**UNIVERSITY OF NEVADA, RENO**

# Dimensions of Antarctic microbial life revealed through microscopic, cultivation-based, molecular phylogenetic and environmental genomic characterization

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in

**Biochemistry**

by

**Emanuele Kuhn**

Dr. Alison E. Murray/Dissertation Advisor

May, 2014

University of Nevada, Reno
Statewide • Worldwide

THE GRADUATE SCHOOL

We recommend that the dissertation
prepared under our supervision by

**EMANUELE KUHN**

entitled

**Dimensions Of Antarctic Microbial Life Revealed Through Microscopic, Cultivation-Based, Molecular Phylogenetic And Environmental Genomic Characterization**

be accepted in partial fulfillment of the
requirements for the degree of

**DOCTOR OF PHILOSOPHY**

Alison E. Murray, Ph. D., Advisor

Gary Blomquist, Ph. D., Committee Member

David Shintani, Ph. D., Committee Member

Joseph J. Grzymski, Ph. D., Committee Member

Christian Fritsen, Ph. D., Graduate School Representative

Marsha H. Read, Ph. D., Dean, Graduate School

May, 2014

# Abstract

Extreme cold temperatures have shaped Antarctic environments and the life that lives within them. Microorganisms affiliated with the three domains of life - Bacteria, Archaea, and Eukaryote - can be found in Antarctic environments from deep subglacial lakes to dry deserts and from deep oceans to cold and dark winter surface seawaters. This dissertation focused on the investigation of the microbial assemblage in two Antarctic environments: Lake Vida, located in the McMurdo Dry Valleys, and the surface seawater from the Antarctic Peninsula.

Lake Vida has a thick (27+ m) ice cover which seals a cryogenic brine reservoir within the lake ice below 16 m. This brine's environment challenges the conditions for the existence of life. Despite the perceived challenges of aphotic, anoxic and freezing conditions, the brine contained an abundant assemblage ($6.13 \pm 1.65 \times 10^7$ cells $mL^{-1}$) of ultra-small cells $0.192 \pm 0.065$ μm in diameter and a less abundant assemblage ($1.47 \pm 0.25 \times 10^5$ cells $mL^{-1}$) of microbial cells ranging from > 0.2 to 1.5 μm in length. Scanning electron microscopy provided supporting evidence for cell membranes associated with the ~ 0.2 μm cells and helped discern a second smaller size class of particles ($0.084 \pm 0.063$ μm). 16S rRNA clone library analyses indicated that the ultra-small cell-size assemblage was dominated by the *Proteobacteria*-affiliated genera *Herbaspirillum, Pseudoalteromonas,* and *Marinobacter*. Cultivation efforts of the 0.1 – 0.2 μm size fraction led to the isolation of *Actinobacteria*-affiliated genera *Microbacterium* and *Kocuria*. Based on phylogenetic relatedness and microscopic observations, we

hypothesize that the ultra-small cells in Lake Vida brine are ultramicrocells that are likely in a reduced size state as a result of environmental stress or life-cycle related conditions.

The previously unexplored deeper ice of Lake Vida (from 18 to 27 m) revealed an ice column banded by sediment layers up to 15 cm in length and a diverse and cell-rich microbial assemblage with cell counts ranging from $3.73 \times 10^4$ to $8.58 \times 10^6$ cells $mL^{-1}$. Illumina tag sequencing (iTag) targeting the 16S rRNA gene and RNA (as cDNA) indicated that the microbial assemblage from the lake ice and four sediment layers below 21 m was dominated by organisms capable of the reduction and oxidation of sulfur compounds in addition to high molecular weight complex polymer degradation. *Proteobacteria, Bacteroidetes*, *Actinobacteria*, and *Firmicutes*-affiliated genera were the most abundant bacterial phyla detected. The distribution of the microbial assemblage within the ice and sediment layers was correlated with the presence of sediment particles, total dissolved solids (TDS), total carbon, $SO_4^{2-}$, and $Na^+$ concentration. Chemolithoautotrophic and heterotrophic genera such as *Sulfurovum, Desulfocapsa, Lutibacter,* and *Desulforomonas* dominated the ice segments and heterotrophic genera such as *Cellulomonas* and *Conexibacter* dominated the sediment layers. Sediment layer cDNA iTag sequences indicated that the taxa carrying the potential for metabolic capacity were mostly the *Proteobacteria*-affiliated genera *Pseudoalteromonas* and *Vibrio*. Representatives of these genera are often identified in Antarctic marine environments such as sea-ice, seawater, and marine sediment and adapted to low temperatures and high salinity. Therefore, the detection of abundant and considerably diverse microbial assemblages in Lake Vida brine, ice and sediment layers indicates that life is likely sustained in isolated deep, icy and dark anoxic environments. This also suggests that if similar conditions are found elsewhere beyond Earth, there is the possibility to find life.

The second subject in this dissertation research was to conduct an in-depth assessment of the environmental genomics of the archaeal phylum Thaumarchaeota within Antarctic surface seawaters during the wintertime. Thaumarchaeota (earlier identified as Marine Group I.1a) composes 10-30% of the bacterioplankton during the

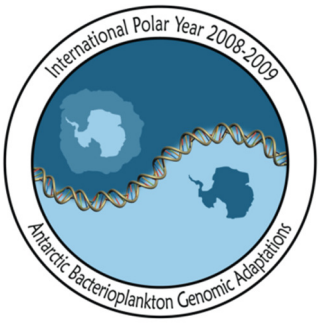
winter in Antarctic surface seawaters playing a noteworthy role in carbon fixation coupled to ammonia oxidation. By comparative genomic analyses, we found that the Antarctic Group I.1a genome fragments represent a unique low diversity Group I.1a cluster affiliated with a "*Ca.* Nitrosopumilus" that has not yet been cultivated. Antarctic Group I.1a exhibited highly conserved genes in a rearranged genomic structure when compared to the genome of *Nitrosopumilus maritimus* SCM1indicating a high frequency of recombination (30% of the Antarctic Group I.1a open reading frames). Overall, our results provided insights into the genomic variability of a thaumarchaeal Antarctic assemblage indicating very low diversity that motivates the importance of acquiring the Antarctic strain in pure culture for further analysis of its physiological capacities and evolutionary history.

