus sp. strain NKBG 15041c	Marine
Synechococcus sp. UTEX 2973	Freshwater
Synechocystis sp. PCC 6803	Freshwater
Synechocystis sp. PCC 7509	Stone
Xenococcus sp. PCC 7305	Marine

Table 5.3b The sample source of non-heterocystous filamentous cyanobacteria strains

Non-heterocystous filamentous strains	Sample source
Arthrospira platensis NIES-39	Freshwater
Arthrospira platensis Paraca	Freshwater
Arthrospira sp. PCC 8005	Freshwater
Crinalium epipsammum PCC 933	Soil
Cyanobacterium ESFC-1	Marine
Geitlerinema sp. PCC 7105	Freshwater
Geitlerinema sp. PCC 7407	Freshwater
Leptolyngbya boryana dg5	Freshwater
Leptolyngbya sp. Heron Island J	Marine
Leptolyngbya sp. JSC-1	Hot spring
Leptolyngbya sp. KIOST-1	Freshwater
Leptolyngbya sp. NIES 2104	Soil
Leptolyngbya sp. NIES 3755	Soil
Leptolyngbya sp. O-77	Hot spring
Leptolyngbya sp. PCC 6406	Freshwater
Leptolyngbya sp. PCC 7375	Marine
Leptolyngbya sp. PCC 7376	Freshwater
Limnorphis robusta CS951	Freshwater
Lyngbya aestuarii BL-J	Marine
Lyngbya confervoides	Marine
Lyngbya sp. PCC 8106	Marine
Coleofasciculus chthonoplastes PCC 7420	Marine
Microcoleus sp. PCC 7113	Soil
Microcoleus vaginatus FGP-2	Desert soil
Moorea producens 3L	Marine
Nodosilinea nodulosa PCC 7104	Soil
Oscillatoria acuminata PCC 6304	Soil
Oscillatoria formosa PCC 6407	Freshwater
Oscillatoria nigro-viridis PCC 7112	Soil
Oscillatoria sp. PCC 10802	Freshwater
Oscillatoria sp. PCC 6506	Freshwater
Oscillatoriales cyanobacterium JSC 12	Freshwater
Oscillatoriales cyanobacterium MTP1	Freshwater
Phordium sp. OSCR	Freshwater
Phormidesmis priestleyi	Freshwater
Phormidesmis priestleyi Ana	Freshwater
Planktothricoides sp. SR001	Freshwater
Planktothrix agardhii NIVA-CYA 18	Freshwater
Planktothrix prolifica NIVA-CYA 406	Freshwater
Planktothrix rubescens NIVA-CYA 407	Freshwater
Prochlorothrix hollandica PCC 9006	Freshwater
Pseudanabaena sp. PCC 6802	Freshwater
Pseudanabaena sp. PCC 7429	Freshwater
Spirulina subsalsa PCC 9445	Freshwater

Table 5.3c The sample source of heterocystous filamentous cyanobacteria strains

Heterocystous filamentous strains	Sample source
Anabaena cylindrica PCC 7122	Freshwater
Anabaena sp. 90	Freshwater
Anabaena sp. PCC 7108	Marine
Anabaena sp. wa102	Freshwater
Anabaena variabilis ATCC 29413	Freshwater
Aphanizomenon flos-aquae	Marine
Aphanizomenon flos-aquae 2012/KM1/D3	Marine
Aphanizomenon ovalisporum UAM-MAO	Freshwater
Calothrix sp. 336/3	Freshwater
Calothrix sp. PCC 6303	Freshwater
Calothrix sp. PCC 7103	Terrestrial
Calothrix sp. PCC 7507	Freshwater
Chlorogloeopsos fritschii	Hot spring
Cylindrospermopsis raciborskii CS-505	Freshwater
Cylindrospermopsis sp. CR12	Freshwater
Cylindrospermum stagnale PCC 7417	Soil
Dolichpspermum circinale	Marine
Ficherella sp. PCC 9605	Freshwater
Fischerella muscicola SAG 1427	Freshwater
Fischerella sp. JSC-11	Freshwater
Fischerella sp. NIES 3754	Hot spring
Fischerella sp. PCC 9339	Freshwater
Fischerella sp. PCC 9431	Freshwater
Fischerella sp. PCC 9605	Freshwater
Fischerella thermalis	Hot spring
Hapalosiphon sp. MRB220	Hot spring
Hassallia byssoidea VB512170	Stone
Mastigocladopsis repens	Stone
Mastigocladus laminosus PCC 7702	Hot spring
Mastigocladus laminosus UU774	Hot spring
Mastigocoleus testarum BC008	Marine
Microchaete sp. PCC 7126	Freshwater
Nodularia piscinale CENA21	Freshwater
Nodularia spumigena CCY9414/PCC7310)	Marine
Nostoc azollae 0708	Freshwater
Nostoc piscinale CENA21	Soil
Nostoc punctiforme PCC 73102	Soil
Nostoc sp. NIES-3756	Soil
Nostoc sp. PCC 7107	Freshwater
Nostoc sp. PCC 7120	Freshwater
Nostoc sp. PCC 7524	Hot spring
Mastigocladus laminosus UU774	Hot spring
Mastigocoleus testarum BC008	Marine
Microchaete sp. PCC 7126	Freshwater

(continued)

Heterocystous filamentous strains	Sample source
Nodularia piscinale CENA21	Freshwater
Nodularia spumigena CCY9414/PCC7310)	Marine
Nostoc azollae 0708	Freshwater
Nostoc piscinale CENA21	Soil
Nostoc punctiforme PCC 73102	Soil
Nostoc sp. NIES-3756	Soil
Nostoc sp. PCC 7107	Freshwater
Nostoc sp. PCC 7120	Freshwater
Nostoc sp. PCC 7524	Hot spring
Raphidiopsis brookii D9	Freshwater
Richelia intracellularis HH01	Marine
Richelia intracellularis MH01	Marine
Rivularia sp. PCC 7116	Marine
Scytonema hofmanni PCC 7110	Stone
Scytonema hofmanni PCC 9009	Stone
Scytonema millei VB511283	Stone
Scytonema tolypothrichoides VB 61278	Stone
Tolypothrix bouteillei VB521301	Stone
Tolypothrix campylonmoides VB511288	Stone
Tolypothrix sp. PCC 7601	Freshwater
Trichodesmium erythraeum IMS101	Marine

Table 5.4 The enriched culture composition and purification efficiency

	Cyanobacterial OTUs	Reference taxonomy of each cyanobacterial OTU (%)*	OTUs after sorting (%)	OTUs after antibiotics (%)
B5E3	OTU2 (89.2%)	OTU 2: uncultured cyanobacteria	ND	ND
CBM	OTU 2 (61.2%)	OTU 2: uncultured cyanobacteria	80.8	35.7
	OTU 123 (15%)	OTU 123: *Chroococcidiopsis* sp. CC1 (93)		
THF	OTU 3 (53.1%)	OTU 3: *Leptolyngbya* sp. (92)	ND	ND
LS2	OTU 129 (43.5%)	OTU 129: *Chroococcidiopsis* sp. CC3 (98)	ND	45.5
	OTU 1 (10%)	OTU 1: *Chroococcidiopsis* sp. CC3 (97)		
LS5	OTU 1 (64.0%)	OTU 1: *Chroococcidiopsis* sp. CC3 (97)	ND	ND
Q11A	OTU 1 (54.2%)	OTU 1: *Chroococcidiopsis* sp. CC3 (97)	ND	ND
RQEc	OTU 1 (52.9%)	OTU 1: *Chroococcidiopsis* sp. CC3 (97)	ND	ND
	OTU 129 (6%)	OTU 129: *Chroococcidiopsis* sp. CC3 (98)		
SROKI	OTU 1 (37.2%)	OTU 1: *Chroococcidiopsis* sp. CC3 (97)	ND	ND
	OTU 129 (4.5%)	OTU 129: *Chroococcidiopsis* sp. CC3 (98)		

* The 16S similarity of reference taxonomy strains BLAST against the Greengene database
ND: not determined

Table 5.5 The assembly statistics of cyanobacteria cultures using SPAdes

	B5E3	CBM	THF	LS2*	LS5	Q11A	Q11B	RQEc*	SORK
# of contig (>= 0 bp)	29884	21334	19308	18775	23431	19499	12461	18154	35812
# of contig (>= 1000 bp)	1354	2837	1462	3697	2667	3516	2564	2159	6531
Total length (>= 0 bp)	16078553	16257733	14782198	24227109	19748737	19318406	14336562	18307288	33954657
Total length (>= 1000 bp)	1734678	6859896	6331697	16602110	9406598	14336562	9288567	10310725	18513231
# of contigs	12642	9990	6325	9345	10160	8969	6432	8001	19455
Largest contig	6199	54899	77828	85805	80902	56887	77328	88137	94998
Total length	9162823	11745573	9572501	20450887	14116771	15087364	11931163	14235470	27399418
GC %	57.29	58.24	55.23	59.48	58.37	59.06	56.48	56.75	61.34
N50	708	1209	2724	7194	2015	2747	3563	6064	1647
L50	4821	1936	342	628	1011	1055	511	511	2674

*LS2 and RQEc contigs assembled by SPAdes were subjected to PROKKA for annotation

Table 5.6 The assembly statistics of cyanobacteria cultures using MEGAHIT

	B5E3	CBM	THF*	LS2	LS5	Q11A	Q11B	RQEc	SORK
# of contig (>= 0 bp)	26767	13864	8939	13834	14718	12659	9508	11844	28255
# of contig (>= 1000 bp)	891	2367	1202	3815	2302	3376	2357	2085	5375
Total length (>= 0 bp)	1360656	11848903	9626528	20820040	14516236	15340578	12202394	14476385	27927115
Total length (>= 1000 bp)	11121806	5796358	5702048	11508082	8258131	10503942	8417087	9534024	15681280
# of contigs	10377	7716	4528	8509	7458	7569	5718	6215	16565
Largest contig	5243	42000	61674	58550	50536	51568	50549	57818	55911
Total length	7315112	9587791	7941696	18792120	11727458	13395487	10757784	12331969	23400707
GC %	57.02	57.87	53.97	59.2	56.66	58.14	55.5	55.96	60.63
N50	575	1302	6108	5212	2712	2787	3663	5676	1725
L50	6971	1438	258	821	740	2946	521	540	2220

* THF contigs assembled by MEGAHIT were subjected to PROKKA for annotation

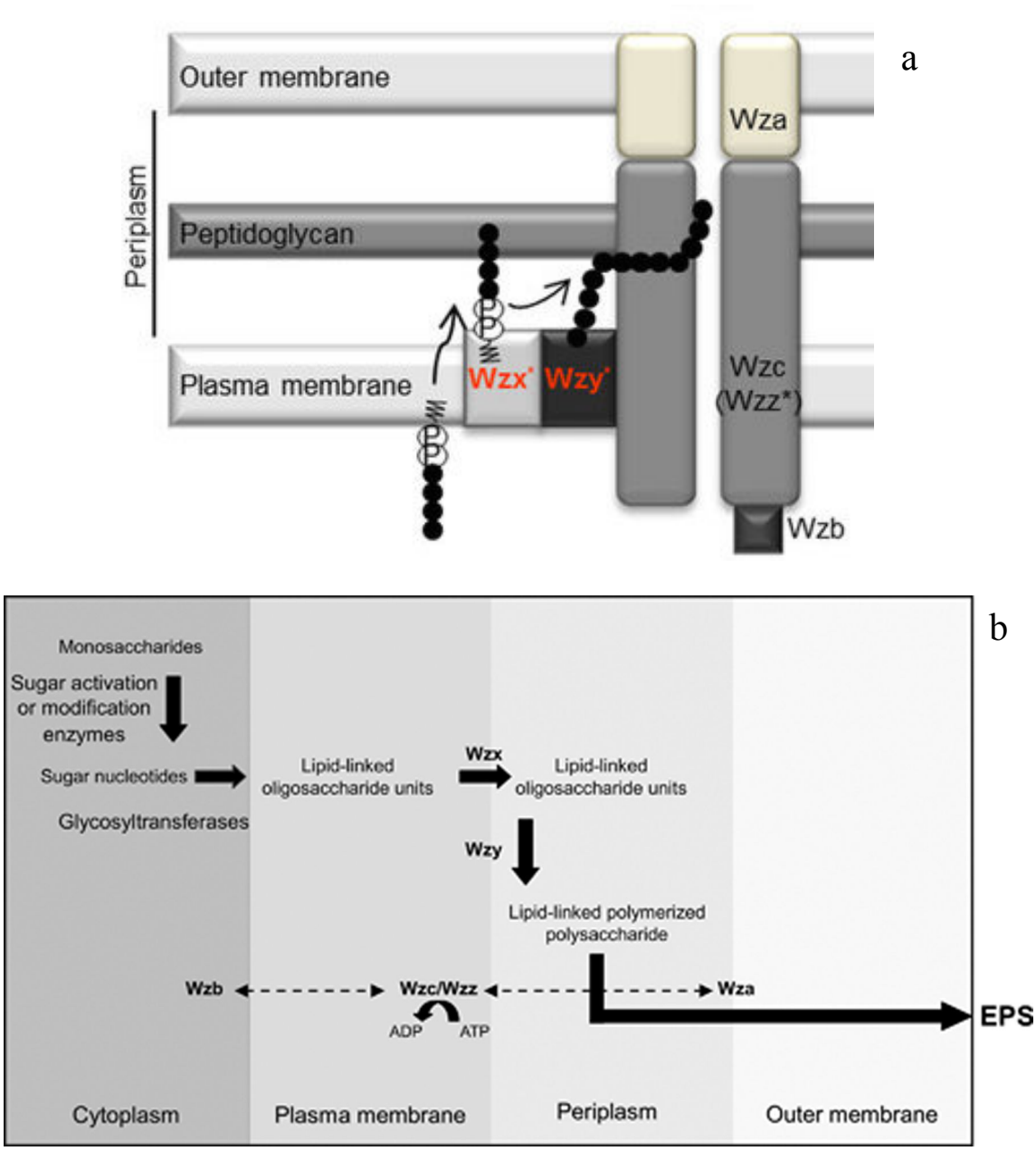


Fig 5.1 The putative Wzy-dependent synthesis pathway**. a** the scheme of each protein involved in Wzy-dependent synthesis pathway. **b** the steps of EPS synthesis in cytosol and the putative function of each protein in Wzy-dependent pathway.

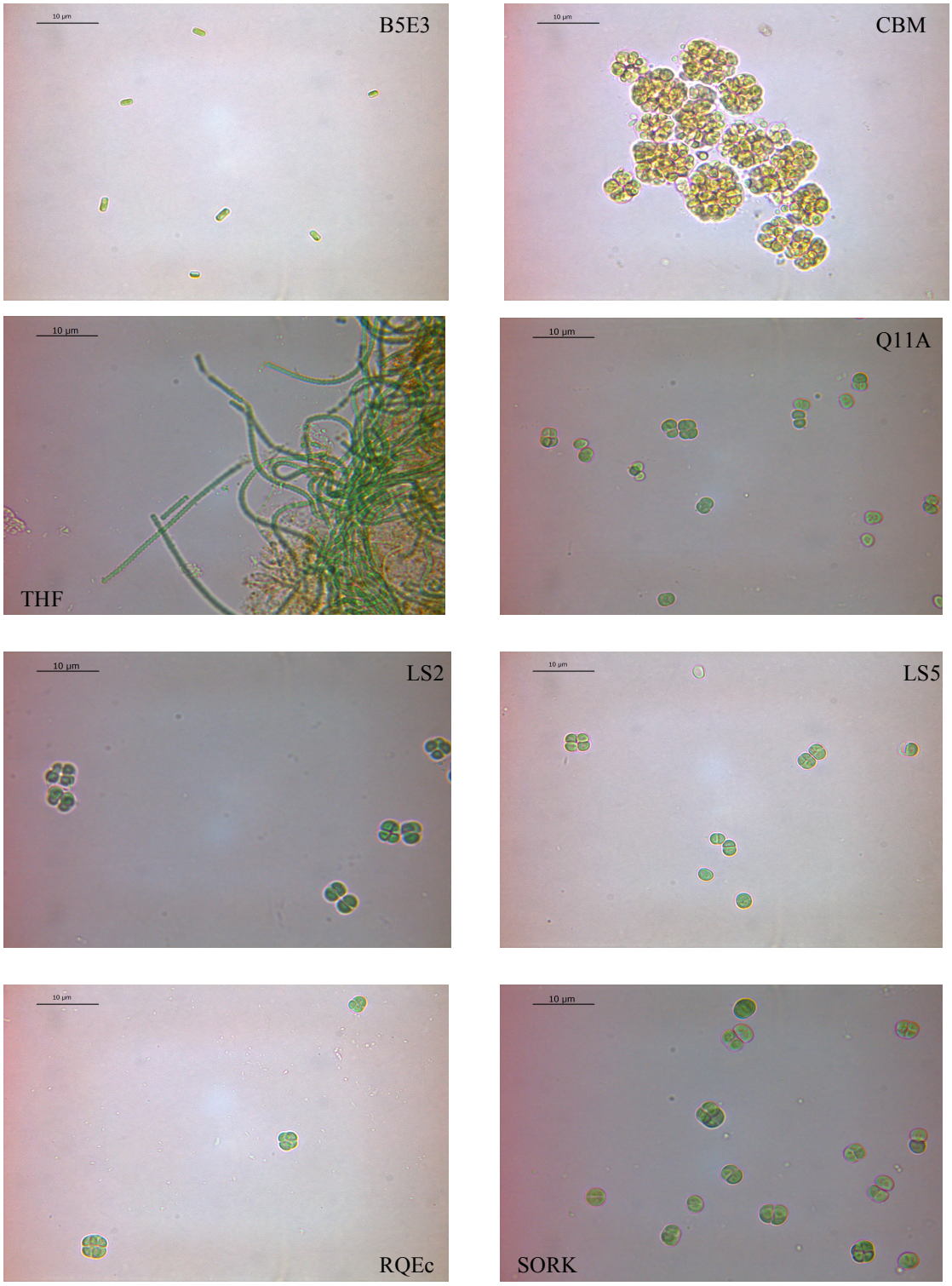


Fig 5.2 Morphology of cyanobacteria isolates from Dry Valleys (B5E3, CBM and THF), Tibet deserts (LS2, LS5 and Q11A) and China deserts (RQEc and SORK).

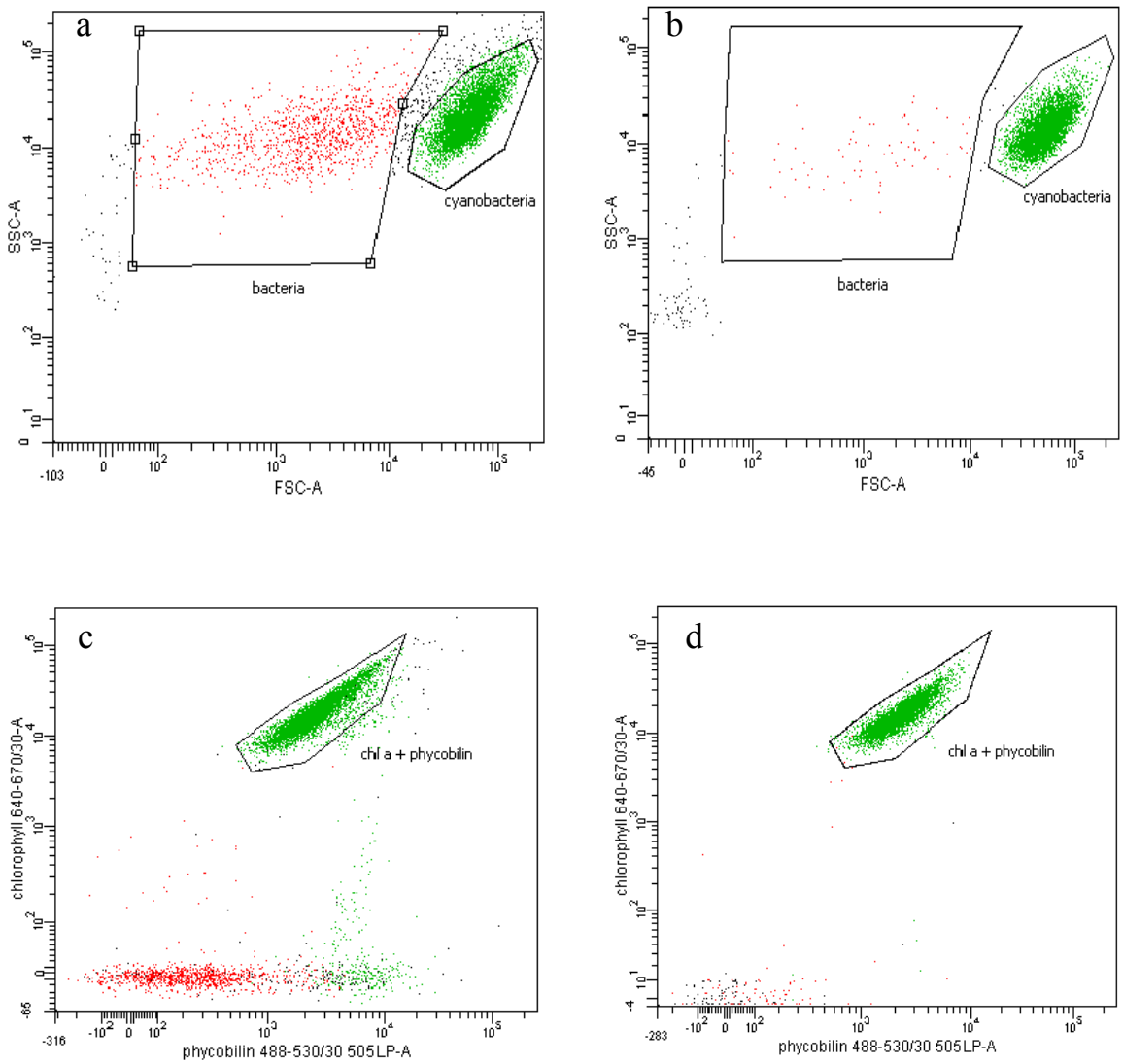


Fig 5.3 Flow cytometric analysis of the Dry Valley culture CBM. A doublet discrimination gate of FSC-Area vs. SSC-Area showed the separate populations of bacteria and cyanobacteria (**a & b**); Phycobilin vs. Chl*a* fluorescence cytogram indicated that cells exhibited high fluorescence of both pigments also fell within the FSC-SSC gates were cyanobacteria (**c & d**). The approximate cell count within the bacterial and cyanobacterial gates was recorded before (**a, c**) and after sorting (**b**, **d**). Cyanobacteria purity increased from 73.7% to 94.9% whist bacteria population was down to 4.9% after sorting.