# Acknowledgments











*They are always around you*

*To give you the support that you need*

*So you can do the best of your best.*

# Content

## Chapter I



## Chapter II

## Chapter III



## Chapter IV

## Chapter V

# List of Tables

# List of Figures

# Chapter I

## Introduction: Life under subzero conditions and study sites

Cold environments, below 15°C, are the most abundant physicochemical conditions on Earth and beyond. For instance, 10% of the Earth's surface is covered in ice, 20-25% of the terrestrial surface is permafrost, and 90% of the volume of the oceans is at or below 5°C (Rodrigues and Tiedge, 2008). Six of our solar system's planets and their moons, as well as most dwarf planets, are permanently cold. Also, 33% of detected exoplanets are orbiting their parent stars beyond the habitable zone, where temperatures are thought to be too cold for life to exist. Recent studies have found evidence of water on Mars and icy moons in the outer solar system (Maltagliati *et al*., 2011). Mars hosts a permafrost layer, thick water ice deposits at the poles, and possibly experiences surface water flow in the form of brines (Davila *et al.*, 2010; Glamoclija *et al*., 2010). The

probability of the existence of liquid water or ice in polar sheets or within permafrost soils on these celestial bodies suggests the possibility for finding life outside our planet.

Microorganisms that survive or thrive in cold environments are the most abundant extremophiles (organisms that live in extreme conditions) on Earth in terms of biomass, diversity, and distribution (D'Amico *et al.*, 2006). Organisms that grow well at temperatures ranging from 20°C to below 0°C with an optimum growth rate at 15°C or below are known as psychrophiles (Morita, 1975). Psychrophiles have developed cellular and metabolic strategies relying on their genome content and gene expression to cope with the following challenges dictated by freezing temperatures: ice crystal formation, low water activity, limited nutrient availability, osmotic stress, increase in structural rigidity of cellular components, and decrease in enzymatic catalytic rates. Some of their cellular and metabolic strategies include increasing the expression of helper-proteins (which assist in the folding of DNA, RNA, and other proteins), accumulation of compatible solutes, producing ice-binding proteins and cryoprotectans, reducing cell size, preserving cell integrity inside capsules, surrounding cells with exopolysaccharides, and aggregating cells (Thieringer, 1998; Hébraud and Potier, 1999; Yu and Owttrim, 2000; Srinivas and Ray, 2006).

Life under cold and subzero conditions can exist in multiple metabolic states. For example, microbial cells inhabiting icy environments may be functional and capable of replication (Bakermans, *et al.,* 2003) or, in a minimum activity state, cells can be viable but unable to replicate (Christner, 2002). Microbial cells also may be found in a dormant state where nucleic acid information is protected under layers of distinct coats, such as

polysaccharides and dipicolinic acid. From the dormant state, cellular activity may or may not be recovered in a more conducive environmental condition (Soina *et al.,* 2004). Regardless of the cellular state, all cells have a key factor in common: preservation of their nucleic acid information.

The main methodologies used in this investigation were based on prokaryotic nucleic acid information. 16S rRNA gene amplification techniques and comparative genomic analyses were used to study the microbial assemblage in two distinct Antarctic systems: Lake Vida in the McMurdo Dry Valleys and coastal surface seawater from the west Antarctic Peninsula during winter (Figure 1.1).

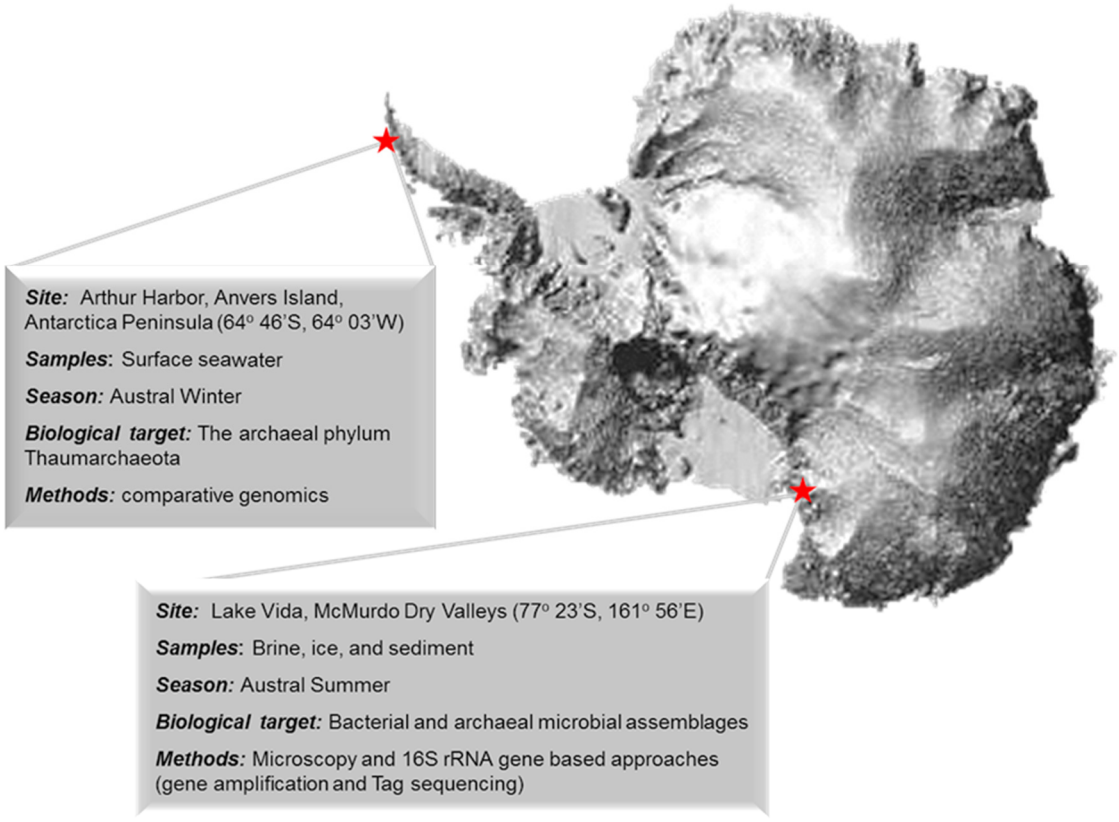



**Figure 1.1.** Study sites and metadata.

The search for life in our solar system is largely focused on ice-covered habitats similar to those found on Earth. Consequently, investigation of life in cold and icy environments is extremely important for the exploration of life beyond our planet. Antarctica, which may appear unique when compared to other ecosystems on Earth, presents physical features that may be similar to those found on icy worlds such as Europa, Enceladus, and Mars. Research into microbial life in Antarctic systems yields important knowledge regarding the persistence of life in extremely low temperatures elsewhere. Therefore, Antarctic environments with their perpetually low temperatures may be the terrestrial environment most similar to extraterrestrial environments that presently are habitable or were potentially habitable at some point in the past.

### 1.1. Lake Vida, McMurdo Dry Valleys, Antarctica

Antarctica is a distinct world in our planet. The continent has more than 100 subglacial lakes, perennially ice-covered lakes, and harbors the coldest and driest polar desert on the planet: the McMurdo Dry Valleys. The McMurdo Dry Valleys, located between 77°20'-77°45' S and 161°00'-164°00' E, is the largest ice-free land area on the continent. High winds, low temperatures (ranging from -30 to -14.8 °C along the valley bottoms), and low mean annual precipitation (<10 cm) make the McMurdo Dry Valleys an extreme environment for life and an excellent extraterrestrial analog (Doran *et al*., 2003). Even in these extreme conditions, liquid water persists and microbial life can be found in the Dry Valleys inhabiting environments such as ice, snow, lakes, glaciers, ephemeral streams, soils, and rocks.

The McMurdo Dry Valleys comprise three main ice-free valleys: Taylor Valley, Wright Valley, and Victoria Valley (Figure 1.2). Lakes or briny ponds located in all three valleys are surrounded by rocks, arid soils, and glaciers. Most of the dry valley lakes are perennially covered with three to six meters of ice. These ice covers allow limited atmospheric exchange, limited light penetration, and prevent physical mixing of the water in the lake, leading to long-term vertical stratification of the water column (Wharton *et al*., 1993). Unlike other lakes in the region, Lake Vida – located in Victoria Valley – does not present the same hydrological structure of other ice-covered lakes. The 6.8 $km^2$ lake has a thick ice cover (>27 meters) which contains a brine reservoir trapped within the ice below 16 m and has well sorted sediment layers up to 30 cm long banding salty ice below 20 m. Therefore, Lake Vida provides an opportunity to study a model of what environmental conditions might exist in other icy worlds (isolated, cold, and salty) in three different environments including brine, ice, and frozen sediment.

Up until now, Lake Vida microbial assemblage composition have been investigated in the upper portion of the lake, from 4.8 to 15.9 m depth (Mosier *et al*., 2007) and in the encapsulated brine (Murray *et al.,* 2012). The microbial assemblages within the upper portion of the ice cover are distinct and dependent on depth. The phylum *Actinobacteria* is dominant (42%) nearest the ice surface (Mosier *et al*., 2007). The deepest ice core segment (15.9 m depth) previously collected and analyzed is dominated by the class *Gammaproteobacteria* (52%) (Mosier *et al*., 2007). Even with extreme conditions of freezing temperatures and high salinity, Lake Vida's brine is inhabited by a heterogeneous microbial assemblage. The brine microbial assemblage is composed of

two cell size classes: cells >0.2 to 1.5 µm in diameter and cells ~0.2 µm in diameter. These two cell-size populations are composed of at least eight bacterial phyla including *Proteobacteria* (*Gammaproteobacteria*, *Betaproteobacteria, Epsilonproteobacteria,* and *Deltaproteobacteria*), *Lentisphaerae*, *Firmicutes*, *Spirochaetea*, *Bacteroidetes*, *Actinobacteria*, *Verrucomicrobia*, and candidate Division TM7 (Murray *et al*., 2012). *Bacteroidetes* is one of the most abundant groups detected in the upper portion of the lake ice, as well as in the brine.