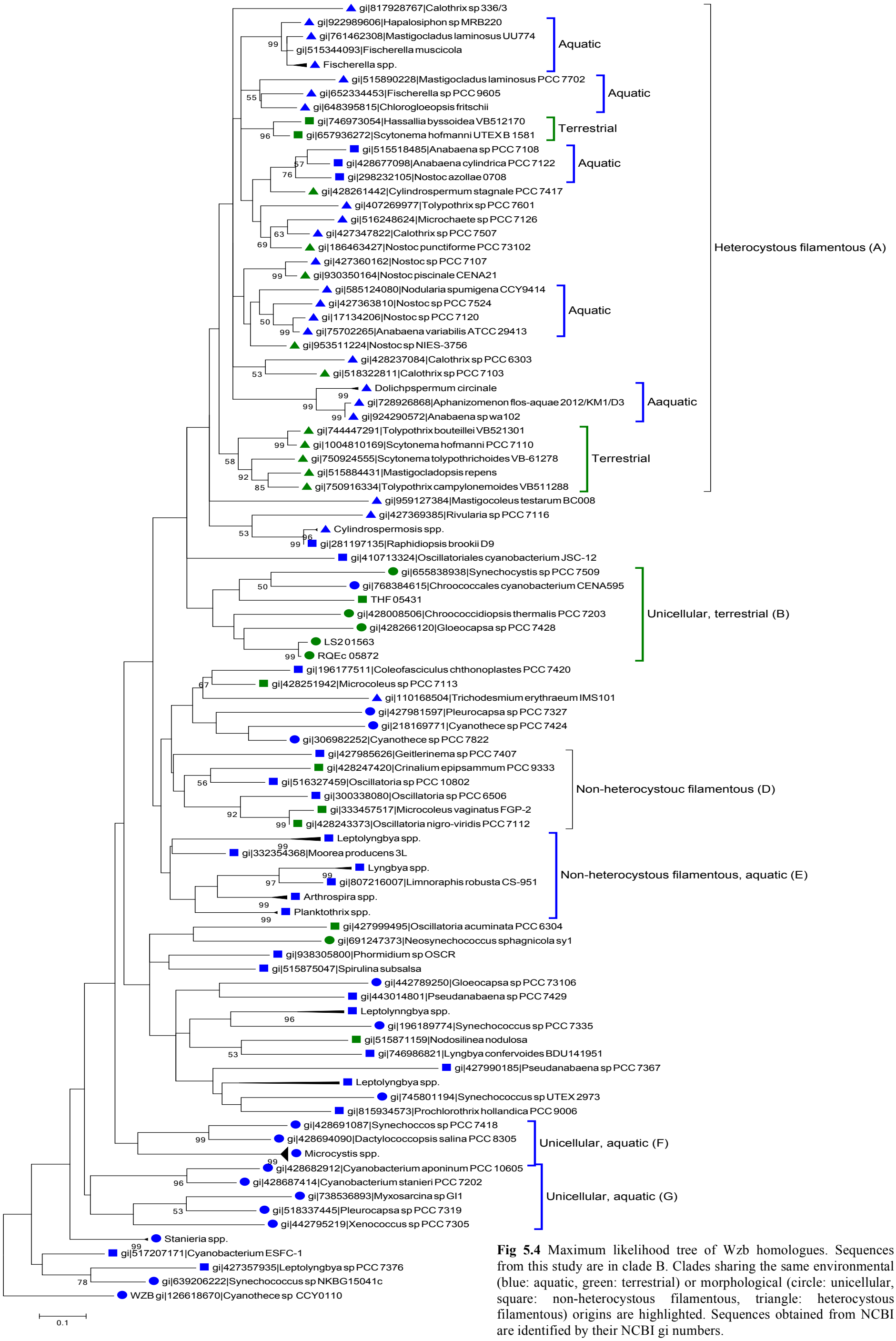


Fig 5.4 Maximum likelihood tree of Wzb homologues. Sequences from this study are in clade B. Clades sharing the same environmental (blue: aquatic, green: terrestrial) or morphological (circle: unicellular, square: non-heterocystous filamentous, triangle: heterocystous filamentous) origins are highlighted. Sequences obtained from NCBI are identified by their NCBI gi numbers.

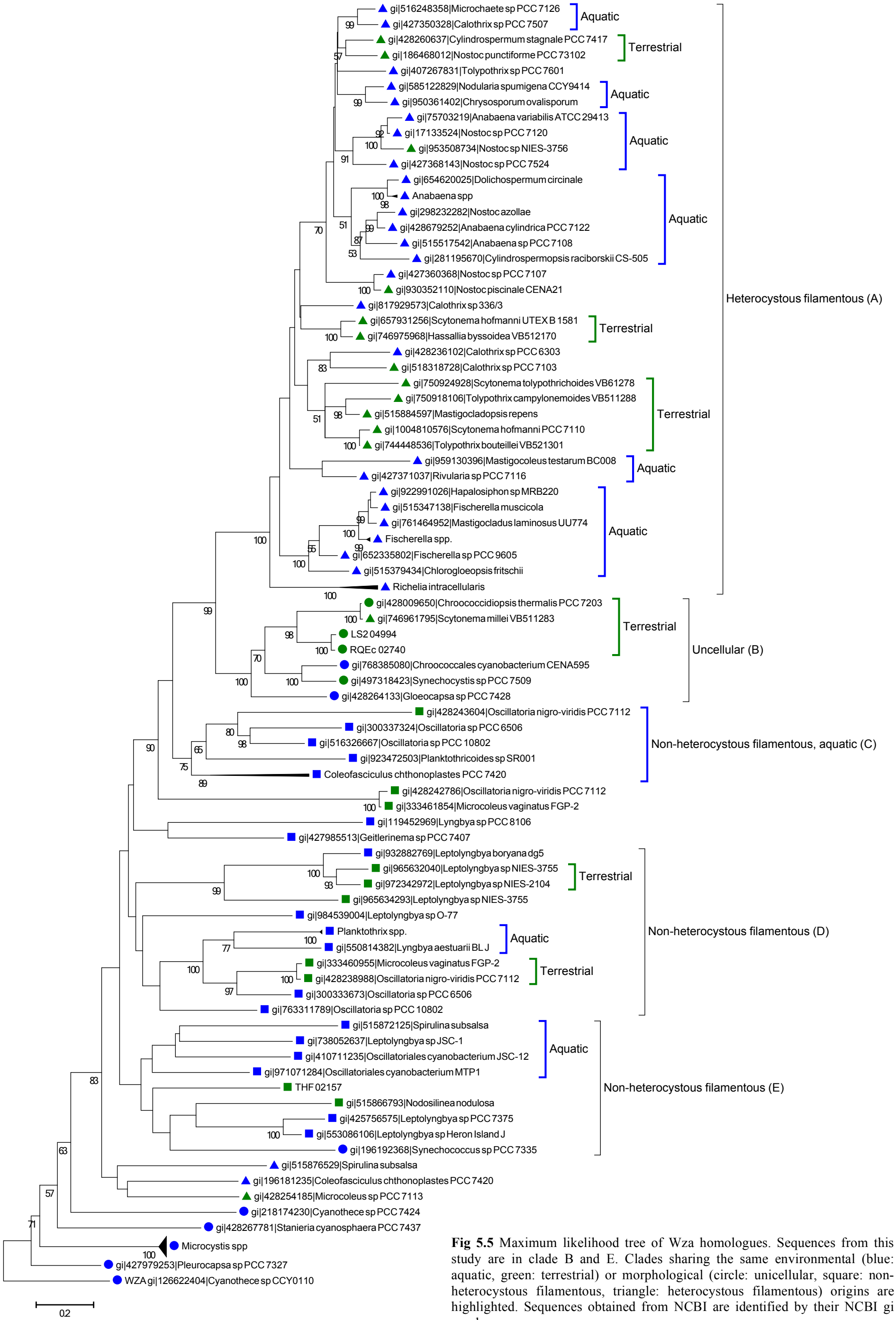


Fig 5.5 Maximum likelihood tree of Wza homologues. Sequences from this study are in clade B and E. Clades sharing the same environmental (blue: aquatic, green: terrestrial) or morphological (circle: unicellular, square: non-heterocystous filamentous, triangle: heterocystous filamentous) origins are highlighted. Sequences obtained from NCBI are identified by their NCBI gi numbers.

Fig 5.6 Maxium likelihood tree of Wzx homologues. Sequences from this study are in clade A to D, and F and G. Clades sharing the same environmental (blue: aquatic, green: terrestrial) or morphological (circle: unicellular, square: non-heterocystous filamentous, triangle: heterocystous filamentous) origins are highlighted. Sequences with polysacc_synt domain are labelled. Others are sequences with Polysacc_synt_3 domain. Sequences obtained from NCBI are identified by their NCBI gi numbers

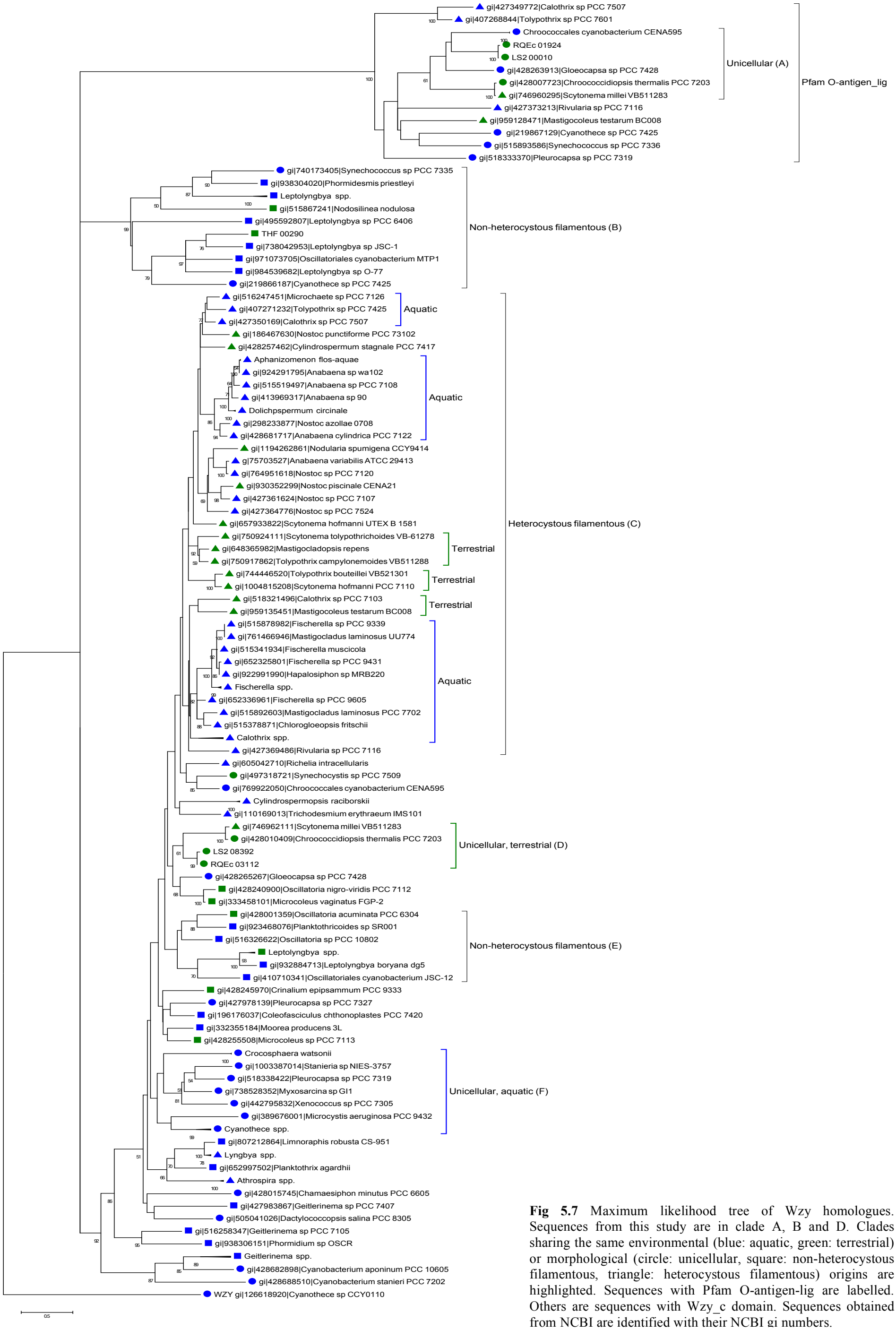


Fig 5.7 Maximum likelihood tree of Wzy homologues. Sequences from this study are in clade A, B and D. Clades sharing the same environmental (blue: aquatic, green: terrestrial) or morphological (circle: unicellular, square: non-heterocystous filamentous, triangle: heterocystous filamentous) origins are highlighted. Sequences with Pfam O-antigen-lig are labelled. Others are sequences with Wzy_c domain. Sequences obtained from NCBI are identified with their NCBI gi numbers.

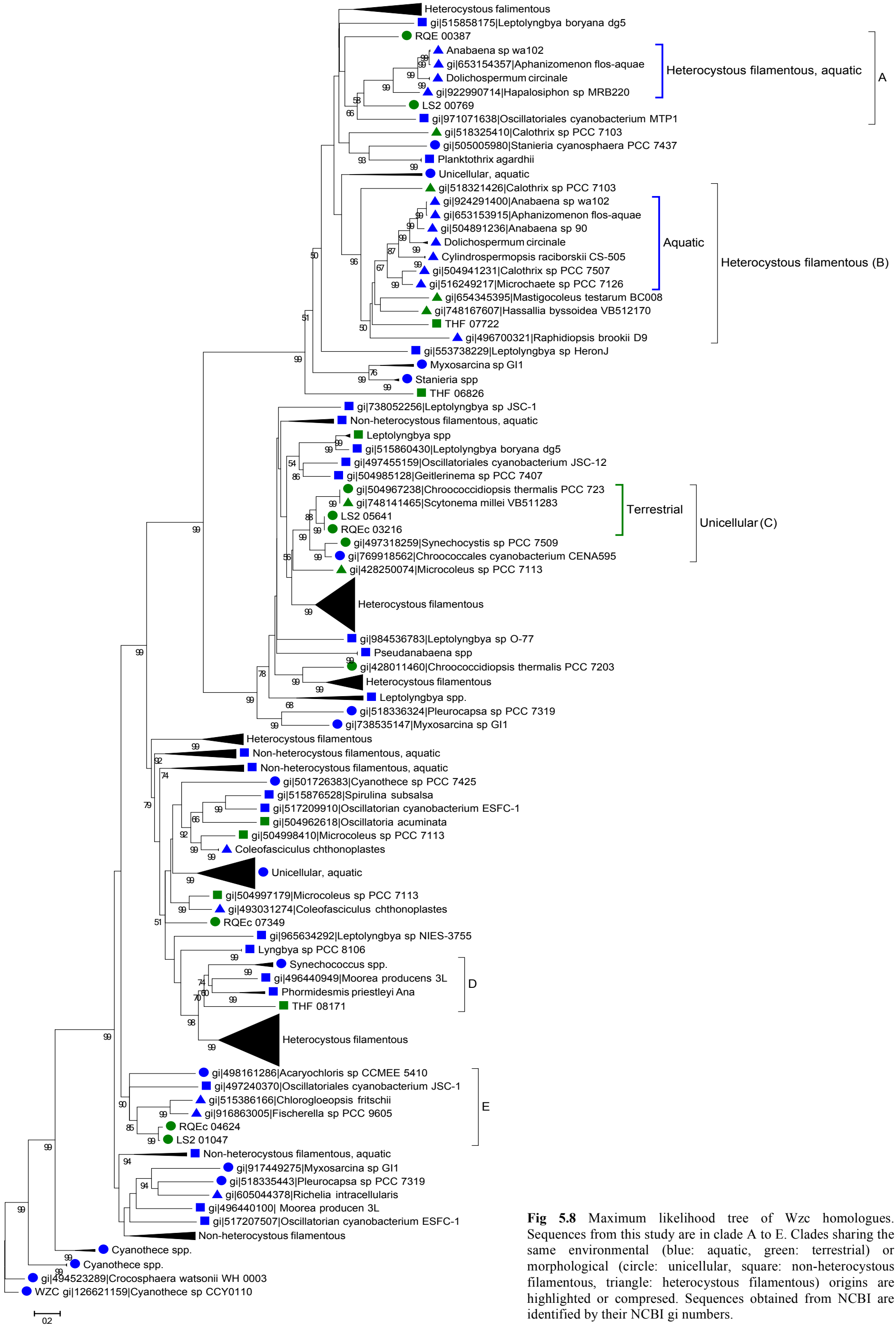


Fig 5.8 Maximum likelihood tree of Wzc homologues. Sequences from this study are in clade A to E. Clades sharing the same environmental (blue: aquatic, green: terrestrial) or morphological (circle: unicellular, square: non-heterocystous filamentous, triangle: heterocystous filamentous) origins are highlighted or compresed. Sequences obtained from NCBI are identified by their NCBI gi numbers.

Chapter 6 Thesis conclusions

This thesis describes the use of two metagenomic approaches, DNA functional microarrays (GeoChip) and next generation sequencing (Miseq, Illummina) to elucidate the functional gene diversity and potential ecological role of microbial communities in the McMurdo Dry Valleys ecosystem of Antarctica.

The literature review begins with a discussion of the different definitions of drylands and an introduction to their ecosystems. In hyper-arid deserts such as the McMurdo Dry Valleys, higher plant and animal life is restricted by severe environmental stress (desiccation, thermal extreme and UV stress). Microorganisms have been discovered in cryptic habitats within and beneath rocks, as a stress avoidance strategy, and they comprise the most abundant biomass in this ecosystem. The microbial diversity and community structure of hypoliths, endoliths and soils have been widely surveyed using 16S rRNA gene clone libraries and next-generation sequencing. A major research gap (addressed by this thesis) is the question of what the functional roles of these microbial communities are, and also the cellular mechanisms that mediate stress tolerance.

In this thesis the putative functional capacity of microbial communities in Miers Valley, a long-term study site in the McMurdo Dry Valleys, was investigated. Miers Valley is a maritime-influenced valley and dominated by granite and moraine substrates where chasmoendoliths and hypoliths occurred, respectively. In this oligotrophic ecosystem, the putative carbon and nitrogen cycling pathways were established using GeoChip analysis. This demonstrated that among different habitats (hypolith, chasmoendolith and soil) the major contributors of each specific metabolism were discernible. An overall summary of how the GeoChip investigation informed the possible geobiological transformations in the Dry Valleys is shown in Figure 6.1 & 6.2. In carbon transformations, the signatures of different forms of *rubisco* indicated that archaea, Proteobacteria, Bacteroidetes and Cyanobacteria (in hypoliths especially) were

responsible for carbon fixation; methane oxidation was primarily contributed by Proteobacteria and Verrucomicrobia; anaerobic pathways, such as acetogenesis were largely contributed by archaea, and methanogenesis in lithic habitats was performed by Proteobacteria, Actinobacteria and Firmicutes. Putative nitrogen cycling pathways were all identified, and in general Proteobacteria were involved in most of the pathways. Nitrogen fixation and denitrification were indicated by archaeal phyla. Fungi and *Deinococcus*-Thermus conducted ammonification and nitrification, respectively. Assimilatory nitrate reduction is largely performed by Firmicutes and Verrucomicrobia whereas DNRA was conducted by Bacteroidetes, Firmicutes and Lentisphaerae. Planctomycetes contributed ANAMMOX, as expected.

Stress response genes were very widespread in Miers Valley bacterial communities, especially in phyla Actinobacteria, Bacteroidetes, Cyanobacteria and Proteobacteria. On the other hand, fungal and archaeal phyla largely displayed pathways related to nutrient limitation. Further insight will come from landscape-scale *in situ* studies, and interrogation of the cellular basis for stress tolerance *in vitro*. However, in natural environments where microorganisms are usually exposed to diverse signals and environmental factors (abiotic and spatial restrictions). Thus, understanding the intracellular process, community interaction and functionality will be challenging (Ackermann, 2013).

In McKelvey Valley functional genes were also widespread (Chan et al., 2013) but the less extreme Miers Valley supported relatively greater diversity in all pathways. This leads to a hypothesis that less selective pressure may actually introduce multiple redundancies in metabolic pathways. Furthermore, this may also indicate a more profound difference in the way community stress tolerance is mediated: with individual taxa eliciting their own responses under more extreme stress, but in less extreme environments the community-wide benefits accrue from responses among a few taxa.

The GeoChip also identified antibiotic resistance genes and phage diversity, providing a new insight of biotic controls on Dry Valleys microbial communities, although a metagenome study has suggested that interspecific competition might not be a major driver in shaping community structures (Fierer et al., 2012). However, the distinct phage diversity might indicate a possible bottom-up controls on communities structure and biomass among Dry Valleys.

The GeoChip provides a broad range of functional gene screening by detecting gene loci (He et al., 2007; Vieites, Guazzaroni, Beloqui, Golyshin, & Ferrer, 2009), but there are some limitations of this functional microarray. First, novel genes cannot be identified and acquired as functional genes that targeted by GeoChip probes are derived from pre-existing DNA sequences. The second issue is specificity and sensitivity (He, Deng, & Zhou, 2012). Apart from the hybridization condition, probe designing in microarray is challenging. Longer probes provide stronger signal but also lower the specificity to homologous gene within a sample, causing cross-hybridization issues (Roh, Abell, Kim, Nam, & Bae, 2010). Oligonucleotide probes designed in GeoChip are in length of 50bp for the best confidence of detection (He et al., 2007). Another limitation of microarray is the requirement of large amount of DNA extracted from the environmental samples: GeoChip 3.0 requiring at least 2 μg of environmental DNA (He et al., 2010; Roh et al., 2010). This can be problematic for polar soils samples where there is low recoverable DNA.

GeoChip also contains a phylogenetic marker, *gyrB* gene, providing a better evolutionary resolution at species/ strain level (He et al., 2010). However, novel species and true microbial community structure cannot be revealed. Therefore, the NGS technology has complemented these major disadvantages of microarray and is suitable for discovering novel gene diversity in microbial communities, requiring no information of any known DNA sequences by massively parallel sequencing 16S rRNA genes,

functional genes or whole genomes with low DNA input. Preparation of libraries for large numbers of samples in NGS can be time consuming, although the advantage of multiplexing using multiple barcodes may offset this (Roh et al., 2010). In this thesis, an attempt was made to utilize *de novo* metagenome sequencing in chapter five, mainly focusing on identification of cyannobacterial EPS synthesis genes devried from deserts.

Protein homologues (WZA, WZB, WZC, WZX and WZY) involved in WZY-dependent pathways and a single homologue (KpsM) involved in ABC transporter pathways from Tibetan (strain LS2, *Chroococcidiopsis*-like), Chinese (strain ROEc, *Chroococcidiopsis*-like) and the Dry Valleys (strain THF, *Lyptolyngbya*-like) cyanobacteria were identified (Fig 6.3). Homologues related to synthase-dependant pathways were failed to identify might be due to low coverage of sequence output or poor assemblies of contigs. Comprehensive phylogenetic analyses of sequences related to WZY-dependant pathways have indicated that these genes may not be instrumental in adaptation to environmental stress.