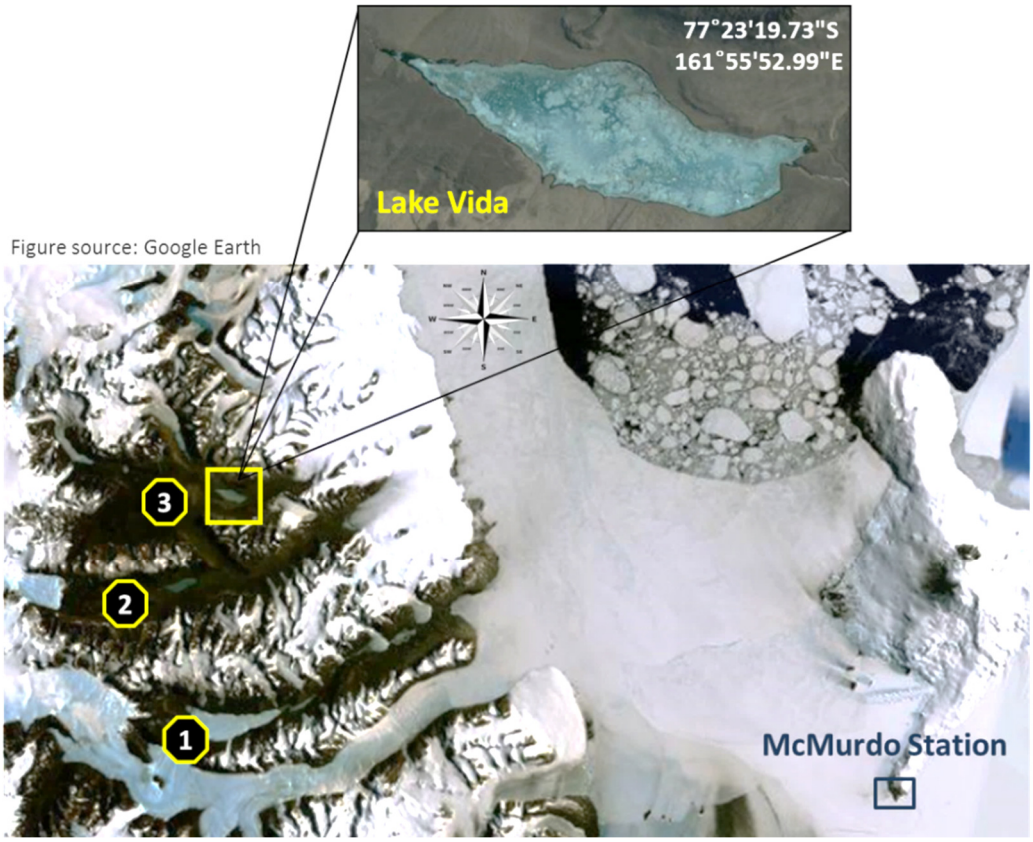



**Figure 1.2.** McMurdo Dry Valleys, Antarctica. The three major valleys that constitute the Dry Valleys are Taylor Valley (1), Wright Valley (2), and Victoria Valley (3). Lake Vida (yellow rectangle) is located in Victoria Valley, the northern-most area of the Dry Valleys. Figure source: Google Earth.

The last expedition to Lake Vida (November, 2010) recovered 27 m of ice, intercalated with four sorted sediment layers below 21 m, and brine from the core borehole. Characterization of life in the recently collected brine, ice, and frozen sediment is the focus of chapters II and III in this dissertation. Chapter II describes the nature of the abundant "cell-like particles" with ~0.2 μm in diameter found in the lake's brine (Murray *et al*., 2012), and chapter III presents findings of the microbial assemblage which inhabits the deep ice – from 18 to 27 m depth – and sediment layers of the lake.

## 1.2. Antarctic Peninsula coastal surface seawater and its Thaumarchaeota Group I.1a assemblage

Archaea is one of the three domains of life proposed by Carl Woese and colleagues in 1977 (Figure 1.3; Woese and Fox, 1977). Until 2008, the domain Archaea was constituted by four phyla: Euryarchaeota, Crenarchaeota, Nanoarchaeota, and Korarchaeota. The Korarchaeota phylum is represented only by environmental DNA sequences detected in high temperature habitats (Barns *et al*., 1996). The Nanoarchaeota phylum has only one known member, *Nanoarchaeum equitans,* which is a symbiotic hyperthermophile with a genome of 0.45 megabases in size (Huber *et al.*, 2002). The Euryarchaeota phylum is comprised of a wide variety of morphologically, metabolically, and physiologically diverse species – including various methanogens, extreme halophiles, thermoacidophiles, and sulfate-reducing microbes. The Crenarchaeota phylum was long thought to consist solely of extreme thermophiles and acidophiles. In 1992, marine

ecologists detected for the first time members of the phylum Crenarchaeota in a non-extreme environment: the ocean (DeLong, 1992; Fuhrman *et al*., 1992). Analyses of a concatenated data set of 53 ribosomal proteins from the first available genome of a mesophilic Crenarchaeota, *Cenarchaeum symbiosum,* showed that *C. symbiosum* branched off before the separation of Crenarchaeota and Euryarchaeota in a phylogenetic tree (Brochier-Armanet *et al*., 2008). Based on this phylogenetic analyses, the Marine Group I Crenarchaeota was reclassified as an new Archaea phylum: the Thaumarchaeota (Brochier-Armanet *et al*., 2008; Thaumarchaeota Group I.1a, hereafter).

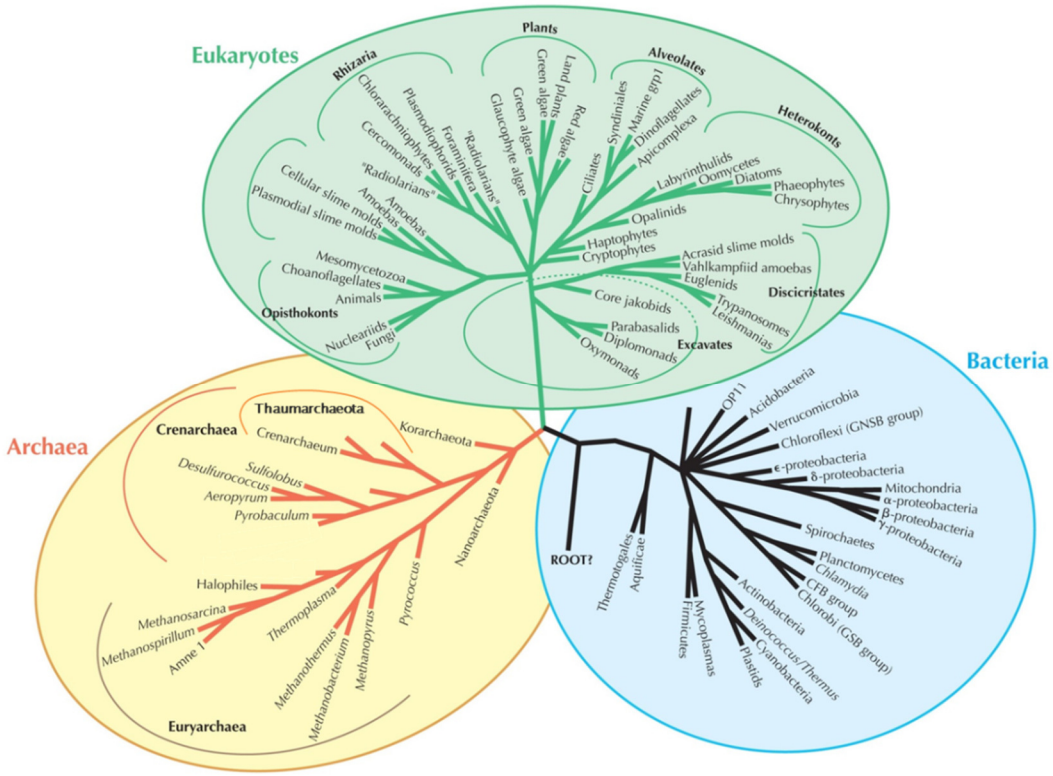



**Figure 1.3.** Representation of the tree of life. The three domains system – which classifies life into Archaea, Bacteria, and Eukaryotes – was based in the evolutionary relationship of a cellular molecule found in all organisms: the small-subunit of the ribosomal RNA (Woese and Fox, 1977). The three domains are shown in different colors. Figure adapted from Barton *et al.,* 2007.

In marine environments, Thaumarchaeota Group I.1a (Group I.1a) accounts for up to 40% of the bacterioplankton in deep ocean waters (Karner *et al*., 2001) and is the most abundant inhabitant in polar surface seawater during winter (Murray *et al*., 1998; Grzymski *et al*., 2012). In 2005, cultivation of a mesophilic crenarchaeon (belonging to the new phylum Thaumarchaeota Group I.1a) was first reported: *Nitrosopumilus maritimus* SCM1 (Könneke *et al*., 2005). *N. maritimus* SCM1 was not just the first Thaumarchaeota cultivated; it also was the first archaea known to be able to chemolithoautotrophically oxidize ammonium aerobically to nitrite (Könneke *et al*., 2005).

Comparative genomic analysis of *N. maritimus* SCM1 to environmental genome fragments showed a high similarity of globally distributed Group I.1a to the genome of *N. maritimus* SCM1. Therefore, Group I.1a is not only the dominant bacterioplankton assemblage in winter Antarctic coastal surface seawater; it also is representative of the most plentiful archaeal group known on Earth (Walker *et al*., 2010).

A study comparing bacterioplankton metagenomic data from coastal surface seawater from the west Antarctic Peninsula during austral summer and austral winter has reported a shift in microbial community composition and genome-encoded capacity between the two seasons (Grzymski *et al.*, 2012). The winter Antarctic bacterioplankton assemblage was represented by 20% of Group I.1a closely affiliated with the Antarctic archaeal fosmid 74A4 (Béjà *et al*., 2002) and to *N. maritimus* SCM1 (Könneke *et al*., 2005) and was identified as the major player for carbon fixation coupled to ammonia oxidation (Grzymski *et al.*, 2012).

Motivated by detection of the abundant environmental genome fragments affiliated with Group I.1a in the Antarctic winter surface seawater (Grzymski *et al.*, 2012), chapter IV presents a comparative genomic study between the Antarctic Group I.1a-related genome fragments to the genome from the five Thaumarchaeota Group I.1a genomes. We hypothesize that the Antarctic Group I.1a represents a Thaumarchaeota genome not yet represented by the genome references available.

### 1.3. Objectives

The focus of this dissertation is to explore microbial life of Antarctic environments from a standpoint of microbial assemblage characterization by combining microscopy observations, cultivation-based methods, molecular phylogenetic analysis, and comparative genomics.

The specific objectives of this dissertation are:

i. Characterize and identify the nature of the abundant cell-like particles with ~0.2 μm in diameter existent in Lake Vida's icy brine.
ii. Characterize the composition and distribution of the microbial assemblage along the lower portion of Lake Vida's thick ice cover – from 18 to 27 m depth.
iii. Characterize the Thaumarchaeota Group I.1a assemblage that inhabits Antarctic surface seawater during winter seasons by applying comparative environmental genomics analyses.

## 1.4. References


Bakermans, C., Tsapin, A. I., Souza-Egipsy, V., Gilichinsky, D. A, & Nealson, K. H. (2003). Reproduction and metabolism at -10ºC of bacteria isolated from Siberian permafrost. *Environmental Microbiology, 5*(4), 321–326.

Barns, S. M., Delwiche, C. F., Palmer, J. D., & Pace, N. R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proceedings of the National Academy of Sciences, 93*(17), 9188–9193.

Barton, N. H., Briggs, D. E. G., Eisen, J. A., Goldstein, D. B. & Patel, N.H. (2007). *Evolution*. Cold Spring Harbor Laboratory Press, 833 pp.

Béjà, O., Koonin, E. V., Aravind, L., Taylor, L. T., Seitz, H., Stein, J. L., Bensen, D. C., et al. (2002). Comparative genomic analysis of archaeal genotypic variants of a single population in two different oceanic provinces. *Applied and Environmental Microbiology*, *68*(1), 335–345.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., & Forterre, P. (2008). Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Reviews in Microbiology*, *6*(3), 245–252.

Christner, B. C. (2002). Incorporation of DNA and protein precursors into macromolecules by bacteria at -15ºC. *Applied and Environmental Microbiology, 68*(12), 6435–6438.