The original sequencing strategy in this study was culture-based isolation. The environmental sample was enriched and followed by single colony isolation. The axenic culture was then subjected to genome sequencing to obtain a clonal genome (Blainey, 2013). Unfortunately, the axenic cultures of cyanobacteria in this study could not be obtained by colony isolation and antibiotic selection. Pure cyanobacterial cells failed to collect using flow cytometery due to the presence of EPS. A further consideration is therefore removal of EPS since several studies have reported the methods for removing EPS from bacteria (Jachlewski et al., 2015), algae (Staats, De Winder, Stal, & Mur, 1999) and cyanobacteria (Klock, Wieland, Seifert, & Michaelis, 2007; Tamaru et al., 2005) physically without cell impairment. In future, these EPS-removed cyanobacterial cells will be applied to single cell genome sequencing to obtain single-cell genome (Blainey, 2013). The advantages of single-cell sequencing are cultivation-free

experimental work and no compositing data from multiple cells or strains, but this approach does involve demanding sample preparation and contamination is another key challenge (Blainey, 2013). Nevertheless, since EPS synthesis genes were identified in this thesis, it would be interesting to test which genes were essential and what the cellular response would be to environmental stress *in vitro* under deletion of these genes.

Overall this thesis has made a significant novel contribution to understanding the putative functional role of Antarctic cryptic communities in rock and soil, providing new insight on the diversity of functional pathways and how they may reflect the differential community assembly observed between these niches. Perhaps the most exciting discoveries were the high level of functional redundancy in microbial communities, implying that communities may be strongly resilient to change. The differential adaptation to stress at the community level between soil and rock substrates also provides new insight on how physiological traits impact community assembly in this extreme environment. With the increasing access and cost-effectiveness of metagenomic techniques, future research to more fully elucidate the functional role of these communities will undoubtedly provide new advances.

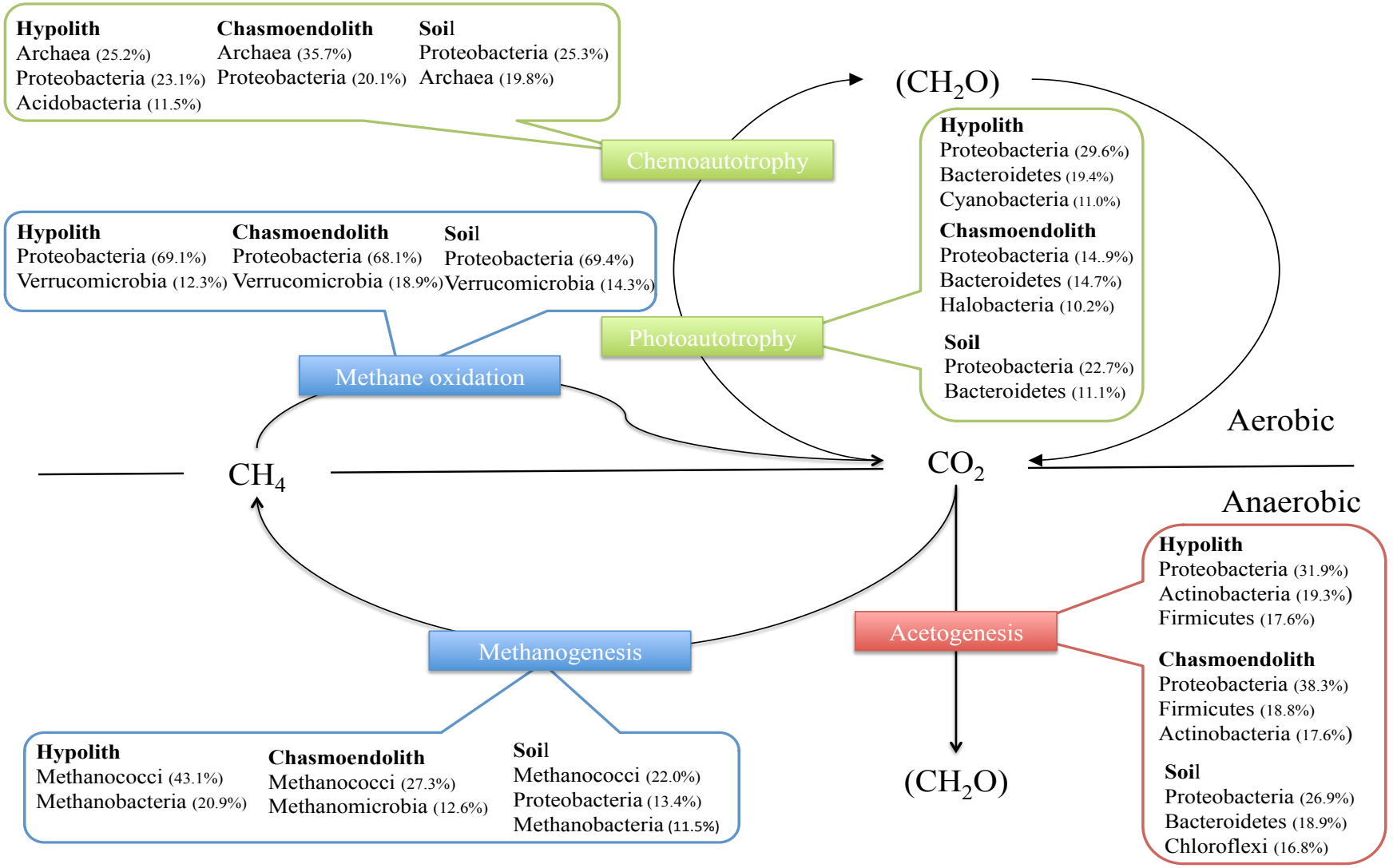


Fig 6.1 Diagram to illustrate overall carbon cycling in hypolithic, chasmoendolithic and soil communities in Miers Valley. This synthesis was based upon all the results present in this thesis. Only the signal intensity of specific pathway under the given taxa over than 10% was shown.

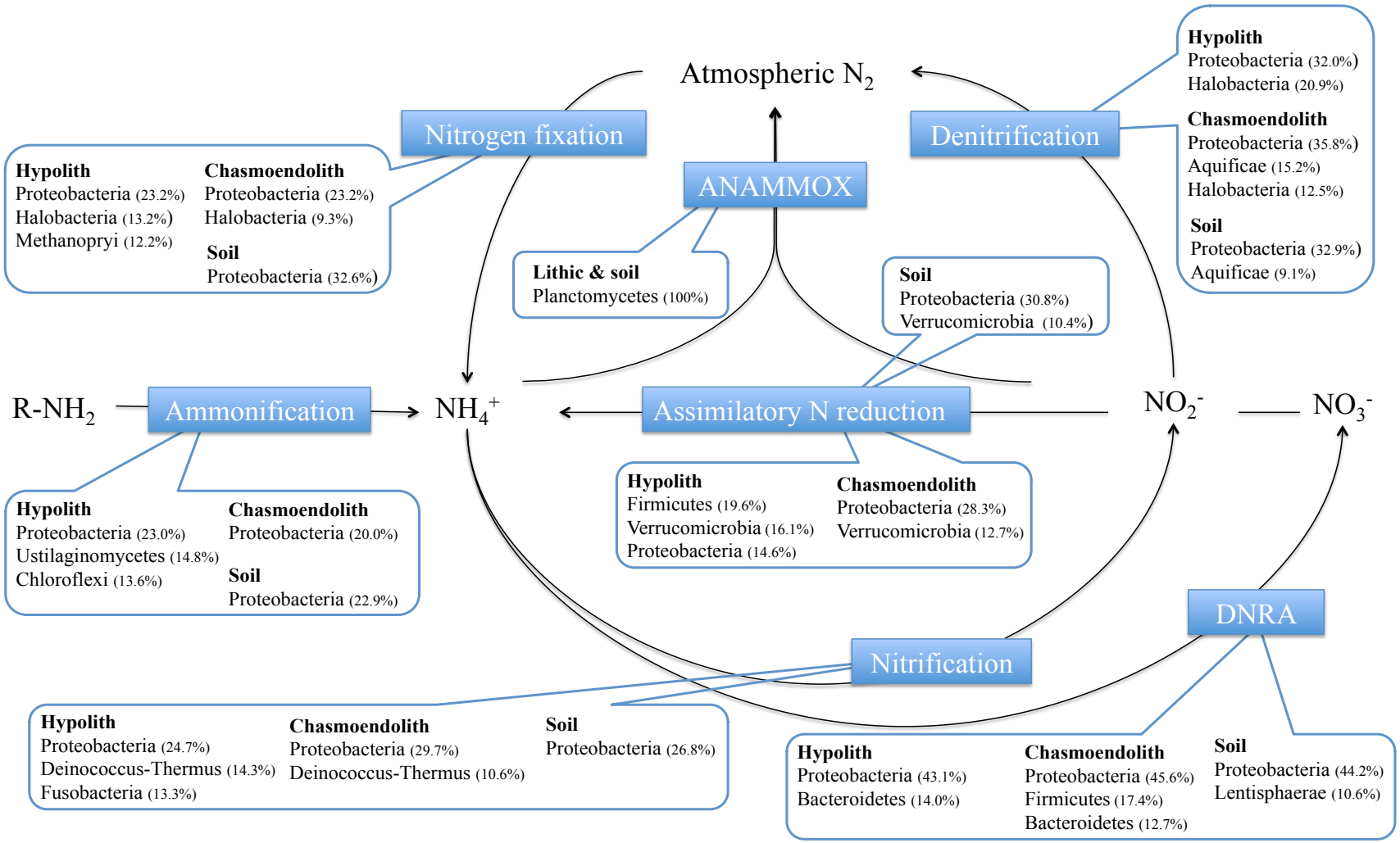


Fig 6.2 Diagram to illustrate overall nitrogen cycling in hypolithic, chasmoendolithic and soil communities in Miers Valley. This synthesis was based upon all the results present in this thesis. Only the signal intensity of specific pathway under the given taxa over than 10% was shown.

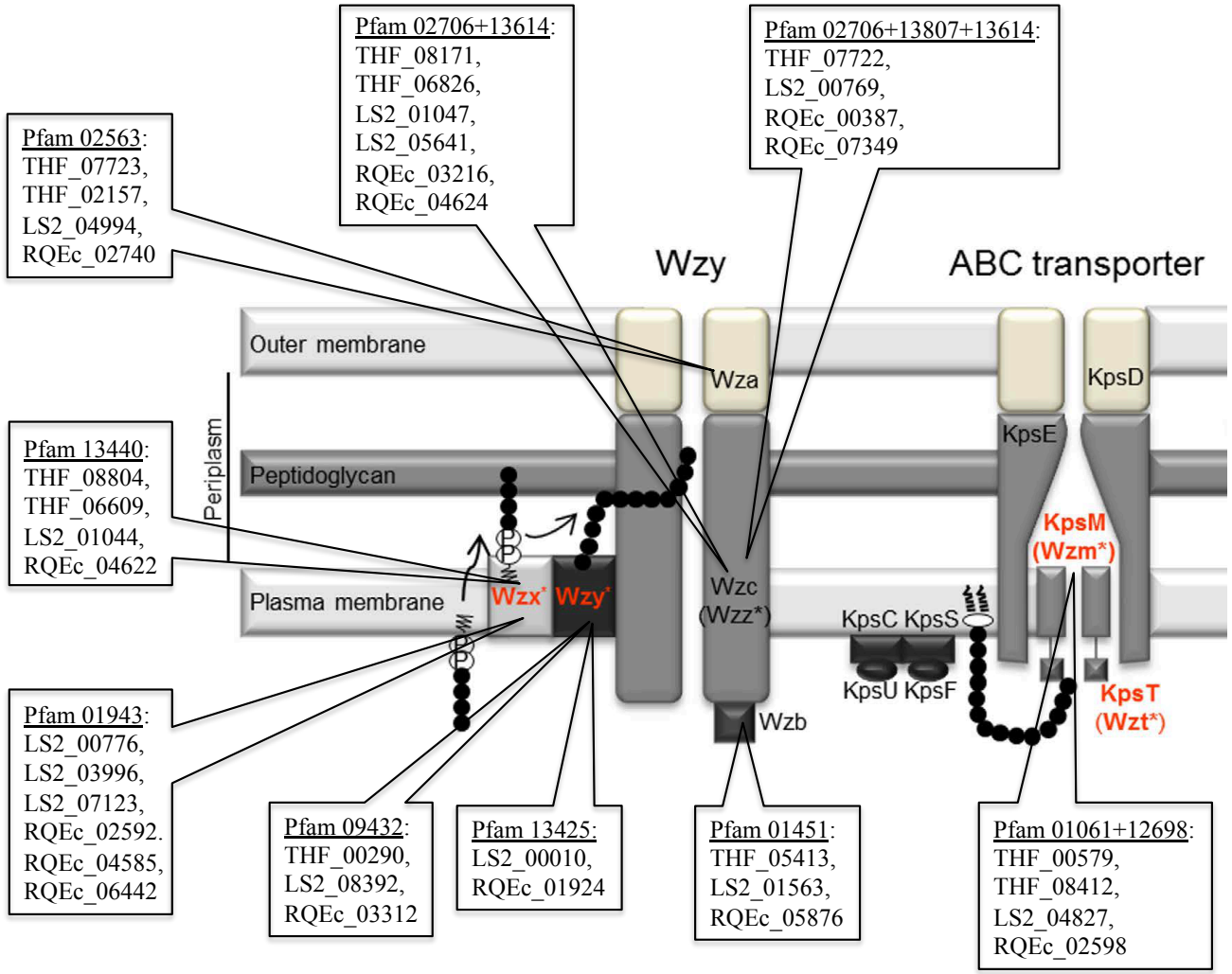


Fig 6.3 Diagram to illustrate contigs identified as putative protein homologues involved in WZY-dependent and ABC transporter pathway. Session number of Pfam functional domains is also indicated.

References

Ackermann, M. (2013). Microbial individuality in the natural environment. *The ISME Journal*, *7*(3), 465–7. http://doi.org/10.1038/ismej.2012.131

Adriaenssens, E. M., Van Zyl, L., De Maayer, P., Rubagotti, E., Rybicki, E., Tuffin, M., & Cowan, D. a. (2014). Metagenomic analysis of the viral community in Namib Desert hypoliths. *Environmental Microbiology*, *17*(2), 480–95. http://doi.org/10.1111/1462-2920.12528

Aislabie, J. M., Chhour, K.-L., Saul, D. J., Miyauchi, S., Ayton, J., Paetzold, R. F., & Balks, M. R. (2006). Dominant bacteria in soils of Marble Point and Wright Valley, Victoria Land, Antarctica. *Soil Biology and Biochemistry*, *38*(10), 3041–3056. Journal Article. http://doi.org/http://dx.doi.org/10.1016/j.soilbio.2006.02.018

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–10. http://doi.org/10.1016/S0022-2836(05)80360-2

Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, *59*(1), 143–69.

Arenz, B. E., & Blanchette, R. a. (2011). Distribution and abundance of soil fungi in Antarctica at sites on the Peninsula, Ross Sea Region and McMurdo Dry Valleys. *Soil Biology and Biochemistry*, *43*(2), 308–315. http://doi.org/10.1016/j.soilbio.2010.10.016

Arenz, B. E., Held, B. W., Jurgens, J. A., Farrell, R. L., & Blanchette, R. A. (2006). Fungal diversity in soils and historic wood from the Ross Sea Region of Antarctica. *Soil Biology and Biochemistry*, *38*(10), 3057–3064. http://doi.org/10.1016/j.soilbio.2006.01.016

Azúa-Bustos, A., González-Silva, C., Mancilla, R. A., Salas, L., Gómez-Silva, B., McKay, C. P., & Vicuña, R. (2011). Hypolithic cyanobacteria supported mainly by fog in the coastal range of the Atacama Desert. *Microbial Ecology*, *61*(3), 568–81. http://doi.org/10.1007/s00248-010-9784-5

Azua-Bustos, A., Urrejola, C., & Vicuña, R. (2012). Life at the dry edge: microorganisms of the Atacama Desert. *FEBS Letters*, *586*(18), 2939–45. http://doi.org/10.1016/j.febslet.2012.07.025

Badger, M. R., & Bek, E. J. (2008). Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. *Journal of Experimental Botany*, *59*(7), 1525–41. http://doi.org/10.1093/jxb/erm297

Bahl, J., Lau, M. C. Y., Smith, G. J. D., Vijaykrishna, D., Cary, S. C., Lacap, D. C., … Pointing, S. B. (2011). Ancient origins determine global biogeography of hot and cold desert cyanobacteria. *Nature Communications*, *2*, 163. http://doi.org/10.1038/ncomms1167

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., …

Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, *19*(5), 455–77. http://doi.org/10.1089/cmb.2012.0021

Bargagli, R. (2005). Antarctica: geomorphology and climate trends. In R. Bargagli (Ed.), *Antarctic ecosystems: environmental contamination, climate change and human impact* (pp. 1–41). Book, New York: Springer.

Belnap, J. (2003). Comparative structure of physical and biological soil crusts. In J. Belnap & O. L. Lange (Eds.), *Biological Soil Crusts: Structure, Function, and Management* (pp. 177–192). Berline: Springer-Verlag.

Belnap, J. (2006). The potential roles of biological soil crusts in dryland hydrologic cycles. *Hydrological Processes*, *20*(15), 3159–3178. http://doi.org/10.1002/hyp.6325

Belnap, J., Büdel, B., & Lange, O. L. (2003). Biological soil crusts: characteristics and distribution. In J. Belnap & O. L. Lange (Eds.), *Biological Soil Crusts: Structure, Function, and Management* (pp. 3–30). Berlin: Springer-Verlag. http://doi.org/10.1007/978-3-642-56475-8_1

Berg, I. A., Kockelkorn, D., Ramos-Vera, W. H., Say, R. F., Zarzycki, J., Hügler, M., … Fuchs, G. (2010). Autotrophic carbon fixation in archaea. *Nature Reviews. Microbiology*, *8*(6), 447–60. http://doi.org/10.1038/nrmicro2365

Billi, D., Friedmann, E. I., Hofer, K. G., Caiola, M. G., & Ocampo-Friedmann, R. (2000). Ionizing-radiation resistance in the desiccation-tolerant cyanobacterium Chroococcidiopsis. *Applied and Environmental Microbiology*, *66*(4), 1489–92.

Billi, D., & Potts, M. (2002). Life and death of dried prokaryotes. *Research in Microbiology*, *153*(1), 7–12.

Billi, D., Viaggiu, E., Cockell, C. S., Rabbow, E., Horneck, G., & Onofri, S. (2011). Damage escape and repair in dried Chroococcidiopsis spp. from hot and cold deserts exposed to simulated space and martian conditions. *Astrobiology*, *11*(1), 65–73. http://doi.org/10.1089/ast.2009.0430

Blainey, P. C. (2013). The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiology Reviews*, *37*(3), 407–27. http://doi.org/10.1111/1574-6976.12015

Blaut, M. (1994). Metabolism of methanogens. *Antonie van Leeuwenhoek*, *66*(1–3), 187–208.

Boor, K. J. (2006). Bacterial stress responses: what doesn't kill them can make then stronger. *PLoS Biology*, *4*(1), e23. http://doi.org/10.1371/journal.pbio.0040023

Bottos, E. M., Woo, A. C., Zawar-Reza, P., Pointing, S. B., & Cary, S. C. (2014). Airborne bacterial populations above desert soils of the McMurdo Dry Valleys, Antarctica. *Microbial Ecology*, *67*(1), 120–8. http://doi.org/10.1007/s00248-013-0296-y

Bowker, M. A., Belnap, J., Büdel, B., Sannier, C., Pietrasiak, N., Eldridge, D. J., &

Rivera-Aguilar, V. (2016). Controls on Distribution Patterns of Biological Soil Crusts at Micro- to Global Scales (pp. 173–197). http://doi.org/10.1007/978-3-319-30214-0_10

Boyd, W. L., Rothenbg, I., & Boyd, J. W. (1970). Soil microorganisms at Paradise Harbor, Antarctica. *Ecology*, *51*(6), 1040–1045. Journal Article. http://doi.org/10.2307/1933630

Breitbart, M., & Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology*, *13*(6), 278–84. http://doi.org/10.1016/j.tim.2005.04.003

Breitbart, M., Wegley, L., Leeds, S., Schoenfeld, T., & Rohwer, F. (2004). Phage Community Dynamics in Hot Springs. *Appl Environ Microbiol*, *70*(3), 1633–1640. http://doi.org/10.1128/AEM.70.3.1633

Buchen-Osmond, C. (2003). The universal virus database ICTVdB. *Computing in Science & Engineering*, *5*(3), 16–25. http://doi.org/10.1109/MCISE.2003.1196303

Büdel, B., Bendix, J., Bicker, F. R., & Allan Green, T. G. (2008). Dewfall as a water source frequently activates the endolithic cyanobacterial communities in the granites of Taylor Valley, Antarctica. *Journal of Phycology*, *44*(6), 1415–1424. http://doi.org/10.1111/j.1529-8817.2008.00608.x

Budel, B., Weber, B., Kuhl, M., Pfanz, H., Sultemeyer, D., & Wessels, D. (2004). Reshaping of sandstone surfaces by cryptoendolithic cyanobacteria: bioalkalization causes chemical weathering in arid landscapes. *Geobiology*, *2*(4), 261–268. http://doi.org/10.1111/j.1472-4677.2004.00040.x

Bugmann, H. K. M., & Solomon, A. M. (2000). Explaing forest composition and biomass across multiple biogeographical regions. *Ecological Applications*, *10*(1), 95–114. http://doi.org/10.1890/1051-0761(2000)010[0095:EFCABA]2.0.CO;2

Cáceres, L., Gómez-Silva, B., Garró, X., Rodríguez, V., Monardes, V., & McKay, C. P. (2007). Relative humidity patterns and fog water precipitation in the Atacama Desert and biological implications. *Journal of Geophysical Research*, *112*(G4), G04S14. http://doi.org/10.1029/2006JG000344

Calvin, M., & Benson, A. A. (1948). The Path of Carbon in Photosynthesis. *Science*, *107*(2784), 476–480. http://doi.org/10.1126/science.107.2784.476

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*(5), 335–6. http://doi.org/10.1038/nmeth.f.303

Caruso, T., Chan, Y., Lacap, D. C., Lau, M. C. Y., McKay, C. P., & Pointing, S. B. (2011). Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. *The ISME Journal*, *5*(9), 1406–1413. http://doi.org/10.1038/ismej.2011.21

Cary, S. C., McDonald, I. R., Barrett, J. E., & Cowan, D. A. (2010). On the rocks: the microbiology of Antarctic Dry Valley soils. *Nature Reviews. Microbiology*,

8(2), 129–38. Journal Article. http://doi.org/10.1038/nrmicro2281

Casanueva, A., Tuffin, M., Cary, C., & Cowan, D. A. (2010). Molecular adaptations to psychrophily: the impact of "omic" technologies. *Trends in Microbiology*, *18*(8), 374–81. http://doi.org/10.1016/j.tim.2010.05.002

Cellamare, M., Rolland, A., & Jacquet, S. (2009). Flow cytometry sorting of freshwater phytoplankton. *Journal of Applied Phycology*, *22*(1), 87–100. http://doi.org/10.1007/s10811-009-9439-4

Chan, Y., Lacap, D. C., Lau, M. C. Y., Ha, K. Y., Warren-Rhodes, K. a, Cockell, C. S., … Pointing, S. B. (2012). Hypolithic microbial communities: between a rock and a hard place. *Environmental Microbiology*, *14*(9), 2272–82. http://doi.org/10.1111/j.1462-2920.2012.02821.x

Chan, Y., Van Nostrand, J. D., Zhou, J., Pointing, S. B., & Farrell, R. L. (2013). Functional ecology of an Antarctic Dry Valley. *Proceedings of the National Academy of Sciences*, *110*(22), 8990–8995. Journal Article. http://doi.org/10.1073/pnas.1300643110

Chung, H. J., Bang, W., & Drake, M. A. (2006). Stress Response of Escherichia coli. *Comprehensive Reviews in Food Science and Food Safety*, *5*(3), 52–64. http://doi.org/10.1111/j.1541-4337.2006.00002.x

Clark, A. G. (1994). Invasion and maintenance of a gene duplication. *Proceedings of the National Academy of Sciences*, *91*(8), 2950–2954. http://doi.org/10.1073/pnas.91.8.2950

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Austral Ecology*, *18*(1), 117–143. http://doi.org/10.1111/j.1442-9993.1993.tb00438.x

Cockell, C. S., & Stokes, M. D. (2004). Ecology: widespread colonization by polar hypoliths. *Nature*, *431*(7007), 414. http://doi.org/10.1038/431414a

Cockell, C. S., & Stokes, M. D. (2006). Hypolithic Colonization of Opaque Rocks in the Arctic and Antarctic Polar Desert. *Arctic, Antarctic, and Alpine Research*, *38*(3), 335–342.

Cordero, R. R., Seckmeyer, G., Damiani, A., Riechelmann, S., Rayas, J., Labbe, F., & Laroze, D. (2014). The world's highest levels of surface UV. *Photochemical & Photobiological Sciences : Official Journal of the European Photochemistry Association and the European Society for Photobiology*, *13*(1), 70–81. http://doi.org/10.1039/c3pp50221j

Cowan, D. A., Makhalanyane, T. P., Dennis, P. G., & Hopkins, D. W. (2014). Microbial ecology and biogeochemistry of continental Antarctic soils. *Frontiers in Microbiology*, *5*(April), 154. http://doi.org/10.3389/fmicb.2014.00154

Cowan, D. A., Pointing, S. B., Stevens, M. I., Craig Cary, S., Stomeo, F., & Tuffin, I. M. (2011). Distribution and abiotic influences on hypolithic microbial communities in an Antarctic Dry Valley. *Polar Biology*, *34*(2), 307–311. http://doi.org/10.1007/s00300-010-0872-2

Cowan, D. A., Sohm, J. A., Makhalanyane, T. P., Capone, D. G., Green, T. G. A., Cary, S.

C., & Tuffin, I. M. (2011). Hypolithic communities: Important nitrogen sources in Antarctic desert soils. *Environmental Microbiology Reports*, *3*(5), 581–586. http://doi.org/10.1111/j.1758-2229.2011.00266.x

Cowan, D. A., & Tow, L. A. (2004). Endangered Antarctic environments. *Annual Review of Microbiology*, *58*, 649–690. http://doi.org/10.1146/annurev.micro.57.030502.090811

Cox, M. M., & Battista, J. R. (2005). Deinococcus radiodurans - the consummate survivor. *Nature Reviews. Microbiology*, *3*(11), 882–92. http://doi.org/10.1038/nrmicro1264

Crits-Christoph, A., Robinson, C. K., Barnum, T., Fricke, W. F., Davila, A. F., Jedynak, B., … Diruggiero, J. (2013). Colonization patterns of soil microbial communities in the Atacama Desert. *Microbiome*, *1*(1), 28. http://doi.org/10.1186/2049-2618-1-28

Crombie, A. T., & Murrell, J. C. (2014). Trace-gas metabolic versatility of the facultative methanotroph Methylocella silvestris. *Nature*, *510*(7503), 148–151. http://doi.org/10.1038/nature13192

d'Angelo-Picard, C., Faure, D., Penot, I., & Dessaux, Y. (2005). Diversity of N-acyl homoserine lactone-producing and -degrading bacteria in soil and tobacco rhizosphere. *Environmental Microbiology*, *7*(11), 1796–808. http://doi.org/10.1111/j.1462-2920.2005.00886.x

Dall'Olmo, G., & Karnieli, A. (2010). Monitoring phenological cycles of desert ecosystems using NDVI and LST data derived from NOAA-AVHRR imagery. *International Journal of Remote Sensing*, *23*(19), 4055–4071. http://doi.org/10.1080/01431160110115988

De la Torre, J. R., Goebel, B. M., Friedmann, E. I., & Pace, N. R. (2003). Microbial diversity of cryptoendolithic communities from the McMurdo Dry Valleys, Antarctica. *Applied and Environmental Microbiology*, *69*(7), 3858–3867.

de Los Ríos, A., Ascaso, C., Wierzchos, J., Fernández-Valiente, E., & Quesada, A. (2004). Microstructural characterization of cyanobacterial mats from the McMurdo Ice Shelf, Antarctica. *Applied and Environmental Microbiology*, *70*(1), 569–80.

De los Ríos, A., Cary, C., Cowan, D. D., de los Rios, A., Cary, C., & Cowan, D. D. (2014). The spatial structures of hypolithic communities in the Dry Valleys of East Antarctica. *Polar Biology*, *37*(12), 1823–1833. http://doi.org/10.1007/s00300-014-1564-0

de los Ríos, A., Grube, M., Sancho, L. G., Ascaso, C., Ascaso, C., Wierzchos, J., … Wynn-Williams, D. (2007). Ultrastructural and genetic characteristics of endolithic cyanobacterial biofilms colonizing Antarctic granite rocks. *FEMS Microbiology Ecology*, *59*(2), 386–95. http://doi.org/10.1111/j.1574-6941.2006.00256.x

De Los Ríos, A., Wierzchos, J., & Ascaso, C. (2014). The lithic microbial ecosystems of Antarctica's McMurdo Dry Valleys. *Antarctic Science*, *26*(5), 459–477. http://doi.org/10.1017/S0954102014000194

De Philippis, R., & Vincenzini, M. (1998). Exocellular polysaccharides from

cyanobacteria and their possible applications. *FEMS Microbiology Reviews*, *22*(3), 151–175. http://doi.org/10.1111/j.1574-6976.1998.tb00365.x

DiRuggiero, J., Wierzchos, J., Robinson, C. K., Souterre, T., Ravel, J., Artieda, O., … Ascaso, C. (2013). Microbial colonisation of chasmoendolithic habitats in the hyper-arid zone of the Atacama Desert. *Biogeosciences*, *10*(4), 2439–2450. http://doi.org/10.5194/bg-10-2439-2013

Dong, H., Rech, J. A., Jiang, H., Sun, H., & Buck, B. J. (2007). Endolithic cyanobacteria in soil gypsum: Occurrences in Atacama (Chile), Mojave (United States), and Al-Jafr Basin (Jordan) Deserts. *Journal of Geophysical Research*, *112*(G2), G02030. http://doi.org/10.1029/2006JG000385

Doran, P. T., McKay, C. P., Clow, G. D., Dana, G. L., Fountain, A. G., Nylen, T., & Lyons, W. B. (2002). Valley floor climate observations from the McMurdo Dry Valleys, Antarctica, 1986–2000. *Journal of Geophysical Research: Atmospheres*, *107*(D24), 4772. Journal Article. http://doi.org/10.1029/2001JD002045

Doran, P. T., Priscu, J. C., Lyons, W. B., Walsh, J. E., Fountain, A. G., McKnight, D. M., … Parsons, A. N. (2002). Antarctic climate cooling and terrestrial ecosystem response. *Nature*, *415*(6871), 517–520. Journal Article. http://doi.org/10.1038/nature710

Doran, P. T., Wharton, R. A., & Lyons, W. B. (1994). Paleolimnology of the McMurdo Dry Valleys, Antarctica. *Journal of Paleolimnology*, *10*(2), 85–114. http://doi.org/10.1007/BF00682507

Drlica, K., & Zhao, X. (1997). DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiology and Molecular Biology Reviews : MMBR*, *61*(3), 377–392.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–7. http://doi.org/10.1093/nar/gkh340

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–1. http://doi.org/10.1093/bioinformatics/btq461

Ehling-Schulz, M., Bilger, W., & Scherer, S. (1997). UV-B-induced synthesis of photoprotective pigments and extracellular polysaccharides in the terrestrial cyanobacterium Nostoc commune. *J. Bacteriol.*, *179*(6), 1940–1945.

Elbert, W., Weber, B., Burrows, S., Steinkamp, J., Büdel, B., Andreae, M. O., & Pöschl, U. (2012). Contribution of cryptogamic covers to the global cycles of carbon and nitrogen. *Nature Geoscience*, *5*(7), 459–462. http://doi.org/10.1038/ngeo1486

Espie, G. S., & Kimber, M. S. (2011). Carboxysomes: cyanobacterial RubisCO comes in small packages. *Photosynthesis Research*, *109*(1–3), 7–20. http://doi.org/10.1007/s11120-011-9656-y

Ewing, S. A., Sutter, B., Owen, J., Nishiizumi, K., Sharp, W., Cliff, S. S., … Amundson, R. (2006). A threshold in soil formation at Earth's arid–hyperarid transition. *Geochimica et Cosmochimica Acta*, *70*(21), 5293–5322.

http://doi.org/10.1016/j.gca.2006.08.020

Feng, Y.-N., Zhang, Z.-C., Feng, J.-L., & Qiu, B.-S. (2012). Effects of UV-B radiation and periodic desiccation on the morphogenesis of the edible terrestrial cyanobacterium Nostoc flagelliforme. *Applied and Environmental Microbiology*, *78*(19), 7075–81. http://doi.org/10.1128/AEM.01427-12

Ferris, M. J., & Hirsch, C. F. (1991). Method for isolation and purification of cyanobacteria. *Applied and Environmental Microbiology*, *57*(5), 1448–1452.

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., … Caporaso, J. G. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(52), 21390–5. http://doi.org/10.1073/pnas.1215210110

Fierer, N., Strickland, M. S., Liptzin, D., Bradford, M. A., & Cleveland, C. C. (2009). Global patterns in belowground communities. *Ecology Letters*, *12*(11), 1238–49. http://doi.org/10.1111/j.1461-0248.2009.01360.x

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2015). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285. http://doi.org/10.1093/nar/gkv1344

Fisher, M. L., Allen, R., Luo, Y., & Curtiss, R. (2013). Export of extracellular polysaccharides modulates adherence of the Cyanobacterium synechocystis. *PloS One*, *8*(9), e74514. http://doi.org/10.1371/journal.pone.0074514

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., … Martiny, A. C. (2013). Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(24), 9824–9. http://doi.org/10.1073/pnas.1307701110

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y., & Postlethwait, J. (1999). Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics*, *151*(4), 1531–1545.

Fountain, A. G., Nylen, T. H., Monaghan, A., Basagic, H. J., & Bromwich, D. (2010). Snow in the McMurdo Dry Valleys, Antarctica. *International Journal of Climatology*, *30*(5), 633–642. Journal Article. http://doi.org/10.1002/joc.1933

Fowler, D., Coyle, M., Skiba, U., Sutton, M. A., Cape, J. N., Reis, S., … Voss, M. (2013). The global nitrogen cycle in the twenty-first century. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *368*(1621).

Friedmann, E. I. (1980). Endolithic microbial life in hot and cold deserts. *Origins of Life*, *10*(3), 223–235. JOUR. http://doi.org/10.1007/BF00928400

Friedmann, E. I. (1982). Endolithic microorganisms in the Antarctic cold desert. *Science (New York, N.Y.)*, *215*(4536), 1045–53. http://doi.org/10.1126/science.215.4536.1045

Friedmann, E. I., Kappen, L., Meyer, M. A., & Nienow, J. A. (1993). Long-term

productivity in the cryptoendolithic microbial community of the Ross Desert, Antarctica. *Microbial Ecology*, *25*(1). http://doi.org/10.1007/BF00182129

Friedmann, E. I., & Kibler, A. P. (1980). Nitrogen economy of endolithic microbial communities in hot and cold deserts. *Microbial Ecology*, *6*(2), 95–108. http://doi.org/10.1007/BF02010548

Friedmann, E. I., & Ocampo, R. (1976). Endolithic blue-green algae in the Dry Valleys: primary producers in the antarctic desert ecosystem. *Science (New York, N.Y.)*, *193*(4259), 1247–1249. http://doi.org/10.1126/science.193.4259.1247

Gao, Q., & Garcia-Pichel, F. (2011). Microbial ultraviolet sunscreens. *Nature Reviews Microbiology*. http://doi.org/10.1038/nrmicro2649

Gaston, K. J. (2000). Global patterns in biodiversity. *Nature*, *405*(6783), 220–227. http://doi.org/10.1038/35012228

Glennie, K. W. (1987). Desert sedimentary environments, present and past—A summary. *Sedimentary Geology*, *50*(1–3), 135–165. http://doi.org/10.1016/0037-0738(87)90031-5

Golubic, S., Friedmann, E. I., & Schneider, J. (1981). The lithobiontic ecological niche, with special reference to microorganisms. *Journal of Sedimentary Research*, *51*(2), 475–478. http://doi.org/10.1306/212f7cb6-2b24-11d7-8648000102c1865d

Gorbushina, A. A. (2007). Life on the rocks. *Environmental Microbiology*, *9*(7), 1613–31. http://doi.org/10.1111/j.1462-2920.2007.01301.x

Gougoulias, C., Clark, J. M., & Shaw, L. J. (2014). The role of soil microbes in the global carbon cycle: tracking the below-ground microbial processing of plant-derived carbon for manipulating carbon dynamics in agricultural systems. *Journal of the Science of Food and Agriculture*, *94*(12), 2362–71. http://doi.org/10.1002/jsfa.6577

Grant, W. D. (2004). Life at low water activity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1448), 1249-66–7. http://doi.org/10.1098/rstb.2004.1502

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, *29*(8), 1072–5. http://doi.org/10.1093/bioinformatics/btt086

Hartley, A. J. (2003). Andean uplift and climate change. *Journal of the Geological Society*, *160*(1), 7–10. http://doi.org/10.1144/0016-764902-083

He, Z., Deng, Y., Van Nostrand, J. D., Tu, Q., Xu, M., Hemme, C. L., … Zhou, J. (2010). GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *The ISME Journal*, *4*(9), 1167–1179. http://doi.org/10.1038/ismej.2010.46

He, Z., Deng, Y., & Zhou, J. (2012). Development of functional gene microarrays for microbial community analysis. *Curr Opin Biotechnol*, *23*(1), 49–55. Journal Article.

http://doi.org/10.1016/j.copbio.2011.11.001

He, Z., Gentry, T. J., Schadt, C. W., Wu, L., Liebich, J., Chong, S. C., … Zhou, J. (2007). GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *The ISME Journal*, *1*(1), 67–77. http://doi.org/10.1038/ismej.2007.2

Hernandez, D., François, P., Farinelli, L., Osterås, M., & Schrenzel, J. (2008). De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Research*, *18*(5), 802–809. http://doi.org/10.1101/gr.072033.107

Herrero, A., Muro-Pastor, A. M., & Flores, E. (2001). Nitrogen Control in Cyanobacteria. *Journal of Bacteriology*, *183*(2), 411–425. http://doi.org/10.1128/JB.183.2.411-425.2001

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews. Microbiology*, *8*(1), 15–25. http://doi.org/10.1038/nrmicro2259

Hirsch, P., Eckhardt, F. E. W., & Palmer Jr., R. J. (1995). Fungi active in weathering of rock and stone monuments. *Canadian Journal of Botany*, *73*(S1), 1384–1390. http://doi.org/10.1139/b95-401

Hopkins, D. W., Sparrow, A. D., Gregorich, E. G., Elberling, B., Novis, P., Fraser, F., … Greenfield, L. G. (2009). Isotopic evidence for the provenance and turnover of organic carbon by soil microorganisms in the Antarctic dry valleys. *Environmental Microbiology*, *11*(3), 597–608. http://doi.org/10.1111/j.1462-2920.2008.01830.x

Hopkins, D. W., Sparrow, A. D., Novis, P. M., Gregorich, E. G., Elberling, B., & Greenfield, L. G. (2006). Controls on the distribution of productivity and organic resources in Antarctic Dry Valley soils. *Proceedings. Biological Sciences / The Royal Society*, *273*(1602), 2687–95. http://doi.org/10.1098/rspb.2006.3595

Hopkins, D. W., Sparrow, A. D., Shillam, L. L., English, L. C., Dennis, P. G., Novis, P., … Greenfield, L. G. (2008). Enzymatic activities and microbial communities in an Antarctic dry valley soil: Responses to C and N supplementation. *Soil Biology and Biochemistry*, *40*(9), 2130–2136. http://doi.org/10.1016/j.soilbio.2008.03.022

Horowitz, N. H., Bauman, A. J., Cameron, R. E., Geiger, P. J., Hubbard, J. S., Shulman, G. P., … Westberg, K. (1969). Sterile soil from Antarctica: organic analysis. *Science (New York, N.Y.)*, *164*(883), 1054–1056.

Horowitz, N. H., Cameron, R. E., & Hubbard, J. S. (1972). Microbiology of the Dry Valleys of Antarctica. *Science (New York, N.Y.)*, *176*(4032), 242–245.

Houston, J., & Hartley, A. J. (2003). The central Andean west-slope rainshadow and its potential contribution to the origin of hyper-aridity in the Atacama Desert. *International Journal of Climatology*, *23*(12), 1453–1464. http://doi.org/10.1002/joc.938

Huang, C.-H., Chang, M.-T., & Huang, L. (2013). Use of novel species-specific PCR

primers targeted to DNA gyrase subunit B (gyrB) gene for species identification of the Cronobacter sakazakii and Cronobacter dublinensis. *Molecular and Cellular Probes*, *27*(1), 15–8. http://doi.org/10.1016/j.mcp.2012.08.004

Hulo, C., de Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., & Le Mercier, P. (2011). ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Research*, *39*(Database issue), D576-82. http://doi.org/10.1093/nar/gkq901

Islam, S. T., & Lam, J. S. (2014). Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. *Canadian Journal of Microbiology*, *60*(11), 697–716.

Jachlewski, S., Jachlewski, W. D., Linne, U., Bräsen, C., Wingender, J., & Siebers, B. (2015). Isolation of Extracellular Polymeric Substances from Biofilms of the Thermoacidophilic Archaeon Sulfolobus acidocaldarius. *Frontiers in Bioengineering and Biotechnology*, *3*, 123. http://doi.org/10.3389/fbioe.2015.00123

Jetten, M. S. M. (2008). The microbial nitrogen cycle. *Environmental Microbiology*, *10*(11), 2903–2909. http://doi.org/10.1111/j.1462-2920.2008.01786.x

Jetten, M. S. M., Niftrik, L. van, Strous, M., Kartal, B., Keltjens, J. T., & Op den Camp, H. J. M. (2009). Biochemistry and molecular biology of anammox bacteria. *Critical Reviews in Biochemistry and Molecular Biology*, *44*(2–3), 65–84. http://doi.org/10.1080/10409230902722783

Jittawuttipoka, T., Planchon, M., Spalla, O., Benzerara, K., Guyot, F., Cassier-Chauvat, C., & Chauvat, F. (2013). Multidisciplinary evidences that Synechocystis PCC6803 exopolysaccharides operate in cell sedimentation and protection against salt and metal stresses. *PloS One*, *8*(2), e55564. http://doi.org/10.1371/journal.pone.0055564

Jones, S. (2008). Microbial physiology: Five ways to cycle carbon. *Nature Reviews Microbiology*, *6*(2), 94–95. http://doi.org/10.1038/nrmicro1847

Kasai, H., Ezaki, T., & Harayama, S. (2000). Differentiation of phylogenetically related slowly growing mycobacteria by their gyrB sequences. *Journal of Clinical Microbiology*, *38*(1), 301–8.

Katoh, H., Furukawa, J., Tomita-Yokotani, K., & Nishi, Y. (2012). Isolation and purification of an axenic diazotrophic drought-tolerant cyanobacterium, Nostoc commune, from natural cyanobacterial crusts and its utilization for field research on soils polluted with radioisotopes. *Biochimica et Biophysica Acta*, *1817*(8), 1499–1505. http://doi.org/10.1016/j.bbabio.2012.02.039

Kerr, B., Riley, M. A., Feldman, M. W., & Bohannan, B. J. M. (2002). Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature*, *418*(6894), 171–4. http://doi.org/10.1038/nature00823

Keshari, N., & Adhikary, S. P. (2013). Characterization of cyanobacteria isolated from biofilms on stone monuments at Santiniketan, India. *Biofouling*, *29*(5), 525–36.

http://doi.org/10.1080/08927014.2013.794224

Khan, N., Tuffin, M., Stafford, W., Cary, C., Lacap, D., Pointing, S., & Cowan, D. (2011). Hypolithic microbial communities of quartz rocks from Miers Valley, McMurdo Dry Valleys, Antarctica. *Polar Biology*, *34*(11), 1657–1668. Journal Article. http://doi.org/10.1007/s00300-011-1061-7

Kier, G., Mutke, J., Dinerstein, E., Ricketts, T. H., Küper, W., Kreft, H., & Barthlott, W. (2005). Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography*, *32*(7), 1107–1116. http://doi.org/10.1111/j.1365-2699.2005.01272.x

Klock, J.-H., Wieland, A., Seifert, R., & Michaelis, W. (2007). Extracellular polymeric substances (EPS) from cyanobacterial mats: characterisation and isolation method optimisation. *Marine Biology*, *152*(5), 1077–1085. http://doi.org/10.1007/s00227-007-0754-5

Kondrashov, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings. Biological Sciences / The Royal Society*, *279*(1749), 5048–57. http://doi.org/10.1098/rspb.2012.1108

Koskella, B., & Brockhurst, M. A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews*, *38*(5), 916–31. http://doi.org/10.1111/1574-6976.12072

Lacap, D. C., Warren-Rhodes, K. A., McKay, C. P., & Pointing, S. B. (2011). Cyanobacteria and chloroflexi-dominated hypolithic colonization of quartz at the hyper-arid core of the Atacama Desert, Chile. *Extremophiles : Life under Extreme Conditions*, *15*(1), 31–8. http://doi.org/10.1007/s00792-010-0334-3

Laity, J. (2008). *Deserts and Desert Environments*. Chichester: Wiley-Blackwell.