D'Amico, S., Collins, T., Marx, J. C., Feller, G., & Gerday, C. (2006). Psychrophilic microorganisms: challenges for life. *EMBO Reports*, *7*(4), 385–389.

Davila, A. F., Skidmore, M., Fairén, A. G., Cockell, C., & Schulze-Makuch, D. (2010). New priorities in the robotic exploration of Mars: the case for in situ search for extant life. *Astrobiology, 10*(7), 705–710.

DeLong, E. F. (1992). Archaea in coastal marine environments. *Proceedings of the National Academy of Sciences*, *89*(12), 5685–5689.

DeLong, E. F. (2003). Oceans of Archaea. *American Society for Microbiology News*, *69*(10), 503–511.

Doran, P. T., Fritsen, C. H., McKay, C. P., Priscu, J. C., & Adams, E. E. (2003). Formation and character of an ancient 19-m ice cover and underlying trapped brine in an "ice-sealed" east Antarctic lake. *Proceedings of the National Academy of Sciences,100*(1), 26–31.

Fuhrman, J. A., McCallum, K., & Davis, A. A. (1992). Novel major archaebacterial group from marine plankton. *Nature*, *356*(6365), 148–149.

Glamoclija, M., Marinangeli, L., & Komatsu, G. (2010). Harmakhis Vallis Source Region, Mars: insights into the recent geothermal history based on geological mapping. *Planetary and Space Science*, *59*(11-12), 1179–1194.

Grzymski, J. J., Riesenfeld, C. S., Williams, T. J., Dussaq, A. M., Ducklow, H., Erickson, M., Cavicchioli, R., et al. (2012). A metagenomic assessment of winter and summer bacterioplankton from Antarctica Peninsula coastal surface waters. *The ISME Journal*, *6*(10), 1901–1915.

Hébraud, M., & Potier, P. (1999). Cold shock response and low temperature adaptation in psychrotrophic bacteria. *Journal of Molecular Microbiology and Biotechnology*, *1*(2), 211–219.

Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., & Stetter, K. O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature*, *417*(6884), 63–67.

Karner, M. B., DeLong, E. F., & Karl, D. M. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature*, *409*(6819), 507–510.

Könneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., & Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*, *437*(7058), 543–546.

Maltagliati, L., Montmessin, F., Fedorova, A, Korablev, O., Forget, F., & Bertaux, J. L. (2011). Evidence of water vapor in excess of saturation in the atmosphere of Mars. *Science*, *333*(6051), 1868–1871.

Morita, R. Y. (1975). Psychrophilic Bacteria. *Bacterilogical Reviews, 39*(2), 144–167.

Mosier, A. C., Murray, A. E., & Fritsen, C. H. (2007). Microbiota within the perennial ice cover of Lake Vida, Antarctica. *Fems Micriobiology Ecology*, *59*(2), 274–288.

Murray, A. E., Kenig, F., Fritsen, C. H., McKay, C. P., Cawley, K. M., Edwards, R., Kuhn, E., et al. (2012). Microbial life at -13 °C in the brine of an ice-sealed Antarctic lake. *Proceedings of the National Academy of Sciences,109*(50), 20626–20631.

Murray, A. E., Preston, C. M., Massana, R., Taylor, L. T., Blakis, A, Wu, K., & DeLong, E. F. (1998). Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. *Applied and Environmental Microbiology*, *64*(7), 2585–2595.

Preston, C. M., Wu, K. Y., Molinski, T. F., & DeLong, E. F. (1996). A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proceedings of the National Academy of Sciences,93*(13), 6241–6246.

Rodrigues, D. F., & Tiedje, J. M. (2008). Coping with our cold planet. *Applied and Environmental Microbiology*, *74*(6), 1677–1686.

Soina, V. S., Mulyukin, A. L., Demkina, E. V., Vorobyova, E. A., & El-Registan, G. I. (2004). The structure of resting bacterial populations in soil and subsoil permafrost. *Astrobiology, 4*(3), 345-358.

Srinivas, U. K., & Ray, M. K. (2006). Cold-stress response of low temperature adapted bacteria. *Molecular Biology*, *661*(2), 1–23.

Thieringer, H. A., Jones, P.G., & Inouye, M. (1998). Cold shock and adaptation. *BioEssays, 20*(1), 49–57.

Walker, C. B., De La Torre, J. R., Klotz, M. G., Urakawa, H., Pinel, N., Arp, D. J., Brochier-Armanet, C., et al. (2010). *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proceedings of the National Academy of Sciences,107*(19), 8818–8823.

Wharton, R. A., McKay, C. P., Clow, G. D., & Andersen, D. T. (1993) Perennial ice covers and their influence on Antarctica lake ecosystems. *In* Physical and Biogeochemical Processes in Antarctic Lakes. Friedmann, E.I., and Green, W.J. (Eds). Washington, D.C.: American Geophysical Union, 53–70.

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences, 74*(11), 5088-5090.

Yu, E., & Owttrim, G. W. (2000). Characterization of the cold stress-induced cyanobacterial DEAD-box protein CrhC as an RNA helicase. *Nucleic Acids Research*, *28*(20), 3926–3934.