Larrain, H., Velásquez, F., Cereceda, P., Espejo, R., Pinto, R., Osses, P., & Schemenauer, R. . (2002). Fog measurements at the site “Falda Verde” north of Chañaral compared with other fog stations of Chile. *Atmospheric Research*, *64*(1–4), 273–284. http://doi.org/10.1016/S0169-8095(02)00098-4

Laws, A., Gu, Y., & Marshall, V. (2001). Biosynthesis, characterisation, and design of bacterial exopolysaccharides from lactic acid bacteria. *Biotechnology Advances*, *19*(8), 597–625. http://doi.org/10.1016/S0734-9750(01)00084-2

Laybourn-Parry, J. (2009). Microbiology. No place too cold. *Science (New York, N.Y.)*, *324*(5934), 1521–2. http://doi.org/10.1126/science.1173645

Lee, C. K., Barbier, B. A., Bottos, E. M., McDonald, I. R., Cary, S. C., & Barbier, B. A. (2012). The inter-valley soil comparative survey: the ecology of Dry Valley edaphic microbial communities. *ISME J*, *6*(5), 1046–1057. Journal Article. http://doi.org/10.1038/ismej.2011.170

Lester, E. D., Satomi, M., & Ponce, A. (2007). Microflora of extreme arid Atacama Desert soils. *Soil Biology and Biochemistry*, *39*(2), 704–708. http://doi.org/10.1016/j.soilbio.2006.09.020

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: An ultra-fast

single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. http://doi.org/10.1093/bioinformatics/btv033

Lillesand, T. M., & Kiefer, R. W. (1994). *Remote sensing and image interpretation*. Wiley & Sons.

Looijesteijn, P. J., Trapet, L., de Vries, E., Abee, T., & Hugenholtz, J. (2001). Physiological function of exopolysaccharides produced by Lactococcus lactis. *International Journal of Food Microbiology*, *64*(1–2), 71–80. http://doi.org/10.1016/S0168-1605(00)00437-2

López-Bueno, A., Tamames, J., Velázquez, D., Moya, A., Quesada, A., & Alcamí, A. (2009). High diversity of the viral community from an Antarctic lake. *Science*, *326*(5954), 858–61. http://doi.org/10.1126/science.1179287

Lwoff, A., Horne, R., & Tournier, P. (1962). A System of Viruses. *Cold Spring Harbor Symposia on Quantitative Biology*, *27*(0), 51–55. http://doi.org/10.1101/SQB.1962.027.001.008

Maalcke, W. J., Reimann, J., de Vries, S., Butt, J. N., Dietl, A., Kip, N., … Kartal, B. (2016). Characterization of Anammox Hydrazine Dehydrogenase, a Key N2-producing Enzyme in the Global Nitrogen Cycle. *The Journal of Biological Chemistry*, *291*(33), 17077–92. http://doi.org/10.1074/jbc.M116.735530

MacClune, K. L., Fountain, A. G., Kargel, J. S., & MacAyeal, D. R. (2003). Glaciers of the McMurdo dry valleys: Terrestrial analog for Martian polar sublimation. *Journal of Geophysical Research*, *108*(E4), 5031. http://doi.org/10.1029/2002JE001878

MacLean, D., Jones, J. D. G., & Studholme, D. J. (2009). Application of "next-generation" sequencing technologies to microbial genetics. *Nature Reviews Microbiology*, *7*(2), 96–97. http://doi.org/10.1038/nrmicro2088

Magalhães, C. M., Machado, A., Frank-Fahle, B., Lee, C. K., & Cary, S. C. (2014). The ecological dichotomy of ammonia-oxidizing archaea and bacteria in the hyper-arid soils of the Antarctic Dry Valleys. *Frontiers in Microbiology*, *5*, 515. http://doi.org/10.3389/fmicb.2014.00515

Magalhães, C., Stevens, M. I., Cary, S. C., Ball, B. a, Storey, B. C., Wall, D. H., … Ruprecht, U. (2012). At limits of life: multidisciplinary insights reveal environmental constraints on biotic diversity in continental antarctica. *PloS One*, *7*(9), e44578. http://doi.org/10.1371/journal.pone.0044578

Mager, D. M., & Thomas, A. D. (2011). Extracellular polysaccharides from cyanobacterial soil crusts: A review of their role in dryland soil processes. *Journal of Arid Environments*, *75*(2), 91–97. http://doi.org/10.1016/j.jaridenv.2010.10.001

Makhalanyane, T. P., Pointing, S., & Cowan, D. A. (2014). *Antarctic Terrestrial Microbiology*. (D. A. Cowan, Ed.). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-45213-0

Makhalanyane, T. P., Valverde, A., Birkeland, N.-K., Cary, S. C., Marla Tuffin, I., & Cowan, D. A. (2013). Evidence for successional development in Antarctic

hypolithic bacterial communities. *The ISME Journal*, *7*(11), 2080–2090. http://doi.org/10.1038/ismej.2013.94

Makhalanyane, T. P., Valverde, A., Gunnigle, E., Frossard, A., Ramond, J.-B., & Cowan, D. A. (2015). Microbial ecology of hot desert edaphic systems. *FEMS Microbiology Reviews*, *39*(2), 203–21. http://doi.org/10.1093/femsre/fuu011

Makhalanyane, T. P., Valverde, A., Lacap, D. C., Pointing, S. B., Tuffin, M. I., & Cowan, D. A. (2013). Evidence of species recruitment and development of hot desert hypolithic communities. *Environmental Microbiology Reports*, *5*(2), 219–24. http://doi.org/10.1111/1758-2229.12003

Makhalanyane, T., Pointing, S., & Cowan, D. (2014). Lithobionts: Cryptic and Refuge Niches. In D. A. Cowan (Ed.), *Antarctic Terrestrial Microbiology SE - 9* (pp. 163–179). CHAP, Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-45213-0_9

Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., … Kyrpides, N. C. (2014). IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Research*, *42*(Database issue), D560-7. http://doi.org/10.1093/nar/gkt963

Matsumoto, G., Chikazawa, K., MURAYAMA, H., TORII, T., FUKUSHIMA, H., & HANYA, T. (1983). Distribution and correlation mercury in Antarctic dry valley of total organic carbon and soils, sediments and organisms. *Geochemical Journal*, *17*, 247–255.

Mazor, G., Kidron, G. J., Vonshak, A., & Abeliovich, A. (1996). The role of cyanobacterial exopolysaccharides in structuring desert microbial crusts. *FEMS Microbiology Ecology*, *21*(2), 121–130. http://doi.org/10.1111/j.1574-6941.1996.tb00339.x

McKay, C. P., & Friedmann, E. I. (1985). The cryptoendolithic microbial environment in the Antarctic cold desert: temperature variations in nature. *Polar Biology*, *4*, 19–25.

McKay, Friedmann, E. I., Gómez-Silva, B., Cáceres-Villanueva, L., Andersen, D. T., & Landheim, R. (2004). Temperature and Moisture Conditions for Life in the Extreme Arid Region of the Atacama Desert: Four Years of Observations Including the El Niño of 1997–1998.

McKenzie, R., Smale, D., Bodeker, G., & Claude, H. (2003). Ozone profile differences between Europe and New Zealand: Effects on surface UV irradiance and its estimation from satellite sensors. *Journal of Geophysical Research: Atmospheres*, *108*(D6), 4179. Journal Article. http://doi.org/10.1029/2002JD002770

McKnight, D. M., Niyogi, D. K., Alger, A. S., Bomblies, A., Conovitz, P. A., & Tate, C. M. (1999). Dry Valley streams in Antarctica: Ecosystems waiting for water. *BioScience*, *49*(12), 985–995.

McKnight, D. M., Runkel, R. L., Tate, C. M., Duff, J. H., & Moorhead, D. L. (2004). 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, *23*(2), 171–188. http://doi.org/10.1899/0887-3593(2004)023<0171:INAPDO>2.0.CO;2

Meiges, P. (1953). World Distribution of Arid and Semi-arid Homoclimates. In UNESCO (Ed.), *Arid Zone Hyreology* (pp. 203–209). Paris: United Nations.

Mohr, R., Voss, B., Schliep, M., Kurz, T., Maldener, I., Adams, D. G., … Hess, W. R. (2010). A new chlorophyll d-containing cyanobacterium: evidence for niche adaptation in the genus Acaryochloris. *The ISME Journal*, *4*(11), 1456–69. http://doi.org/10.1038/ismej.2010.67

Murrell, J. C., & Jetten, M. S. M. (2009). The microbial methane cycle. *Environmental Microbiology Reports*, *1*(5), 279–284. http://doi.org/10.1111/j.1758-2229.2009.00089.x

Nash, T. H. III, White, S. L. Marsh, J. E. (1977). Lichen and moss distribution and biomass in hot desert ecosystems. *The Bryologist*, *80*, 470–470.

Neiderberger, T. D., Sohm, J. A., Gunderson, T. E., Parker, A. E., Tirindelli, J. J., Capone, D. G., … Cary, S. C. (2015). Microbial community composition of transiently wetted Antarctic Dry Valley soils. *Frontiers in Microbiology*, *6*, 9. http://doi.org/10.3389/fmicb.2015.00009

Nelson, M. B., Martiny, A. C., & Martiny, J. B. H. (2016). Global biogeography of microbial nitrogen-cycling traits in soil. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 8033–40. http://doi.org/10.1073/pnas.1601070113

Ng, K. W., Pointing, S. B., & Dvornyk, V. (2013). Patterns of nucleotide diversity of the ldpA circadian gene in closely related species of cyanobacteria from extreme cold deserts. *Applied and Environmental Microbiology*, *79*(5), 1516–22. http://doi.org/10.1128/AEM.03439-12

Nicolaus, B., Panico, A., Lama, L., Romano, I., Manca, M. C., Giulio, A. De, & Gambacorta, A. (1999). Chemical composition and production of exopolysaccharides from representative members of heterocystous and non-heterocystous cyanobacteria. *Phytochemistry*, *4*(52), 639–647.

Niederberger, T. D., McDonald, I. R., Hacker, A. L., Soo, R. M., Barrett, J. E., Wall, D. H., & Cary, S. C. (2008). Microbial community composition in soils of Northern Victoria Land, Antarctica. *Environmental Microbiology*, *10*(7), 1713–24. http://doi.org/10.1111/j.1462-2920.2008.01593.x

Niederberger, T. D., Sohm, J. A., Tirindelli, J., Gunderson, T., Capone, D. G., Carpenter, E. J., & Cary, S. C. (2012). Diverse and highly active diazotrophic assemblages inhabit ephemerally wetted soils of the Antarctic Dry Valleys. *FEMS Microbiology Ecology*, *82*(2), 376–90. http://doi.org/10.1111/j.1574-6941.2012.01390.x

Nylen, T. H., Fountain, A. G., & Doran, P. T. (2004). Climatology of katabatic winds in the McMurdo dry valleys, southern Victoria Land, Antarctica. *Journal of Geophysical Research: Atmospheres*, *109*(D3), D03114. Journal Article. http://doi.org/10.1029/2003JD003937

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D.,

Powell, G. V. N., Underwood, E. C., … Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, *51*(11), 933. http://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Systems*, *4*, 2. http://doi.org/10.1186/1746-1448-4-2

Orlando, J., Carú, M., Pommerenke, B., & Braker, G. (2012). Diversity and activity of denitrifiers of chilean arid soil ecosystems. *Frontiers in Microbiology*, *3*, 101. http://doi.org/10.3389/fmicb.2012.00101

Pace, M. L., & Cole, J. J. (1994). Comparative and experimental approaches to top-down and bottom-up regulation of bacteria. *Microbial Ecology*, *28*(2), 181–93. http://doi.org/10.1007/BF00166807

Padan, E., & Shilo, M. (1973). Cyanophages-viruses attacking blue-green algae. *Bacteriological Reviews*, *37*(3), 343–70.

Palmer, R. J., & Friedmann, E. I. (1990). Water relations and photosynthesis in the cryptoendolithic microbial habitat of hot and cold deserts. *Microbial Ecology*, *19*(1), 111–8. http://doi.org/10.1007/BF02015057

Partensky, F., Hess, W. R., & Vaulot, D. (1999). Prochlorococcus, a marine photosynthetic prokaryote of global significance. *Microbiology and Molecular Biology Reviews : MMBR*, *63*(1), 106–27.

Paul, R., Jinkerson, R. E., Buss, K., Steel, J., Mohr, R., Hess, W. R., … Fromme, P. (2014). Draft Genome Sequence of the Filamentous Cyanobacterium Leptolyngbya sp. Strain Heron Island J, Exhibiting Chromatic Acclimation. *Genome Announcements*, *2*(1), e01166-13. http://doi.org/10.1128/genomeA.01166-13

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*(5), 1633–1644. http://doi.org/10.5194/hess-11-1633-2007

Peeters, K., & Willems, A. (2011). The gyrB gene is a useful phylogenetic marker for exploring the diversity of Flavobacterium strains isolated from terrestrial and aquatic habitats in Antarctica. *FEMS Microbiology Letters*, *321*(2), 130–140. http://doi.org/10.1111/j.1574-6968.2011.02326.x

Pereira, S., Mota, R., Vieira, C. P., Vieira, J., & Tamagnini, P. (2015). Phylum-wide analysis of genes/proteins related to the last steps of assembly and export of extracellular polymeric substances (EPS) in cyanobacteria. *Scientific Reports*, *5*, 14835. http://doi.org/10.1038/srep14835

Pereira, S., Zille, A., Micheletti, E., Moradas-Ferreira, P., De Philippis, R., & Tamagnini, P. (2009). Complexity of cyanobacterial exopolysaccharides: composition, structures, inducing factors and putative genes involved in their biosynthesis and assembly. *FEMS Microbiology Reviews*, *33*(5), 917–41. http://doi.org/10.1111/j.1574-6976.2009.00183.x

Pointing, S. B., & Belnap, J. (2012). Microbial colonization and controls in dryland systems. *Nature Reviews Microbiology*, *10*(9), 654–654.

http://doi.org/10.1038/nrmicro2854

Pointing, S. B., Bollard-Breen, B., & Gillman, L. N. (2014). Diverse cryptic refuges for life during glaciation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(15), 5452–3. http://doi.org/10.1073/pnas.1403594111

Pointing, S. B., Burkhard Büdel, Convey, P., Gillman, L. N. L., Körner, C., Leuzinger, S. S., … Vincent, W. W. F. (2015). Biogeography of photoautotrophs in the high polar biome. *Frontiers in Plant Science*, *6*, 692. http://doi.org/10.3389/fpls.2015.00692

Pointing, S. B., Chan, Y., Lacap, D. C., Lau, M. C. Y., Jurgens, J. A., & Farrell, R. L. (2009). Highly specialized microbial diversity in hyper-arid polar desert. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(47), 19964–9. http://doi.org/10.1073/pnas.0908274106

Pointing, S. B., Warren-Rhodes, K. A., Lacap, D. C., Rhodes, K. L., & McKay, C. P. (2007). Hypolithic community shifts occur as a result of liquid water availability along environmental gradients in China's hot and cold hyperarid deserts. *Environmental Microbiology*, *9*(2), 414–424. http://doi.org/10.1111/j.1462-2920.2006.01153.x

Potts, M. (1994). Desiccation tolerance of prokaryotes. *Microbiological Reviews*, *58*(4), 755–805.

Prentice, M. L., Kleman, J. L., & Stroeven, A. P. (2013). The Composite Glacial Erosional Landscape of the Northern Mcmurdo Dry Valleys: Implications for Antarctic Tertiary Glacial History (pp. 1–38). American Geophysical Union. http://doi.org/10.1029/AR072p0001

Rao, S., Chan, Y., Lacap, D. C., Hyde, K. D., Pointing, S. B., & Farrell, R. L. (2011). Low-diversity fungal assemblage in an Antarctic Dry Valleys soil. *Polar Biology*, *35*(4), 567–574. http://doi.org/10.1007/s00300-011-1102-2

Reeves, P. R., Hobbs, M., Valvano, M. A., Skurnik, M., Whitfield, C., Coplin, D., … Rick, P. D. (1996). Bacterial polysaccharide synthesis and gene nomenclature. *Trends in Microbiology*, *4*(12), 495–503. http://doi.org/10.1016/S0966-842X(97)82912-5

Richter, I., Herbold, C. W., Lee, C. K., McDonald, I. R., Barrett, J. E., & Cary, S. C. (2014). Influence of soil properties on archaeal diversity and distribution in the McMurdo Dry Valleys, Antarctica. *FEMS Microbiology Ecology*, *89*(2), 347–59. http://doi.org/10.1111/1574-6941.12322

Risacher, F., Alonso, H., Salazar, C., McKay, C. P., Friedmann, E. I., Gómez-Silva, B., … Landheim, R. (2003). The origin of brines and salts in Chilean salars: a hydrochemical review. *Earth-Science Reviews*, *3*(3–4), 393–406. http://doi.org/10.1089/153110703769016460

Roberts, I. S. (1996). The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annual Review of Microbiology*, *50*, 285–315. http://doi.org/10.1146/annurev.micro.50.1.285

Roh, S. W., Abell, G. C., Kim, K. H., Nam, Y. D., & Bae, J. W. (2010). Comparing

microarrays and next-generation sequencing technologies for microbial ecology research. *Trends Biotechnol*, *28*(6), 291–299. Journal Article. http://doi.org/10.1016/j.tibtech.2010.03.001

Ronca, S., Ramond, J.-B., Jones, B. E., Seely, M., & Cowan, D. A. (2015). Namib Desert dune/interdune transects exhibit habitat-specific edaphic bacterial communities. *Frontiers in Microbiology*, *6*, 845. http://doi.org/10.3389/fmicb.2015.00845

Roossinck, M. J. (2011). The good viruses: viral mutualistic symbioses. *Nature Reviews. Microbiology*, *9*(2), 99–108. http://doi.org/10.1038/nrmicro2491

Safriel, U., Adeel, Z., Niemeijer, D., Puigdefabregas, J., White, R., Lal, R., … Mcnab, D. (2005). Dryland Systems. *Ecosystems and Human Well-Being: Current State and Trends: Findings of the Condition and Trends Working Group*.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (2ND ed., pp. E3–E4). Cold Spring Harbor Laboratory Press.

Satoh, K., Hirai, M., Nishio, J., Yamaji, T., Kashino, Y., & Koike, H. (2002). Recovery of photosynthetic systems during rewetting is quite rapid in a terrestrial cyanobacterium, Nostoc commune. *Plant and Cell Physiology*, *43*(2), 170–176. http://doi.org/10.1093/pcp/pcf020

SCAR, & Scientific Committee on Antarctic Research. (2004). SCAR Bulletin 155. *Polar Record*, *40*, 371–382.

Schimel, J. P., & Schaeffer, S. M. (2012). Microbial control over carbon cycling in soil. *Frontiers in Microbiology*, *3*, 348. http://doi.org/10.3389/fmicb.2012.00348

Schlesinger, W. H., Pippen, J. S., Wallenstein, M. D., Hofmockel, K. S., Klepeis, D. M., & Mahall, B. E. (2003). Community composition and photosynthesis by photoautotrophs under quartz pebbles, Southern Mojave Desert. *Ecology*, *84*(12), 3222–3231. http://doi.org/10.1890/02-0549

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., … Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–41. http://doi.org/10.1128/AEM.01541-09

Schmidt, H., & Karnieli, A. (2000). Remote sensing of the seasonal variability of vegetation in a semi-arid environment. *Journal of Arid Environments*, *45*(1), 43–59. http://doi.org/10.1006/jare.1999.0607

Seckbach, J., & Oren, A. (2007). Oxygenic photosynthetic microorganisms in extreme environments. In J. Seckbach (Ed.), *Algae and Cyanobacteria in Extreme Environments* (pp. 3–25). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-1-4020-6112-7

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*, *30*(14), 2068–9. http://doi.org/10.1093/bioinformatics/btu153

Sharma, S., Szele, Z., Schilling, R., Munch, J. C., & Schloter, M. (2006). Influence of freeze-thaw stress on the structure and function of microbial communities and denitrifying populations in soil. *Applied and Environmental Microbiology*, *72*(3), 2148–54. http://doi.org/10.1128/AEM.72.3.2148-2154.2006

Shearer, C. A. (1995). Fungal competition. *Canadian Journal of Botany*, *73*(S1), 1259–1264. http://doi.org/10.1139/b95-386

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, *26*(10), 1135–1145. http://doi.org/10.1038/nbt1486

Shima, S., Warkentin, E., Thauer, R. K., & Ermler, U. (2002). Structure and function of enzymes involved in the methanogenic pathway utilizing carbon dioxide and molecular hydrogen. *Journal of Bioscience and Bioengineering*, *93*(6), 519–530. http://doi.org/10.1016/S1389-1723(02)80232-8

Shively, J. M., van Keulen, G., & Meijer, W. G. (1998). Something from almost nothing: carbon dioxide fixation in chemoautotrophs. *Annual Review of Microbiology*, *52*, 191–230. http://doi.org/10.1146/annurev.micro.52.1.191

Siebert, J., Hirsch, P., Hoffmann, B., Gliesche, C. G. C. G., Peissl, K., Jendrach, M., … Peissl, K. (1996). Cryptoendolithic microorganisms from Antarctic sandstone of Linnaeus Terrace (Asgard Range): diversity, properties and interactions. *Biodiversity & Conservation*, *5*(11), 1337–1363. Journal Article. http://doi.org/10.1007/BF00051982

Sifri, C. D. (2008). Quorum sensing: bacteria talk sense. *Clinical Infectious Diseases*, *47*(8), 1070–6. http://doi.org/10.1086/592072

Sinha Häder D.P., R. P., & Photochem Photobiol, S. (2002). UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci*, *1*, 225–236. Journal Article.

Slade, D., & Radman, M. (2011). Oxidative stress resistance in Deinococcus radiodurans. *Microbiology and Molecular Biology Reviews : MMBR*, *75*(1), 133–91. http://doi.org/10.1128/MMBR.00015-10

Sonthiphand, P., Hall, M. W., & Neufeld, J. D. (2014). Biogeography of anaerobic ammonia-oxidizing (anammox) bacteria. *Frontiers in Microbiology*, *5*, 399. http://doi.org/10.3389/fmicb.2014.00399

Staats, N., De Winder, B., Stal, L., & Mur, L. (1999). Isolation and characterization of extracellular polysaccharides from the epipelic diatoms Cylindrotheca closterium and Navicula salinarum. *European Journal of Phycology*, *34*(2), 161–169. http://doi.org/10.1080/09670269910001736212

Stahl, D. A., & de la Torre, J. R. (2012). Physiology and diversity of ammonia-oxidizing archaea. *Annual Review of Microbiology*, *66*, 83–101. http://doi.org/10.1146/annurev-micro-092611-150128

Stanier, R. Y., & Cohen-Bazire, G. (1977). Phototrophic prokaryotes: the cyanobacteria. *Annual Review of Microbiology*, *31*, 225–74. http://doi.org/10.1146/annurev.mi.31.100177.001301

Stanier, R. Y., Deruelles, J., Rippka, R., Herdman, M., & Waterbury, J. B. (1979).

Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *Microbiology*, *111*(1), 1–61. http://doi.org/10.1099/00221287-111-1-1

States, D. J., & Gish, W. (1994). QGB: Combined Use of Sequence Similarity and Codon Bias for Coding Region Identification. *Journal of Computational Biology*, *1*(1), 39–50.

Steindler, L., & Venturi, V. (2007). Detection of quorum-sensing N-acyl homoserine lactone signal molecules by bacterial biosensors. *FEMS Microbiology Letters*, *266*(1), 1–9. http://doi.org/10.1111/j.1574-6968.2006.00501.x

Steven, B., Gallegos-Graves, L. V., Belnap, J., & Kuske, C. R. (2013). Dryland soil microbial communities display spatial biogeographic patterns associated with soil depth and soil parent material. *FEMS Microbiology Ecology*, *86*(1), 101–13. http://doi.org/10.1111/1574-6941.12143

Stomeo, F., Makhalanyane, T. P., Valverde, A., Pointing, S. B., Stevens, M. I., Cary, C. S., … Cowan, D. A. (2012). Abiotic factors influence microbial diversity in permanently cold soil horizons of a maritime-associated Antarctic Dry Valley. *FEMS Microbiology Ecology*, *82*(2), 326–40. http://doi.org/10.1111/j.1574-6941.2012.01360.x

Stomeo, F., Valverde, A., Pointing, S. B., McKay, C. P., Warren-Rhodes, K. a., Tuffin, M. I., … Cowan, D. a. (2013). Hypolithic and soil microbial community assembly along an aridity gradient in the Namib Desert. *Extremophiles*, *17*(2), 329–337. http://doi.org/10.1007/s00792-013-0519-7

Stoughton, R. B. (2005). Applications of DNA microarrays in biology. *Annual Review of Biochemistry*, *74*, 53–82. http://doi.org/10.1146/annurev.biochem.74.082803.133212

Strauss, S. L., Day, T. A., & Garcia-Pichel, F. (2011). Nitrogen cycling in desert biological soil crusts across biogeographic regions in the Southwestern United States. *Biogeochemistry*, *108*(1–3), 171–182. http://doi.org/10.1007/s10533-011-9587-x

Sullivan, M. B., Waterbury, J. B., & Chisholm, S. W. (2003). Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. *Nature*, *424*(6952), 1047–51. http://doi.org/10.1038/nature01929

Suzuki, M., & Giovannoni, S. (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl. Envir. Microbiol.*, *62*(2), 625–630.

Svoma, B. M., DeBiasse, K., Nolte, J., Busby, B., Beeson, C., & Cerveny, R. (2011). Evaluation of the Environmental Protection Agency/National Weather Service Ultraviolet Index Forecast against independent UV measurements: Phoenix Arizona (2000-2006). *International Journal of Climatology*, *31*(9), 1368–1376. http://doi.org/10.1002/joc.2152

Takacs-Vesbach, C., Zeglin, L., Barrett, J. E., Gooseff, M. N., & Priscu, J. C. (2010). Factors promoting microbial diversity in the McMurdo Dry Valleys, Antarctica. In P. T. Doran, W. B. Lyons, & D. M. McKnight (Eds.), *Life in Antarctic Deserts*

and other Cold Dry Environments : Astrobiological Analogues (pp. 221–257). Cambridge: Cambridge University Press.

Tamaru, Y., Takani, Y., Yoshida, T., & Sakamoto, T. (2005). Crucial role of extracellular polysaccharides in desiccation and freezing tolerance in the terrestrial cyanobacterium Nostoc commune. *Applied and Environmental Microbiology*, *71*(11), 7327–33. http://doi.org/10.1128/AEM.71.11.7327-7333.2005

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–9. http://doi.org/10.1093/molbev/mst197

Tan, C. H., Koh, K. S., Xie, C., Zhang, J., Tan, X. H., Lee, G. P., … Kjelleberg, S. (2015). Community quorum sensing signalling and quenching: microbial granular biofilm assembly. *Npj Biofilms and Microbiomes*, *1*, 15006. http://doi.org/10.1038/npjbiofilms.2015.6

Thomas, D. S. G. (2011). Arid zones: their nature and extent. In D. S. G. Thomas (Ed.), *Arid Zone Geomorphology: Process, Form and Change in Drylands* (pp. 3–16). Chichester, UK: John Wiley & Sons, Ltd. http://doi.org/10.1002/9780470710777

Tomitani, A., Knoll, A. H., Cavanaugh, C. M., & Ohno, T. (2006). The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(14), 5442–7. http://doi.org/10.1073/pnas.0600999103

Tracy, C. R., Streten-Joyce, C., Dalton, R., Nussear, K. E., Gibb, K. S., & Christian, K. A. (2010). Microclimate and limits to photosynthesis in a diverse community of hypolithic cyanobacteria in northern Australia. *Environmental Microbiology*, *12*(3), 592–607. http://doi.org/10.1111/j.1462-2920.2009.02098.x

Tu, Q., Yu, H., He, Z., Deng, Y., Wu, L., Van Nostrand, J. D., … Zhou, J. (2014). GeoChip 4: a functional gene-array-based high-throughput environmental technology for microbial community analysis. *Molecular Ecology Resources*. http://doi.org/10.1111/1755-0998.12239

UNEP. (1992). *World atlas of desertification*. London: Edward Arnold.

Valle, A., Bailey, M. J., Whiteley, A. S., & Manefield, M. (2004). N-acyl-l-homoserine lactones (AHLs) affect microbial community composition and function in activated sludge. *Environmental Microbiology*, *6*(4), 424–433. http://doi.org/10.1111/j.1462-2920.2004.00581.x

Vázquez-Martínez, G., Rodriguez, M. H., Hernández-Hernández, F., & Ibarra, J. E. (2004). Strategy to obtain axenic cultures from field-collected samples of the cyanobacterium Phormidium animalis. *Journal of Microbiological Methods*, *57*(1), 115–21. http://doi.org/10.1016/j.mimet.2003.12.003

Vieites, J. M., Guazzaroni, M.-E., Beloqui, A., Golyshin, P. N., & Ferrer, M. (2009). Metagenomics approaches in systems microbiology. *FEMS Microbiology Reviews*, *33*(1), 236–255.

Vishniac, H. S. (1993). The Microbiology of Antarctic Soils. In E. I. Friedmann & A. B. Thistle (Eds.), *Antarctica Microbiology* (pp. 297–341). New York: Wiley-Liss.

Wall, D. H., & Virginia, R. A. (1999). Controls on soil biodiversity: insights from extreme environments. *Applied Soil Ecology*, *13*(2), 137–150. http://doi.org/10.1016/S0929-1393(99)00029-3

Wang, L.-T., Lee, F.-L., Tai, C.-J., & Kasai, H. (2007). Comparison of gyrB gene sequences, 16S rRNA gene sequences and DNA-DNA hybridization in the Bacillus subtilis group. *International Journal of Systematic and Evolutionary Microbiology*, *57*(Pt 8), 1846–50. http://doi.org/10.1099/ijs.0.64685-0

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–7. http://doi.org/10.1128/AEM.00062-07

Wang, X., Yang, F., & von Bodman, S. B. (2012). The genetic and structural basis of two distinct terminal side branch residues in stewartan and amylovoran exopolysaccharides and their potential role in host adaptation. *Molecular Microbiology*, *83*(1), 195–207. http://doi.org/10.1111/j.1365-2958.2011.07926.x

Ward, B. B., & Jensen, M. M. (2014). The microbial nitrogen cycle. *Frontiers in Microbiology*, *5*, 553. http://doi.org/10.3389/fmicb.2014.00553

Warren-Rhodes, K. A., Rhodes, K. L., Boyle, L. N., Pointing, S. B., Chen, Y., Liu, S., … McKay, C. P. (2007). Cyanobacterial ecology across environmental gradients and spatial scales in China's hot and cold deserts. *FEMS Microbiol Ecol*, *61*(3), 470–482. http://doi.org/10.1111/j.1574-6941.2007.00351.x

Warren-Rhodes, K. A., Rhodes, K. L., Pointing, S. B., Ewing, S. A., Lacap, D. C., Gomez-Silva, B., … McKay, C. P. (2006). Hypolithic cyanobacteria, dry limit of photosynthesis, and microbial ecology in the hyperarid Atacama Desert. *Microbial Ecology*, *52*(3), 389–398. http://doi.org/10.1007/s00248-006-9055-7

Weber, B., Wessels, D. C., Deutschewitz, K., Dojani, S., Reichenberger, H., & Büdel, B. (2013). Ecological characterization of soil-inhabiting and hypolithic soil crusts within the Knersvlakte, South Africa. *Ecological Processes*, *2*(1), 8. http://doi.org/10.1186/2192-1709-2-8

Welman, A. D., & Maddox, I. S. (2003). Exopolysaccharides from lactic acid bacteria: perspectives and challenges. *Trends in Biotechnology*, *21*(6), 269–74. http://doi.org/10.1016/S0167-7799(03)00107-0

Wessels, K. ., Prince, S. ., Frost, P. ., & van Zyl, D. (2004). Assessing the effects of human-induced land degradation in the former homelands of northern South Africa with a 1 km AVHRR NDVI time-series. *Remote Sensing of Environment*, *91*(1), 47–67. http://doi.org/10.1016/j.rse.2004.02.005

West, N. E. (1990). Structure and function of microphytic soil crusts in wildland ecosystems of arid to semi-arid regions. *Advances in Ecological Research*, *20*, 179–223. http://doi.org/10.1016/S0065-

2504(08)60055-0

Whitfield, C. (2006). Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. *Annual Review of Biochemistry*, *75*, 39–68. http://doi.org/10.1146/annurev.biochem.75.103004.142545

Whitney, J. C., & Howell, P. L. (2013). Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. *Trends in Microbiology*, *21*(2), 63–72. http://doi.org/10.1016/j.tim.2012.10.001

Whitton, B. A., & Potts, M. (2012). Their diversity in time and space. In B. A. Whitton (Ed.), *Ecology of Cyanobacteria* (pp. 1–13). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-94-007-3855-3

Wierzchos, J., de los Ríos, A., & Ascaso, C. (2012). Microorganisms in desert rocks: the edge of life on Earth. *International Microbiology : The Official Journal of the Spanish Society for Microbiology*, *15*(4), 173–83. http://doi.org/10.2436/20.1501.01.170

Williams, P., Winzer, K., Chan, W. C., & Cámara, M. (2007). Look who's talking: communication and quorum sensing in the bacterial world. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1483), 1119–34. http://doi.org/10.1098/rstb.2007.2039

Willis, L. M., & Whitfield, C. (2013). Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways. *Carbohydrate Research*, *378*, 35–44. http://doi.org/10.1016/j.carres.2013.05.007

Witherow, R. A., Lyons, W. B., Bertler, N. A. N., Welch, K. A., Mayewski, P. A., Sneed, S. B., … Fountain, A. (2006). The aeolian flux of calcium, chloride and nitrate to the McMurdo Dry Valleys landscape: evidence from snow pit analysis. *Antarctic Science*, *18*(4), 497–505. Journal Article.

Wong, F. K. Y., Lacap, D. C., Lau, M. C. Y., Aitchison, J. C., Cowan, D. A., & Pointing, S. B. (2010). Hypolithic microbial community of quartz pavement in the high-altitude tundra of central Tibet. *Microbial Ecology*, *60*(4), 730–739. http://doi.org/10.1007/s00248-010-9653-2

Wong, F. K. Y., Lau, M. C. Y., Lacap, D. C., Aitchison, J. C., Cowan, D. A., & Pointing, S. B. (2009). Endolithic Microbial Colonization of Limestone in a High-altitude Arid Environment. *Microbial Ecology*, *59*(4), 689–699. http://doi.org/10.1007/s00248-009-9607-8

Wood, S. A., Rueckert, A., Cowan, D. A., & Cary, S. C. (2008). Sources of edaphic cyanobacterial diversity in the Dry Valleys of Eastern Antarctica. *The ISME Journal*, *2*(3), 308–320. http://doi.org/10.1038/ismej.2007.104

Woodward, F. I., Lomas, M. R., & Kelly, C. K. (2004). Global climate and the distribution of plant biomes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1450), 1465–76. http://doi.org/10.1098/rstb.2004.1525

World Meteorological Organization. (2011). *Scientific Assessment of Ozone Depletion 2010: Global Ozone Research and Monitoring Project-Report No.*

52. Geneva, Switzerland.

Wright, D. J., Smith, S. C., Joardar, V., Scherer, S., Jervis, J., Warren, A., … Potts, M. (2005). UV irradiation and desiccation modulate the three-dimensional extracellular matrix of Nostoc commune (Cyanobacteria). *The Journal of Biological Chemistry*, *280*(48), 40271–81. http://doi.org/10.1074/jbc.M505961200

Wynn-Williams, D. D. (1990). Ecological Aspects of Antarctic Microbiology (pp. 71–146). Springer US. http://doi.org/10.1007/978-1-4684-7612-5_3

Wynn-Williams, D. D. (2002). Cyanobacteria in Deserts – Life at the Limit? In B. A. Whitton & M. Potts (Eds.), *The Ecology of Cyanobacteria* (pp. 341–366). Dordrecht: Kluwer Academic Publishers. http://doi.org/10.1007/0-306-46855-7

Xia, H., Li, T., Deng, F., & Hu, Z. (2013). Freshwater cyanophages. *Virologica Sinica*, *28*(5), 253–9. http://doi.org/10.1007/s12250-013-3370-1

Xu, X., Khudyakov, I., & Wolk, C. (1997). Lipopolysaccharide dependence of cyanophage sensitivity and aerobic nitrogen fixation in Anabaena sp. strain PCC 7120. *J. Bacteriol.*, *179*(9), 2884–2891.

Yamamoto, S., & Harayama, S. (1995). PCR amplification and direct sequencing of gyrB genes with universal primers and their application to the detection and taxonomic analysis of Pseudomonas putida strains. *Applied and Environmental Microbiology*, *61*(3), 1104–9.

Yamamoto, S., & Harayama, S. (1996). Phylogenetic analysis of Acinetobacter strains based on the nucleotide sequences of gyrB genes and on the amino acid sequences of their products. *International Journal of Systematic Bacteriology*, *46*(2), 506–11. http://doi.org/10.1099/00207713-46-2-506

Yamamoto, S., & Harayama, S. (1998). Phylogenetic relationships of Pseudomonas putida strains deduced from the nucleotide sequences of gyrB, rpoD and 16S rRNA genes. *International Journal of Systematic Bacteriology*, *48 Pt 3*(3), 813–9. http://doi.org/10.1099/00207713-48-3-813

Yergeau, E., Kang, S., He, Z., Zhou, J., & Kowalchuk, G. a. (2007). Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. *The ISME Journal*, *1*(2), 163–79. http://doi.org/10.1038/ismej.2007.24

Yergeau, E., & Kowalchuk, G. a. (2008). Responses of Antarctic soil microbial communities and associated functions to temperature and freeze-thaw cycle frequency. *Environmental Microbiology*, *10*(9), 2223–35. http://doi.org/10.1111/j.1462-2920.2008.01644.x

Yung, C. C. M., Chan, Y., Lacap, D. C., Pérez-Ortega, S., de Los Rios-Murillo, A., Lee, C. K., … Pointing, S. B. (2014). Characterization of Chasmoendolithic Community in Miers Valley, McMurdo Dry Valleys, Antarctica. *Microbial Ecology*. http://doi.org/10.1007/s00248-014-0412-7

Yura, T., Kanemori, M., & Morita, M. T. (2011). Bacterial Stress Responses. In G. T. Storz & R. Henegge (Eds.), *Bacterial Stress Responses* (pp. 3–18). book,

Washington, D.C.: Amer Society for Microbiology.

Zablocki, O., van Zyl, L., Adriaenssens, E. M., Rubagotti, E., Tuffin, M., Cary, C., & Cowan, D. (2014). High diversity of tailed phages, eukaryotic viruses and virophage-like elements in the metaviromes of Antarctic soils. *Applied and Environmental Microbiology*, (August), 1–29. http://doi.org/10.1128/AEM.01525-14

Zhou, J., Kang, S., Schadt, C. W., & Garten, C. T. (2008). Spatial scaling of functional gene diversity across various microbial taxa. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(22), 7768–73. http://doi.org/10.1073/pnas.0709016105

Ziolkowski, L. A., Mykytczuk, N. C. S., Omelon, C. R., Johnson, H., Whyte, L. G., & Slater, G. F. (2013). Arctic gypsum endoliths: a biogeochemical characterization of a viable and active microbial community. *Biogeosciences*, *10*(11), 7661–7675. http://doi.org/10.5194/bg-10-7661-2013

Appendix 1

The sequences of putative Wzb homologues

```
THF_05431       ----------------------MNHLVEQNHLQSQIECDSAGTGSYHIGNPPDRRMAAA-    37
RQEc_05872      MPYKLLFVCLGNICRSPAAENIMNHLIEQNNLSEHIICDSAGTAGYHIGSPPDRRMATAA    60
LS2_01563       MPYKLLFVCLGNICRSPAAENIMNHLIEQNNLSEHIICDSAGTAGYHIGSQPDKRMATAA    60
                                ****:***.*..:* ******..****. **:***:*

THF_05431       AMRRGIQLLGRARQFQKIDFEAFDLILAMDQANYQDILALDPQGKYWHKVKLMCDFCSTH    97
RQEc_05872      ASKLGIKLHGQARQFKKSDFENFDLILAMDRENYQEILFLDPAKQYRDKVRLMCDFCSQH    120
LS2_01563       ASKLGIKLHGQARQFKKSDFENFDLILAMDRENYQDILFLDPAKQYRDKVRLMCDFCSQH    120
                * : **:* *:****:* *** ********: ***:** ***  :* .**:******* *

THF_05431       NDQEVPDPYYGGSDGFNYVIDLLLDACDGLMRYVKEQKQVI-----  138
RQEc_05872      TIKDVPDPYYGGSEGFNRVIDLLLDSCNGLLQNVATNQHLAVSDSP  166
LS2_01563       TTKDVPDPYYGGSEGFNRVIDLLLDSCNGLLQNVATNQHLAVSDSP  166
                . ::**********:*** *******:*:**:: *  ::::
```

Appendix 2

The sequences of putative Wza homologues

```
THF_02157       -----------MISQRFWIYL--PIQSTLLLTVLTLTTIHPAIAA-----------VLSS    36
RQEc_02740      MTRNYYISNSNMVSTGFSKSLTQPVAGLTLLAMMAVAFPSPTRAQLPPNRVTPAPATAPT    60
LS2_04994       -----------MVSTGFSKSLTQPVAGLTLLAMMAVTFPSPTRAQLPPNRVTPAPTTAPT    49
                          *:*  *   *  *: .  **:::::::   *: *           .  :

THF_02157       NQPPL----SNQPLLIAAVTRGDPYTLGAGDRVRIDIFRAPQYSGETQVLADGTLNLPLA    92
RQEc_02740      APPPTAPAPAAAPAAQQAIQVPIDYALGGGDRIRINVFEVPEYTGDYQIPPGGELYLPLI    120
LS2_04994       APPPTAPAPAAAPAAQQAIQVPIDYALGGGDRIRINVFEVPEYTGDYQIPPGGELYLPLI    109
                  **     :  *    *:     *:**.***:**::*..*:*:*: *:   * * ***

THF_02157       GSISVDGMTVDQATAAIAAGYSRFLRRPLITLSLLSRRPLEIGVAGEVSRPGAYSVNVEG    152
RQEc_02740      GGVTVLGLTQAEAAEVIAAKYSRFLKRPLITVSLLAPRPINVVVSGEVTRPGSYTVGLQG    180
LS2_04994       GGVTVLGLTQSEAAEVIAAKYSRFLKRPLITVSLLAPRPINVVVSGEVTRPGSYTVGLQG    169
                *.::* *:*  :*: .*** *****:*****:***: **::: *:***:***:*:* ::*

THF_02157       G-------EFPTINQLLQTAGGTTLSANVRQVEIRRPQ-RDGTDQVINVDLMQLLQTGDL    204
RQEc_02740      GAGDNPGVQYPTIIGALTLAEGVTLAADLRQVQLRRREGRGGPERVTTLDLKQLVTSGSS    240
LS2_04994       GAGDNPGVQYPTIIGALTLAEGVTLAADLRQVQLRRREGRGGPERVTTLDLKQLVTSGSS    229
                *       ::***   *  * *.**:*::***::** : * * ::* .:** **: :*.

THF_02157       KYDIALKDGDTIFVPATETLNVAESGLLSDASFAADNAQPINISVAGEVFRPGPYTVTGS    264
RQEc_02740      PRDITLRDGDTIFVPTATSVNLADIRQFSTASFAAAQNRPRTVTVVGEVNRPGSYVVIGG    300
LS2_04994       PRDITLRDGDTIFVPTATSVNMAEIRQFSTASFATAQNRPRTVTVVGEVNRPGSYVVIGG    289
                  **:*:********:: ::*:*:   :* ****: : :* .::*.*** *** *.* *.

THF_02157       ARTGDAG--------V-PGSGSGGGLPTVTRAIQVAGGIKPTANIREILVRRPTRSGTSQ    315
RQEc_02740      STAAAAPASLQGSASIQAAGQAGGGLPTVSRAIQLAGGITSSADVRNIQVRRKTNNGAEQ    360
LS2_04994       STAAAAPASLQGSGSIQAAGQAGGGLPTVSRAIQLAGGITSTADVRNIQVRRKTNNGAEQ    349
                : :. *         :  .. :*******:****:****. :*::*:* *** *..*:.*

THF_02157       EFKVNLWDLLKSGDANQDPILQEGDAIVVTQIAALNPAEAAALASTSFSPDSIKINVVGE    375
RQEc_02740      RMSLNFWQLLQ-GNPNQDTLVQDGDTIIIPTATDINPAEATTLADASFSPATIQVSVVGE    419
LS2_04994       RMSLNFWQLLQ-GDPNQDTLVQDGDTIIIPTATDINPAESTTLADASFSPATIQVSVVGE    408
                .:.:*:*:**: *: *** ::*:**:*::   : :****:::**.:**** :*::.****

THF_02157       VAQPGTFELPPNTPLNQAILAAGGFNN-RARRASVNLVRLNEDGSVSREEIPVDFTQGIA    434
RQEc_02740      VKNPGLVNLQPNTPLNQAILNAGGFNNSRAKRSTVELVRLNPDGTVTKRSVPVDLSENIN    479
LS2_04994       VRSPGLVNLQPNTPLNQAILNAGGFNNSRAKRSTVELVRLNPDGTVTKRSVPVDLSENIN    468
                * .** .:* ********** ****** **:*::*:***** **:*::..:***::: *

THF_02157       DGKNPSLRNNDVVVVGRSGLATTSDTLGQVLAPL----GNVFSIFQFPFNFLRIFQ       486
RQEc_02740      AKSNPILRQNDIVVVGRSGTARIGDALGSVFGPIIAPALGLFNIFR----------       525
LS2_04994       AKTNPILRQNDIVVVGRSGTARIGDALGSVFGPIIAPALGLFNIFR----------       514
                  .** **:**:******* *  .*:**.*:.*:      :*.**:
```