# Chapter II

## The ultra-small microbial assemblage of Lake Vida encapsulated brine


### Abstract

Lake Vida's anoxic brine, that permeates the lake's perennial ice below 16 m, contains abundant very small (≤ 0.2 μm) particles mixed with a less abundant population of larger particles in the length size range from > 0.2 to 1.5 μm. DNA fluorescence staining, electron microscopy (EM) observations and elemental analysis indicated that these ultra-small particles are cells. Genomic DNA with high molecular weight was extracted from the ultra-small cells and a continuous electron-dense layer surrounding a less electron-dense region was observed by EM. These features and structures are representative of intact coccoid bacterial cells with diameters of 0.192 ± 0.065 μm. Morphological variability included the presence of diplococci cells, the presence of iron-rich capsular structures surrounding cell external surfaces, and cell aggregates. Smaller unidentified

nanoparticles ranging from 0.020 to 0.140 μm in diameter were observed among the brine cells and ultra-small cells. A 16S rRNA gene clone library from the brine 0.1 – 0.2 μm size fraction resulted in the detection of a range of Bacteria-affiliated sequences that were dominated by those most closely affiliated with the *Proteobacteria Pseudoalteromonas* and *Herbaspirillum*. Cultivation efforts of the 0.1 – 0.2 μm size fraction led to the isolation of *Actinobacteria* affiliated cells *Microbacterium* sp. and *Kocuria* sp. Based on phylogenetic relatedness and microscopic observations, we concluded that the ultra-small cells in Lake Vida brine are ultramicrocells that are likely in a reduced-size state as a result of environmental stress or life-cycle related conditions.

## 2.1. Introduction

One of the coldest stable liquid cryo-environments known on Earth is the recently described Lake Vida brine (Doran *et al.*, 2003; Murray *et al.*, 2012). Lake Vida, situated in Victoria Valley, East Antarctica, is one of the largest lakes in the McMurdo Dry Valleys. Most of the dry valley lakes are perennially covered with three to six meters of ice which allows for limited atmospheric exchange and light penetration. However, the thick (27+ m) perennial ice of Lake Vida blocks sunlight below 12 m and creates a sealed brine reservoir below 16 m depth. This brine has not exchanged gases with the atmosphere for an estimated 2,800 years (Doran *et al.*, 2003). The interstitial, anoxic brine (with salinity range from 188 to 210) reveals a unique ecosystem with a temperature of -13.4°C, pH of 6.2, high solute concentrations [e.g. $NH_4^+$ (3.8 mM), Fe (>

300 μM), $NO_3^-$ (905 μM), and high levels of dissolved organic carbon (580 mg $L^{-1}$) (Murray *et al.*, 2012).

Lake Vida brine assemblage is dominated by Bacteria (~$10^7$ cells $mL^{-1}$; Murray *et al.*, 2012). Eight bacterial phyla were identified from a 16S rRNA gene clone library from brine-cells collected by filtration on 0.2 μm pore-sized filters: *Proteobacteria* (classes *Gamma*, *Delta*, and *Epsilon*), *Lentisphaerae*, *Firmicutes*, *Spirochaetea*, *Bacteroidetes*, *Actinobacteria*, *Verrucomicrobia*, and Candidate Division TM7 (Murray *et al.*, 2012). According to the authors, the brine microbial assemblage is composed of two cell-size classes. One class consists of typical aquatic bacteria with sizes of 0.5–1.5 μm. The second cell-size class consists of ultra-small "cell-like particles" of ~0.2 μm in diameter, with 100x the abundance of the larger cells.

Ultra-small microorganisms (organisms small enough to pass through 0.2 μm pore-sized filters) have been observed inhabiting different environments as parasites, symbionts, and free-living cells. Phylogenetic, morphological, and (meta)genomic characterizations of cultivated and uncultivated ultra-small microbial cells from glacial ice (Karl, 1999; Priscu, 1999; Miteva *et al.*, 2005), seawater (Haller *et al.*, 2000; Rappé *et al.*, 2002; Giovannoni *et al.*, 2005), permafrost (Dmitriev *et al.*, 2001; 2004; 2008), soil (Bae *et al.*, 1972; Iizuka *et al.*, 1998; Duda *et al.*, 2009), subsurface and deep-ground water (Miyoshi *et al.*, 2005), freshwater (Hahn *et al.*, 2003), acid mine drainage (Comolli *et al.*, 2009; Baker *et al.*, 2010), deep-sea hydrothermal vents (Huber *et al.*, 2002; Naganuma *et al.*, 2007; Nakai *et al.*, 2011), and insect guts (Geissinger *et al.*, 2009) show how prevalent these tiny microorganisms are on our planet.

Different terminologies are now used to classify ultra-small microorganisms (reviewed by Cavicchioli and Ostrowski, 2003) and two of them are ultramicrobacteria and ultramicrocells (Figure 2.1). The differences between ultramicrobacteria and ultramicrocells rely on their distinct evolutionary paths. Ultramicrobacteria (Torrella and Morita, 1981) is a term used to describe bacterial species with cell volumes less than 0.1 $\mu m^3$ which maintain their cell size and volume regardless of their *in vitro* growth condition (Figure 2.1A). The term ultramicrocells is used to denote cells that have a reduction in size and volume as a consequence of unfavorable environmental factors or cells that change morphology from rod to cocci during the life cycle (Figure 2.1B and 2.1C). Ultramicrocells are usually starved, dormant, under osmotic pressure, or stressed forms of "normal-sized" cells that increase in size and volume when cultivated (Hahn *et al.* 2003; Rutz and Kieft, 2004). A cell's size coordinates cytoplasmic space to harbor cellular genetic elements and ribosomes required to confer a minimum of cellular activity. Examples of strategies to maintain cell integrity and function observed in microbial cells from extremely cold and salty environments include reduction of the cell size and morphological transition to a cyst-like resting cell (Litchfield, 1998; Gilichinsky, 2001; Ponder *et al.*, 2008).