Appendix 3

The sequence of putative Wzc homologues

```
THF_06826       ------------------------------------------------------------M    1
THF_07722       -----------------------------------------------------MRANSEE    7
RQE_00387       ------------------------------------------------------MQSTE    5
LS2_00769       -------------------------------------------------------MEN    3
RQEc_03216      ---------------------------------------------MVRV---TESNLNA--T   12
LS2_05641       ---------------------------------------------MVRV---TESNLNA--T   12
RQEc_04624      MKREQNVQALTLSKNGKQTDQLTPVDTAQSDKLKKKEFNLRPDWTSIQPSIETDEFNEVD   60
LS2_01047       MKHEQNVQALTLRKNGKQTDQLTPVDTAQFDELKERESNLRRDWANIQPSVETDEFDEVD   60
THF_08171       ------MKNTYKPTVESNNH---------------SSPSALVPFQPVQPPG--------E   31
RQEc_07349      MEDQQRTQQVFFEKNGNGKH---------------SQPSLPIYAVPFQENT--------Q   37

THF_06826       NTSYPSFNTPLIIFKRNWLLAICVFVPVFLTVLALGSLQKSTYTAESSLRLVKRNSTAQL   61
THF_07722       TQDSIDLLSYWSILRRRWIPLTGVAASVFGIVVLSTILQKPAFEAVGKLLFSKNDRLSSL   67
RQE_00387       STEYIDLQQYWLILKRRLLPASAVLGFVLLLTALVTFWQKPIYQAQGKLLLKRS-DTSSL   64
LS2_00769       QESSIAFHQYGLILKRQWLPASVVFGSVFALTLLSLLLQKPVYEAQGQILYKKTSPISSY   63
RQEc_03216      AETDAGYGQLFSVLMRRRFWLLGVLGGVLSLATLYTLIAQPTYKSTMQLLIEANYQGRRE   72
LS2_05641       AETDAGYGQLFSVLMRRRFWLLGVLGGVLSLATLYTLIAQPTYKSTMQLLIEANYQGRRE   72
RQEc_04624      KKRGLNLRLLWRTIQRNLLLISGFTTIVTATTFYLGLNSPHIYEGDFRLLVEPITSEAKL  120
LS2_01047       KKRGLNLRPVWRAIQRNILLITGITAIVSTAALYSNLNSPRSYEGNFRLLVEPITTEGKF  120
THF_08171       QKGDWDFKRVLGICKRRWQVVSGVTILVAGYMIGASLSAQPEYLGRFRLLVEPVNAGSNV   91
RQEc_07349      TDDEWDLRQLLTVLRRRALVIGSVAIALSATVCVLTLTRQPQYEGKFQLLVEPVTKESKL   97

THF_06826       TGVGSEFDTIEPLVQS---NSPLSTEVEVLRSLPVIEQTIDRLDLK--TDQGKLMDSQQF  116
THF_07722       TDLTAQNREISGLTQT---INPLDTEAEIIRSAPLIQKTITTLNLK--DAKNLPLSIDQF  122
RQE_00387       TGLGKELGQIDALANSPDSNIPANTQAEIISSAAFAQKVMTELSLK--NEQGEPLKLKEF  122
LS2_00769       TGVGKEVGELDPLQQQG---NPVDTEAVVVLSVPNTQETITRLELK--DKSGAPLKRKQF  118
RQEc_03216      TDPQQLQSQQFAD--SSVKIDDYATQINVMRSSLLIQRAVDILRPEYPTIDVEEIKKNLV  130
LS2_05641       TDPQQLQSQQFAD--SSVKIDDYATQINVMRSSLLIQRAVDILRPEYPTIDVEEIKKNLV  130
RQEc_04624      TDPSVLSRDEQVA--T-AINVDYPTLLEVLQSPGLLSEVAKQVQARYPDVSYDSLSVDLV  177
LS2_01047       TDPSVLSRDGGVV--T-TNSVDYPTLLQVLQSPGLLSKVAKQIQSRYPDVSSDSLS----  173
THF_08171       SELTS--GASEGN--G-DSNLDYDTQIQVLQSPELIEEVLGILQPDYPTLSYRAIV----  142
RQEc_07349      EDLTQIPGINTNL--Q-KEGLDYDSQIKVLQSPELMAPIIKNLSTQYPDITYDSLLE---  151

THF_06826       LDNLSIKP-----VSEADVLRIRYKDRDLARAKRVVNTLMDVYLEASVNSRNRDAIAAEL  171
THF_07722       LKQIKAKG-----IRGTDVLEVSYRGENPQEVAAVVNTLMQNYLNSNIQVNRAEATAARK  177
RQE_00387       QNRLEVSN-----IKGTDVIQISYKSTVPEEARAVVNQLMNLYVQHNIFTNRSQAVAARE  177
LS2_00769       LKQLKVIS-----VKGTDVLQISYRDVDSKRAAAVVNTLMTIYLEGNLLANRTAAGGARK  173
RQEc_03216      LRQVVEE------KVNTKIVEATYRDNDPIKTQRVLEAIYRVYQIYNREQQEQRLKKGLK  184
LS2_05641       LRQVVEE------KVNTKIVEATYRDNDPIKTQRVLEAIYRVYQIYNREQQEQRLKKGLK  184
RQEc_04624      SKNLVIQRIGTNLSDFTKLIEVRYKGEDAAKVQFVLEELAKGYLKYSLENRKTRIGLGVQ  237
LS2_01047       -RDLVVQRIGTNLSDSTKLIEVLYKGEDPKKVQFVLEELAKGYLKYSLEDRKTRIGSGVQ  232
THF_08171       -GGLEISR-----VGKTKILEVQYQGTDAEKMEVVLKQLVKSYLQYSLSQRQTNLRQGIR  196
RQEc_07349      FEDLKINR-----LAETKILEISYQDPDPQKIQFVLQQLAKDYLKYSLQERQTNLRQGIQ  206

THF_06826       FLETQIPEAKSTVERLEKTLRQFREASRIVDLPEEQRVAIADLSDLGRQIVTAESQFADT  231
THF_07722       FINKQLPDVEARVSAAEAALQKFKEANRVVNLEEEAKVAVEDLSKLSEQFTQAQAALADT  237
RQE_00387       FIAAQLPKSKANVQQAEAELREFKEQNKILDLENEAKSAVTVISDLESQIGKAQVALANA  237
LS2_00769       FIDKQLPQAEASVRQADLVLRRFKEQNKVVALDEEAKSAVAITADLQRQVNSTQTQLADA  233
RQEc_03216      FINEQIPEVRKNVDQAEGRLEQFRKNQNLIDPEQQAASIAQSVNAIQQERQVLRAQFNDF  244
LS2_05641       FINEQIPEVRKNVDQAEGRLEQFRKNQNLIDPEQQAASIAQSVNAIQQERQVLRAQFNDF  244
RQEc_04624      FIEDQLPRLQQQVNDLQGKVQTLQQHYTLSDPASDGAAVSQQLREIQAQRLETQRQLQEQ  297
LS2_01047       FIEDQLPRLQQQVNDLQGKLQTLQQHYTLSDPASDGAALSQQLREIQAQRLETQRQLQEQ  292
THF_08171       FVERQLPANQQRVDQLQRELQAFRQNNNLLDPDSQAEQISEQINGLAQQRAEVNQALTIA  256
RQEc_07349      FVDSQRPQLQLRVNSLQKQLQRFRQQYNFIDPEVKAEQLSQQVSAIKLQRLDTEKQLAQA  266

THF_06826       NGQLTLLQNQLD-----------LPIDQAISLTILSQSEGVQSVLKEIWSTEAQLTSELN  280
THF_07722       TSRNQSLQAQIG-----------LNSSQAIALTTLSQSIAVQQVLKDYREVQDQLAVQRT  286
RQE_00387       NSRSTALQKEMG-----------MNAQQATVVNSLSQSPGVQKALTDLQELESQLAVERN  286
LS2_00769       QRQSAVLKNQLN-----------KDLQQAVPASVISQSPGVQDVLKEVQQLESQLAIERS  282
RQEc_03216      QARFGALQQ-----------QLARSPQQALTASRLSQSPRYQSVLNELQKTELALAQRRV  293
LS2_05641       QARFGALQQ-----------QLARSPQQALTASRLSQSPRYQSVLNELQKTELALAQRRV  293
RQEc_04624      KTLYANLQKQLK-----------LAPNEVIAASTLSEDPHYQELLTQLKKVESQIAVESA  346
LS2_01047       KTLYANLQKQLK-----------LAPNEVIAASTLSEDPHYQELLTQLKKVESQIAVESA  341
THF_08171       ------------------ESDLALTQDDTGVVTTLNQAPGYQQIVDELRGIEIQIAEELT  298
RQEc_07349      RTLYAALQGQSGAEFAASSSLYAALQDQPGVGQTQNQASIYQKLLAQLRDIDSQIAAEST  326

THF_06826       RFTQDAPSVQRLRSKIQLLNAELQNRITRIPGQPVAVGTDNPQ-------VSQLQQQLTT  333
THF_07722       RYQADHPEIANLVRKEAALSQQLETRVSQALGVAQPVLNEDLQ-------LSALKQNLAE  339
RQE_00387       RFQENSPVIENLKSKEASLKAVLQERMQGVVG-ERQQFNDNLQ-------IGELQQKLTE  338
LS2_00769       RFREDSPTVVSLKNRKAYLDSLLQERVRGVFGGQRQNFQGSLQ-------SSQFEQKLTE  335
RQEc_03216      VFTDEDPNVQKLLELRQSQIQMLQSEGGRVLGQQQAQANNAGGELLQEGQFSQIDQTLTN  353
LS2_05641       VFTDEDPNVQKLLELRQSQIQMLQSEGGRVLGQQQAQANNAGGELLQEGQFSQIDQTLTN  353
RQEc_04624      QFTEENPALQSLREKQKNLSLLLKQESGKIVGQKLATTTANPQVMTF---QNSLRQGLIK  403
LS2_01047       QFTEENPALQSLREKQKNLSLLLKQESGRIVGQKLSTTTANPQVMTF---QNSLRQGLIK  398
THF_08171       RFGRQSLNIQVLEEKRNNLIPLLQQEAQQAVGGKVAE-----------------------  335
RQEc_07349      RFQEDSPTVQALQNKRENLLPVLQQEASRVLGDKRLE-----------------------  363
```

(continued)

```
THF_06826       DIVKLAGAQQGLRQQLDVLRSQRQAYYQRITQLPRLDQTQRELLRKLEASQSKYLALLKK    393
THF_07722       SLVESEAERLGLASRVSILNAAISNSQARMSVLPKLEKTQQELERSLQVARSTYEQLLKR    399
RQE_00387       EFVKSEVERVVLTNQLTALNNVYSSYQQRANVIPQLQQRQQDLARDVEAARSTYETLLQK    398
LS2_00769       NLVESEARRLGLSSQLAALFKVRDAHRQRIDTLPQLEQKQRELERQLQASQAMYSLLLQK    395
RQEc_03216      QLVDVQTNLSALRARDVSLAQTQQQLRAELDRFPVLLAEYGRLQPLVQINREKLQQLQRA    413
LS2_05641       QLVDVQTNLSALRARDVSLAQTQQQLRAELDRFPVLLAEYGRLQPLVQINREKLQQLQRA    413
RQEc_04624      QMVDTANQAQVLEVRNQAVTQTEAWLEQQVRVFPAITRQYNDLQRQLEIATKTLNQLLIQ    463
LS2_01047       QMVDTANQAQVLEVRNQAVTQTEAWLEQQVRVFPAITRQYNDLQRQLEIATNTLNQLLIQ    458
THF_08171       ----LATQIQTLQIQRDAIDQAQTQVNQSFQQMPGLSRQYTDLQRELQVATESLNRFLAA    391
RQEc_07349      ----VVNQISILKARAAKIAQAESSLNQQIQQLPTLARQYTDLQRELKVATESLNRFLEK    419
                            *  :   :              :* :      *   ::        :

THF_06826       SEEVRIAASLRTGNAEILSPAMEVGGKT------TSRITTY-LSAALLGSLAAIALVYFR    446
THF_07722       LQEVEVAENQNVGNARIVSAALKPVIPI------SPRLPISLAIGAILGLLSGILTAFVL    453
RQE_00387       FQEVRITENQNVGNAQIVDTGRVLEAPV------APRKLLNFGVGGVIGVLLAGFTALAL    452
LS2_00769       LQEARIAENQNVSNARIISSAWVAEDPV------APRKSVYIITGILLGSILAVATVWIL    449
RQEc_03216      QQELSLEIARGGFDWQALEEPQLGRQIG-------PNLQRNLLLGLVAGLMLGGVAAFMR    466
LS2_05641       QQELSLEIARGGFDWQALEEPQLGRQIG-------PNLQRNLLLGLVAGLMLGGVAAFMR    466
RQEc_04624      RETLRVEAAQKEVPWQVVSQPRIPHDAAGNLIPADSDTAKKLAMGVMAGFMLGLGLALLK    523
LS2_01047       RETLRVEVAQKEVPWEVISQPRIPGDAAGNLIPVPSDTAKKLAMGVMAGFMLGLGLALLK    518
THF_08171       RETLQIQASQSEIPWQIVEPPSIAAAAA-------SNWLQDLLLAMAIGGTAGVATAILL    444
RQEc_07349      RESLQIEAAQKEVPWQVIAAPQVPQEPI------SPNVMRNLILGAIAGLLGGMGAALLV    473
                 :   :          . :                          .   *   .   .

THF_06826       EQQDRSLKSSNQAQKIL-GLPVLSEIPIASTKNWLPF------------GGFAS------    487
THF_07722       DAIDKSVKTPEEAQKLA-GYPLLGIIPVVSGKTHGRD----------RT-----------    491
RQE_00387       DARDSSVKTVREARELL-GYTLLGIIPNFEKIDKAAS----------RTRNSEQ------    495
LS2_00769       ETQDKSIKTVDQAREVF-GFTLLGVIPTFKEAEKISD----------RDQDLEQ------    492
RQEc_03216      EVVDDSVHSSDELEKQV-ELPLLGMTPELPRGKASEPIVSL---PFGEPKRLAP------    516
LS2_05641       EVVDDSVHSSDELEKQV-ELPLLGMTPELPRGKASEPIVSL---PFGEPKRLAP------    516
RQEc_04624      EKYRNVFYTAEDIQEAITELPLLEVSIARSAKQ-------------------FHNYAVDAR    565
LS2_01047       EKYRNVFYTAEDIQDAITELPLLEVSLDRSVKQ-------------------FTNYSVGAR    560
THF_08171       DLLSNTYQTVDDLKQKV-KLPILGQLPFQPTLGSDQGNLSIGRSLLARMSNLVPNPVDPE    503
RQEc_07349      ERLDNVFHSPDDLKEVT-KLPLLGVIPFRKHLKQLAY-PSTGAMALTRLRHLHSA-----    526
                :       :  : ..      :*

THF_06826       --------------S------LPHLVVHEAPYSPASEAYRILQTRLS--SIHEPCRTIALT    526
THF_07722       --------------A------GSSLPVRDNPYSSISTNFEMLQTALGFTLSDQPLQVILVT    532
RQE_00387       --------------A------VTKIPVKDTPRSTISETYGMLYANLKFLSSDQEVRVITIT    536
LS2_00769       --------------S------PPEIVVRDSPQSPFSEAYQMLQANLKFLSSDQ-LKVIVVT    532
RQEc_03216      ----------------------WTVQVSNWPPSWESLDLIYKNIQLLNSVSAFQSLMVT    553
LS2_05641       ----------------------WTVQVSNWPPSWESLDLIYKNIQLLNSVSAFQSLMVT    553
RQEc_04624      AIEG-------------------TETDYSDASLFLQTFSSLYTSIRFLASKSPVHSLVVG    606
LS2_01047       AIEG-------------------TKTDNSDASLFLQTFSSLYTSIRFLASKSPVHSLVVG    601
THF_08171       -LAGLTPALPISNSQPNSTIVSLPSNSTYNTSEFVESLRLLYTNLQLRSLGQPVQSIIVS    562
RQEc_07349      ----TEADVEIKPDGPNLVFGNLNNTHDESYFPFVEAFRSLYTNIGFLGSDAPIQSLVIS    582
                                                      :   :         : : :

THF_06826       SAVMGEGKSTVVANLALTFAQVGKRVLVIDSDFRRPIQDKIWGLSNEVGLSNVLSDQVSV    586
THF_07722       SAVSGEGKSFIAANLATAAARMGRRVLLIDADMREPSQHTIWQRINVKGLSDILVEQATL    592
RQE_00387       SAVPKEGKSTVSANLALAIAQLGFRVLLVDADMRRPSQHQFWELSNAVGLSNVIVGQTEL    596
LS2_00769       SCLPKEGKSTVSANLAVAMAQVGRKVLLVDADMRRPMQHHIWELPNQLGLSNLIVGQGEA    592
RQEc_03216      SAVAGEGKSTLALGLALSAARLHQKVLLIDADLRRPNLHKMLNLPNEQGLSTLLVNDATV    613
LS2_05641       SAVAGEGKSTLALGLALSAARLHQKVLLIDADLRRPNLHKMLNLPNEQGLSTLLVNDASV    613
RQEc_04624      SAEPGDGKTTIALELAKAAASMGQRVLLVDANLCQPQLHTRLGLPNLQGLSNILSGNLEP    666
LS2_01047       SAEPGDGKTTIALELAKAAASMGQRVLLVDANLCQPQLHTRLGLPNLQGLSNILSGNLEP    661
THF_08171       SAEPGDGKTTIAVQLAQTVAAMGKRVLLVDTDLRRPSVHTQLQLQNQRGLFEVITGNLPI    622
RQEc_07349      SALHEDGKSTVAMHLAQAAAAMGRRVLLVDADLRLPKVHTRLGLPNEQGLSNVISKNLPV    642
                *.   :**: :   ** : * :  :**::*:::  *  .      *  **  ::  :

THF_06826       EMATW--KITPELEVLPTGAAPANILSLLP--RFNTLLQQ-LKAYDYILIDTPALTLTAD    641
THF_07722       REATH--EVQPNLGLVTSGTMPPNPVALLDSQRFASLIESIRQTYDFVILDTPALNLVPD    650
RQE_00387       KRAIA--EVMPNLYVLTSGVLPPNPLALLNSKRMASLVEVFSQSYDFVILDTPPLNAAAD    654
LS2_00769       RTASK--EVMPNLHVLTAGVTPPNPIALLDSQRMASLVEVFSANYDFVIIDTPALSVAAD    650
RQEc_03216      PTQSSIQSSGSYIDILTAGPTPVDPANLLSSQRMRELMSAFEETYDLILMDAPPVLGMVD    673
LS2_05641       PTQSSIQSSGSYIDILTAGPTPVDPANLLSSQRMRELMSAFEETYDLILMDAPPVLGMVD    673
RQEc_04624      KELIQRSPLEENLFVLTAGQQPPNSTRLLASTRMQPLIEQLQSAFDLVIYDTPHLLGLPD    726
LS2_01047       KELIQRSPLEENLFVLTAGQQPPNSTRLLASTRMQPLIEQLQSAFDLVIYDTPHLLGLPD    721
THF_08171       RKALQNLPHLPLCRVLVAGQVPSDPTQVLASKQMQQYMTAFQQIADLIIYDAPPLASLAD    682
RQEc_07349      LEVIQRAPLWNNLFVLTSGQIPPDPTKLLSSKKMQNIMEQLRQEFDLVIYDTPPLLGLAD    702
                          :: :*  * :   :*   ::   :       * :: *:* :    *

THF_06826       AQALCQLVDGVLLLVRLNWRAMNIQSAEMVNQIFEQSSYRVLGQVINDGTPLNVY----P    697
THF_07722       GLTIGKFVDGILFVVRSQVA--DSIDVIQAKTLLTQTGHRVLGMVANGCLSEKKVGQSYY    708
RQE_00387       ARILGVMTDGILMVVQPQLV--DSASAISAKELLQHSGGNILGMVVNGLNPENEPDSYFY    712
LS2_00769       APILGKIADGVLLVVRPGVV--DTASATYAKEFLEKSGQKVLGQVINGVIPKNEPYSYYY    708
RQEc_03216      AILAASFCSGVVLVSRMGQV--TKTELAQATTM--LSKLNVIGVVANGVNGAATSTYV--    727
LS2_05641       AILAASFCSGVVLVSRMGQV--TKTELAQATTM--LSKLNVIGVVANGVNGAATSNYV--    727
RQEc_04624      ANFLAAHTDGILMVVRVGKT--NRSLLTQVLNKLNSFHLPILGIVANHVEESTNSASD-S    783
LS2_01047       ANFLAAHTDGILMVVRVGKT--NRSLLTQVLNKLNSFHLPILGIVANHVEESTNSASD-S    778
THF_08171       ASLIAPHTDGVLLVVGLGKT--NRSTLTQTLENLTISQVSILGIVCNSLSA---------    731
RQEc_07349      SSLLATYTNGIVLVVRMDKT--DRSVLVQALDQLKFSRANLLGMVCNAAKTNYTPQSY--    758
                .       .*::::               .          ::* * *

THF_06826       LNNTVLLNS-NGSLRRVA-PPSRI-----DLQTQVAMPKKL       731
THF_07722       ANQDSQQRE-----TSSK-VSSKA-------------------      726
RQE_00387       AKEYYSEQDLTAVNDSFK-VGQEI-----NRF----------      738
LS2_00769       FSKQYSAEASVATSANSS-PKS--------------------      729
RQEc_03216      ---PYTREH-GVV----LNKVWGS------------------      743
LS2_05641       ---PYTREH-GVV----LNKVWGS------------------      743
RQEc_04624      HHQHYEQNH-QGSSTFVKKPEFKSTLLTAGKQTDDASP---       820
LS2_01047       HHQHYEQNH-QGSSTFVKKPEFKSTLLTSGKQTDDASA---       815
THF_08171       ----------------------------------------       731
RQEc_07349      ---YYCQPH-TQSARLS------------------------       771
```

Appendix 4

The sequences of putative WZX homologues

```
LS2_00776       ------------------------MGSKFLKNSFYNVI-------------AGAVRIGLAL    24
RQEc_04585      ----------MNLDKLKPWIGRFGKNSLARNTLWMLLAQGMR-------------LVLQA    37
THF_06609       --------------------MVSI-KQLALRGVFWTIASYGISQ--------------IIRF    27
THF_08804       ------------------------------------------------------------    0
RQEc_04622      MHKHLLKFSTSHFPTRQGNMSSI-KKLAIWGAVWTIAGYGASQ--------------ILRF    46
LS2_01044       MHKYLLKFSKHDFPTRLGNMSSL-KKLAIRGMVWTVAGYGASQ--------------ILRL    46
RQEc_06442      ------------------------MKTIFKGARWVWGGRTATAATVQTVLVRVLILATNM    36
LS2_03996       ------------------------MKTIFKGAHWVLGGRTATAATVQTVLVRVLILATNM    36
RQEc_02592      ---------------MNPPASQL-GRRMLNGTIWIFL-------------AEALLLPTGI    31
LS2_07123       ---------------MNPPASQL-GRRMLNGTIWIFL-------------AEALLLPTGI    31

LS2_00776       ITIPVLIRLLKVEGYGVWALASAIVELVVLS-GNGLSVTTTVFVSRDLAQTDSNDGL-SK    82
RQEc_04585      LYFVIIARALGAEQYGAFVGALSFVAILAPFASLGSGNLLIKNV-------SRNRALLNE    90
THF_06609       GSNLVLTRLLFPQVFGLMSLAYVFITGLHLFSDLGIHTSLV------QNKRGTEPVFLNT    81
THF_08804       -----------------MALVNIFVVGLKLFSDVGIGLSVV------QNKRGEEPAFINT    37
RQEc_04622      GSNLILTRLLFPELFGLIALVNIFIVGLHLFSDIGIGPSII------QNKRGDDPAFLNT    100
LS2_01044       GSNLILTRLLFPELFGLMTLVNIFITGLHLFSDIGVGPSII------QNKRGDDPVFLNT    100
RQEc_06442      GTGIITARALGPDGRGEQAAMALWPQFLAYAMTLGLPAA-LRYN--LKRYPDQKTELFSA    93
LS2_03996       GTGIITARALGPDGRGEQAAMALWPQFLAYAMTLGLPAA-LRYN--LKRYPDQKTELFSA    93
RQEc_02592      LTAAFLTRQLGPEGYGLFTLAATLTAWIEWSVTSVFTRATIKFVGEAEDWRPVGA----T    87
LS2_07123       LTAAFLTRKLGPEGYGLFTLAATLTAWIEWSVTSVFTRATIKFVGEAEDWRPVGA----T    87
                                           :

LS2_00776       T--FTIVIGGTLILATLAAFALW-ISAGAIVEFF--PDLKQIQKSAVTQAVQIGGLVVWA    137
RQEc_04585      YWGNALFMISSSG----LALVILVLLISPFFLPKTIPALLI-------FLVAITDLIFYR    139
THF_06609       AWTLQIIRN------VGLWLCCI-IIAIPAANFYNEPQLVW-------VL-PIAGLGT--    124
THF_08804       AWTIQVFRG------VGLWLVCL-LIAYPLSALYNEPQLVW-------LL-PIVGLST--    80
RQEc_04622      AWTLQIIRS------FALWFGCL-LIAWPLAQFYGEPRLSW-------LI-PVVGLNT--    143
LS2_01044       AWTLQVIRS------FGIWLCCV-LMAWPVANFYDMPQLLW-------LI-PLVGLNT--    143
RQEc_06442      ALLLSILLGIAATLVGIVFLPQW-LSQYS-------PEVIR-------FA-QWLMLLS--    135
LS2_03996       ALLLSILLGIAATLVGIVFLPQW-LSQYS-------PEVIR-------FA-QWLMLLS--    135
RQEc_02592      VLRLHLFMG------GGVMLLVW-LFAAPVGKLMGEPVLTT-------YL-RLFALDI--    130
LS2_07123       VLRLHLFMG------GGVMLLVW-LFAALVGKLMGEPVLAT-------YL-RLFALDI--    130
                     :.           :    :          * :               *

LS2_00776       RL------LQQVLIGIEQ-AYQRYGTL-------NFLNTLQWVMLSLGLIGVAW-----L    178
RQEc_04585      ILDIAGQAFQSVL------WLSKTALINILPSVTRLMAALIWVSFFPKSGALEWGCLYLV    193
THF_06609       --LISGFNSTGLASLTRSLSVRELATYELGGQIIGIGVMLILAWFNRSIWAL------LL    176
THF_08804       --ILDGFASTAPFTLGRRLELRKLTLLEISTQAVQVVVMLTWAYFSPTIWAL------AA    132
RQEc_04622      --IISGFNSTALFTLNRRMAISQLAIFELGGQIISLTVMIFWAYFNPTIWAL------VV    195
LS2_01044       --IINGFNSTALYTLNRHMAIAQLAIFELGGQIISLAVMLVWAWFSPSIWAL------VV    195
RQEc_06442      PMILLSGNFVAALEATED--FT---TANQAQYLAPLM---TLVLLGLLAVAQ------IL    181
LS2_03996       PMILLSGNFVAALEATED--FT---TANQAQYLAPLM---TLVLLGLLAVAQ------IL    181
RQEc_02592      PLFCLGYAHRSILVGVGS--FTQRAIATAGRWIARLVLILVLVKIGLSVQGA------IL    182
LS2_07123       PLFCLGYAHRSILVGVGS--FTQRAIATAGRWIARLVLILVLVKLGLSVQGA------IL    182
                                                .      . :     .

LS2_00776       GGRTGELMQWQAITSVVMLVSHLWFVTAFVQSTH-------------LRPILAIK--KGL    223
RQEc_04585      GTAVS------------ALIGVLLVCHYLGVPKL-----AL-----------------SR    219
THF_06609       GTVIS------------SLVQLVWS-HAISPNPPNRLVLEK-----------------KA    206
THF_08804       AGLVS------------GLFKTVWT-HRLIPQYTNRFQWEQ-----------------RS    162
RQEc_04622      GSLVS------------TSIQMVWS-HRLISGSSNRLSWDQ-----------------DA    225
LS2_01044       GGLTS------------AAVQMIWT-HQLIPGTSNRFTWNQ-----------------KA    225
RQEc_06442      TPFTA------------VL-------AYVIPGLP-IFFWLLHRTWKYFRPRWQRL--GSS    219
LS2_03996       TPFTA------------VL-------AYVIPGLP-IFFWLLHRTWKYFRPRWQRL--GSS    219
RQEc_02592      GSVGA------------SLVELVISRFYIRPSLFGHSTFPARKLWSYAVPLFLFALSMRL    230
LS2_07123       GSVGA------------SLVELVISRFYIRPSLFGYSTFPAQKLWSYAIPLFLFALSMRL    230
                    .                       .

LS2_00776       EIAHHSFINWL--ICLGSVAFGRGDRLIVAASLGPEAIGIYAAIIEATSALSSFAAL-PV    280
RQEc_04585      IKPEIREGLYFSVSLSAQTIYNDIDKTMLASLVTLEATGIYAAAYRLIDVAFVPVRSILA    279
THF_06609       VGEIFSFGKWIFLSTALTFFAMQSDRLILGKLLGLQLLGVYGIAVTLAEIPKQVTMAVGG    266
THF_08804       VTEIFSFGRWIFLSTALTFLAEQADRLMLGKLLSLELLGIYGIAVMLSDVPRQLALTLSN    222
RQEc_04622      VNNLFSFGGWIFISTAGTFLAEQTDRLILGKLLSLEMLGVYGIAFVLSDIPRQVLLALSG    285
LS2_01044       AKEIFSFGKWIFLSTAMTFFAEQADRLILGKLLSLEMLGVYGIALTFAELPRSVTLALSG    285
RQEc_06442      YKRLTSYGLRAYGIDLLGTLSGQVDQALVVNLLPPASMGLYVVALSLSRMLSLFHSSIIT    279
LS2_03996       YKRLTSYGLRAYGIDLLGTLSGQVDQALVVNLLPPASMGLYVVALSLSRMLSLFHSSIIT    279
RQEc_02592      YEKLDL-----FALKLLGGTAAQ--------------VGFYGAAQNLSLVPGIFSLSFSP    271
LS2_07123       YEKLDL-----FALKMLGGTAAQ--------------VGFYGAAQNLSLVPGIFSLSFSP    271
                                                    *.*
```

(continued)

```
LS2_00776       EIAHHSFINWL--ICLGSVAFGRGDRLIVAASLGPEAIGIYAAIIEATSALSSFAAL-PV    280
RQEc_04585      IKPEIREGLYFSVSLSAQTIYNDIDKTMLASLVTLEATGIYAAAYRLIDVAFVPVRSILA    279
THF_06609       VGEIFSFGKWIFLSTALTFFAMQSDRLILGKLLGLQLLGVYGIAVTLAEIPKQVTMAVGG    266
THF_08804       VTEIFSFGRWIFLSTALTFLAEQADRLMLGKLLSLELLGIYGIAVMLSDVPRQLALTLSN    222
RQEc_04622      VNNLFSFGGWIFISTAGTFLAEQTDRLILGKLLSLEMLGVYGIAFVLSDIPRQVLLALSG    285
LS2_01044       AKEIFSFGKWIFLSTAMTFFAEQADRLILGKLLSLEMLGVYGIALTFAELPRSVTLALSG    285
RQEc_06442      YKRLTSYGLRAYGIDLLGTLSGQVDQALVVNLLPPASMGLYVVALSLSRMLSLFHSSIIT    279
LS2_03996       YKRLTSYGLRAYGIDLLGTLSGQVDQALVVNLLPPASMGLYVVALSLSRMLSLFHSSIIT    279
RQEc_02592      YEKLDL-----FALKLLGGTAAQ--------------VGFYGAAQNLSLVPGIFSLSFSP    271
LS2_07123       YEKLDL-----FALKMLGGTAAQ--------------VGFYGAAQNLSLVPGIFSLSFSP    271
                                                    *.*

LS2_00776       QPLVPVLSHQSALEDT---NRSNLKQQIKQALKANGLVALGCSGLLIIVA----PVIMQL    333
RQEc_04585      AA----YAKFFQHGTAGISGSVVLAKRL--------LPIAGIYGAIAGIGLLLFAPVVHY    327
THF_06609       KVIFPMFSKFVSLPRP---E---FRSKIRRGRLPILLVTAPALAVMISFG----DILITT    316
THF_08804       RVVLPAASKLADLPRP---E---LRAKILKHRWRLLLVLMGAIALLAGFG----DIIINF    272
RQEc_04622      KIIFPALSKLADFPRE---I---LRAKILQNRQPILVVSALVLAVLVGFG----DILISA    335
LS2_01044       KVIFPAVSKLAELPRE---T---IRAKLLQNRKPILLALAFCLTILVGFG----DLLIKA    335
RQEc_06442      VL----LPKTAARPVE---EVVALTGRAARVGM---VLTVLTAIAVIIPV----PILLRL    325
LS2_03996       VL----LPKTAARPVE---EVVALTGRAARVGT---VLTVLTAIAVIIPV----PILLRL    325
RQEc_02592      LL----LSTLSRTLYA---GDTVLAQKISRDAMRAVIWMLPFAGMTAGAA----PEIVSL    320
LS2_07123       LL----LSTLSRTLYA---GDTVLAQKISRDAMRAVIWMLPFAGMTAGAA----PEIVSL    320
                                  :  :        :                     ::

LS2_00776       LLTGAATGTHIV--LFKVAVIIYALCSLNAVGFYILLSMAVN-LVMFLQLGSGILALILI    390
RQEc_04585      LLGDEYLAAVEALRWLAPLPFLKAMHYFAADTLTGAGFQGVRSVVQVIVAV---FNFLVN    384
THF_06609       LYDNRYIDAAWMIPLLALGIWPIILVTTIDGALFAMGNPTPSTWGFFYSFL---ALAGGI    373
THF_08804       LYDERYGAAAWMLPILALGIWPRLMCATIETSLYAINKMQYTTAANFCRLL---STVFGI    329
RQEc_04622      LYDKRYVQATWMLPLLALGIWPRVLTQTADPSLFAIGKPRYVAYGSFLKFF---FILIGL    392
LS2_01044       LYDDRYIDAAWMLPLLALGIWPRMLCNTNEPSLFALGKPQYSTGGNFTRFL---CTSVGV    392
RQEc_06442      LYGAEFLEAVPVFRILIVEEVIGGTAWVLSQAFMALGRPGTVAILQALGLG---LSVPLM    382
LS2_03996       LYGAEFLKAVPVFRILIVEEVIGGTAWVLSQAFMALGRPGTVAILQALGLG---LSVPLM    382
RQEc_02592      LFGSAFLPTAPLLALLIFGAIALAMISIGTAILTAAGKPNWTFALAGPLLP---LSIAGH    377
LS2_07123       LFGSAFLPTAPLLALLIFGAIALAMISICTAILTAAGKPNWTFALAGPLLP---LSIAGH    377
                *       :      :                :

LS2_00776       AIGGKYFGLLGAIVG-----NIGFFVSWLMILYGLKELKLSS----------------LL    429
RQEc_04585      LWLIPLYSWKGAAWSSLASDAFLMLSLWMMVA-FLYRQQAQRLKENQ-------------    430
THF_06609       WAGFHFFGIVGAVAAVPLSNV-PYYAAV--AH-GLRKEKIDCFDHDLYTTGLLVASSVIL    429
THF_08804       WLGYTLLQVPGAVIAVALNDI-FYYSVV--SY-GLHREGLGCLRQDLKITSLLVILLVAI    385
RQEc_04622      PLGFSLMGLSGAVAVVALNDL-PFYGAV--TY-GLWREGLTAVVQDIKATALLLALLTVV    448
LS2_01044       LLGFSLMGVPGAIIAVTLNDL-SYYVIV--NY-GLWREGLDGLMQDAKATMLLVVLITSA    448
RQEc_06442      ILFIPAYGLEGAGLALLCST-VLRLVLVLIGYPTILKVRPP---------------GLI-    425
LS2_03996       ILFIPAYGLEGAGLALLCST-VLRLVLVLIGYPTILKVRPP---------------GLI-    425
RQEc_02592      ILLIPRLGAIGASIVSTLFASVGALATVLAIY-SLWQVFPP---------------FRTV    421
LS2_07123       ILLIPRLGSIGASIVSTLFASVGALATVLAVY-SLWQVSPP---------------FRTV    421
                         **                      : .

LS2_00776       WLKCLA----FPLSWFVACVLFSFLISTNLSATILTAIIQILVLIGWFLMSQPQYNQF--    483
RQEc_04585      ------------------------------------------------------------    430
THF_06609       IAARCAFGLPLPLLPS-----------------------------------------------    445
THF_08804       MMGRGALGLGSPIDDIWQ--LF-----------------------------------------    405
RQEc_04622      LTGRVILGLGLPIHSLL----------------------------------------------    465
LS2_01044       LVSRLILGLGYPISGLP----------------------------------------------    465
RQEc_06442      ----------MT-KEDWR--FLQ--QVFRLN------KS-----------------------    443
LS2_03996       ----------MT-KEDWR--FLQ--QVFKLN------KS-----------------------    443
RQEc_02592      WQSILICGLAYALAAFWS--TDNFLLLFKLS------VIALLIPLAFFLLGEFSHSEIAL    473
LS2_07123       WQSILICGLAYALAAFWS--TDNFLLLFKLS------VIALLIPLAFFLLGEFSHSEIAL    473


LS2_00776       ---LTKRIRP--------------        490
RQEc_04585      ------------------------        430
THF_06609       ------------------------        445
THF_08804       ------------------------        405
RQEc_04622      ------------------------        465
LS2_01044       ------------------------        465
RQEc_06442      ------------------------        443
LS2_03996       ------------------------        443
RQEc_02592      VRSLLGSIKKYSQKKESNTSSAKD        497
LS2_07123       VRSLLGSIKKYSQKKESNTSSAKD        497
```

Appendix 5

The sequence of putative Wzy homologues

```
RQEc_01924      MTQLYKPQFSSRKAESDNLFWGIWIRWHTLTVGERFVCANIVVLPVWWFVGLYQYMTVIL    60
LS2_00010       MTQLYKPQFSSRKAESDNLFWGIWIRWHTLTVGERFVCANIVLLPVWWFVGLYQYMTVIL    60
THF_00290       --------MQ------KSKI------WQ-------LFRFSLLTLPYLFYVSFVGLF----    29
RQEc_03112      ---MLHPNKR------LQTP------WN-------YAQLGLLIFPLIPWLGALGLF----    34
LS2_08392       ---MLHPNKR------LQTP------WN-------YAQLGLLIFPLIPWLGALGLF----    34
                         .       *.           :: :*   ::.    :

RQEc_01924      LLGVALYEFWQNREIRLKRPSLVVVSLFIFGAYIIVGRFFLLSRVLGALRGTFLIFPFPS    120
LS2_00010       LLGVALYEFWQNREIRLKRPSLVVVSLFIFGAYIIVGRFFLLARVLGALRGTFLIFPFPS    120
THF_00290       ---AVMLID-RFRFGRIALDPLTRNSLAIIGGLLILSCGFAEN-RSEALLQLTNFLPFFL    84
RQEc_03112      ---LALVGTWRQQYHEIIHRPLN-WVLAVLAGWLVITSCLAFY-PLDAFLGLFNFLPFFS    89
LS2_08392       ---LALVGTWRQQYHDIIHRPLN-WVLAVLAGWLVITSGLAFY-PLDAFLGLFNFLPFFS    89
                    .:    : :   :    *    * ::.. :::   :       *:     ::**

RQEc_01924      ACLL-WYIQSKNIRIR---LEVVAWACTVSVVQMLGLWLLIQFVLPDSFFFPPRLNIFAR    176
LS2_00010       ACLL-WYIQSKNIRIR---LEVVAWACTVSVVQMLGLWLLIQFVLPDSFFFPPRLNIFAR    176
THF_00290       FFSVLPYVLKETERLGQLAIDMVIVAIPLSVL-ALGEYLL-------KSTFIPRV-----    131
RQEc_03112      FFAAFSALIKTPAQLRQLAYILVITSVPVVLI-GLGQLFW-------GWVSPVEL-----    136
LS2_08392       FFATFSALIKTPAQLRQLAYILVITSVPVVLI-GLGQLFW-------GWVSPVEL-----    136
                       : .   ::      :*  :  : ::  **  :             .:

RQEc_01924      IAGVEEQTVLDNLLNADAAFDLSNLVPYAPNSNGMLGFNRFSMFFSYPEFFGLVAGFIGI    236
LS2_00010       IAGVEEQTVLDNLLNADAAFDLSNLVPYAPNSNGLLGFNRFSMFFSYPEFFGFVAGFIGI    236
THF_00290       ----------------------VRRIPWVRSLRNAPHKGRAMVMFTHPNSLANYLVLILG    169
RQEc_03112      ----------------------QGILGWVLAPQGNP-PGRIASVFMYTNVLAGYLVIIFI    173
LS2_08392       ----------------------QGILGWVLAPQGNP-PGRIASVFMYTNVLAGYLVIIFI    173
                                     : :.   .      *   .* : : :.    :*

RQEc_01924      LAL-----DIKHRLWSGL-------LFLACVFLIVLS-------ATRIVVVAFPIVVVLY    277
LS2_00010       LAL-----DIKHRLWSWL-------LFLACVFLIVLS-------ATRIVVVAFPIVVVLY    277
THF_00290       LGLGLILYDASRSLDRGLE-PQRYGR-----------------TRRVLLYAGTFSTLVG    210
RQEc_03112      LALGLLITSFQELKGQKNSSPRPSS-----------------LAPRLFLTVAVIGNLVA    215
LS2_08392       LALGLLITSFQEPKDQKNASPRPTLREAACASTSHSSRSRLRVYIPRLFLTVAVIGNLVA    233
                *.*     . ..                                  *:.: .  :  ::

RQEc_01924      YVISNFRKLWGPPIIFG-----------------LMA----VVSFTTLSIPQ--------    308
LS2_00010       YVISNFRKLWGPPIIFG-----------------LMA----VVSFTTLSIPQ--------    308
THF_00290       IFCSGSRNGLIVAVIQIILFCVCIQNNRKILLIGLLSGLGLVVGSAGLGMGRRS------    264
RQEc_03112      LILTNSRNAWAIAVCAGITFAIY--QGWRWLEVGVAAIAGSVLGA---AFGPSPLRQWLR    270
LS2_08392       LILTNSRNAWAIAVCTGITFAIY--QGWRWLEVGVAAIAGSVLGA---AFGPSPLQQWLR    288
                 . :  *:     :                    : :    *:.    .:

RQEc_01924      --TSDLVFSRITDSLQAVNQVRQASTETRLEVYRQTWEGIQENPFWGYQSTGEGGAGTHS    366
LS2_00010       --TSDLVFSRITDSLQAVNQVRQASTETRLEIYRQTWEGIQENPFWGYQSTGEGGAGTHS    366
THF_00290       --LDIMNW----------------ADDPRPRVWSIAWDLIKERPGLGWGLGNYKL-----    301
RQEc_03112      QVIPSFFWQRLTDQMYP----NRPLGSLRKTQWRFAWNLTQQRPWTGWGLRNFTP-----    321
LS2_08392       QVIPSFFWQRLTDQMYP----NRPLGSLRKTQWQFAWNLTQQRPWTGWGLRNFTP-----    339
                     : :                  . *   :  :*:  ::.*  *:

RQEc_01924      VILGNLLYQ--RGLVG----TIIFTFFWISL---FIWFYRTRAGRP-LTCFCVWLLYTLV    416
LS2_00010       VILGNLLYQ--RGLVG----TIIFTFFWISL---FVWLYKTRAGRP-LTCFCVWILYTLV    416
THF_00290       EFLPRLLNQYPSCLVERTYKVIPSNCADVSHPHNFWLLLGSEVGLSIMLGFSLWVGYICF    361
RQEc_03112      LYEAQ-------------------MHAWLGHPHNFFLMLTAEIGIPATILFCGWIIWIVF    362
LS2_08392       LYEAQ-------------------MHAWLGHPHNFFLMLTAEIGIPATILFCGWIIWIVF    380
                    .                     :.    *  :  :. *      *. *: :  .

RQEc_01924      SPTLQLVYEMPI-----------------SSLLI-LLCAVICRPKL---------------    444
LS2_00010       SPTLQLVYEMPI-----------------SSLLI-LLCAVICRPKL---------------    444
THF_00290       RGVRCVMTGK--------LSRWSHSVLLSYLFAFLGCVSFAMLDVTFYDARLNIMNWVLL    413
RQEc_03112      QGIHLLSNWSRIGRTGEGAI-EDRLIFFSYLLAFVACILFNTVDVSLFDLRLNTLGWLLL    421
LS2_08392       QAIQLLSDWSRIGRTGEGAIAQDRLILFSYLLAFAACILFNTVDVSLFDLRLNTLGWLLL    440
                     :                      * *:    *  :   .:

RQEc_01924      -------KSVPNSSPFKKLHHA  459
LS2_00010       -------KSVPNSSPFKKLHHA  459
THF_00290       AGIYTATQEWEES---------  426
RQEc_03112      SAIGGVVYHYRDQRDFGF----  439
LS2_08392       SAIGGVVCHYRDTAPGK-----  457
                          :
```