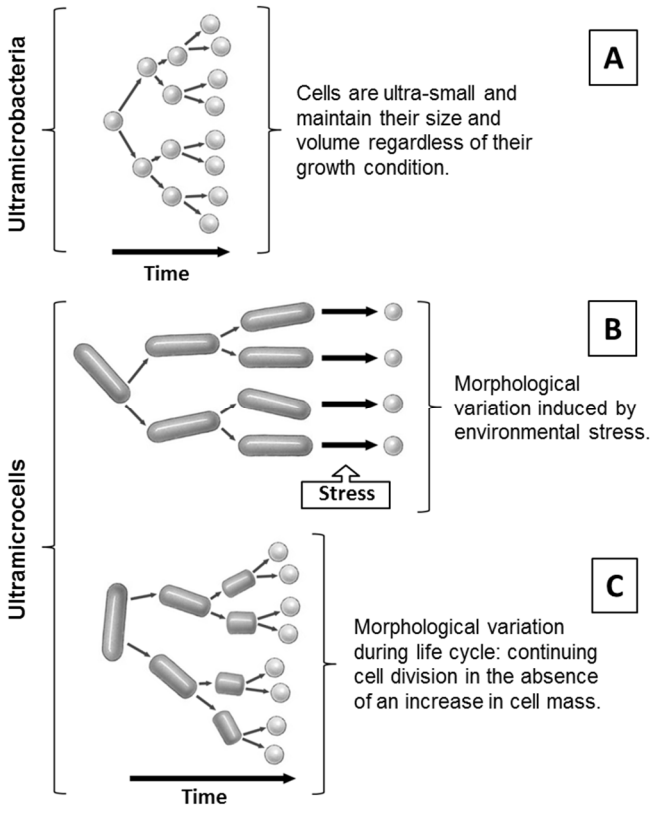



**Figure 2.1.** Diagram demonstrating the distinctions between ultramicrobacteria and ultramicrocells**.** (A) Ultramicorbacteria; (B) Ultramicrocells formed in response to unfavorable environmental conditions, such as hyperosmotic stress; (C) Ultramicrocells formed as consequence of aging: rod-to-sphere transition during late-log and stationary phases of bacteria life cycle.

In this study, microscopy imaging, elemental analyses, DNA-based molecular approaches, and culturing were employed to characterize the ~0.2 µm "cell-like particles" from Lake Vida brine. The results led to the conclusion that the abundant ~0.2 µm cell-like particles are Bacteria related ultramicrocells. Analyses of the 16S rRNA gene indicated that most of the Lake Vida cell fraction ≤ 0.2 µm, are associated with the genera *Pseudoalteromonas* and *Herbaspirillum*. Culturing of the ≤ 0.2 µm brine fraction

led to the isolation of *Actinobacteria*-affiliated strains. The growth state and means for survival of Lake Vida ultramicrocells in the natural brine environment remain under investigation.

## 2.2. Materials and methods

### 2.2.1. Brine sampling and cell size fractionation

Samples were collected in November 2010 from a ~20.4 m hole drilled near the middle of Lake Vida, at 77°23'19.73"S, 161°55'52.99"E. To reduce risk of contamination, instruments used for entry and sampling approaches were cleaned with a dichloromethane-methanol mix (1:1, v:v) (Doran *et al.*, 2008). After drilling, brine infiltrated the borehole until it reached a depth of ~11 m below the ice surface (Figure 2.2). The hole was drained prior to sample collection to allow for flushing with fresh brine. Samples were collected at a depth of 18.4 m using a submersible pump (Proactive S.S. Monsoon, Hamilton, NJ, USA) and polytetrafluoroethylene (PTFE) tubing that guided the brine to two separate PTFE lines. One line was used to collect the brine into two and ten liter HDPE bottles under a nitrogen atmosphere (Figure 2.2, left). The second line (Figure 2.2, right) guided the brine into an anaerobic glove bag (AtmosBag, Aldrich, St. Louis, MO, USA) where samples were collected for microscopy and geochemical analyses. Anoxic conditions were applied during all processes of sampling, filtering, and fixing the brine. Collected samples were stored at -10°C until processed.

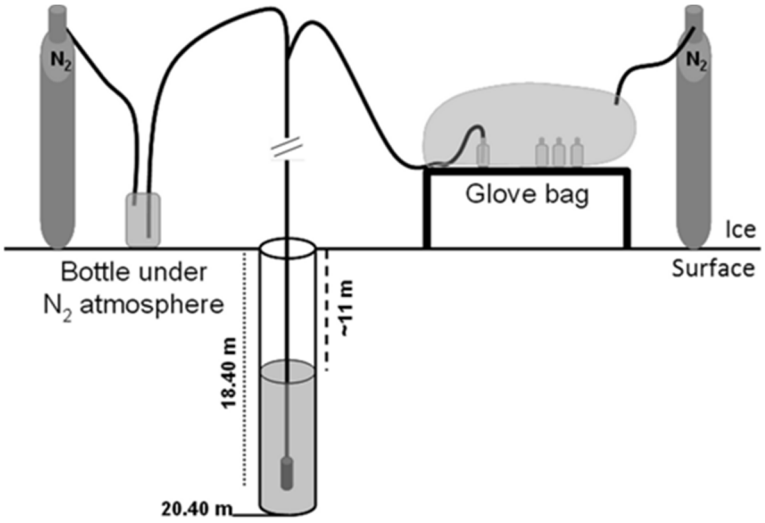



**Figure 2.2.** Diagram of Lake Vida brine sampling system**.** A 3/8" ID PTFE tubing connected to a battery-powered submersible pump was used to bring the brine to the surface where it was directed to different sampling lines. One line was used for collecting the brine in bottles under nitrogen atmosphere (left), and the other line carried the brine into an anaerobic glove bag (right). The depth of the sampling hole was approximately 20.4 m; brine infiltrated the borehole until it reached the depth of approximately 11 m; the pump head was positioned at 18.40 m.

For size fractionation, the brine's microbial cells ≥ 0.2 µm in diameter were collected on 0.22 µm pore-sized Sterivex filters (Millipore, Boston, MA, USA) under anoxic conditions at -10°C, and stored in a sucrose lysis buffer (Massana *et al.*, 1997) at -80°C for further DNA extraction. The filtrate, composed of cells that passed through the 0.22 µm pore-sized filters, was collected in one liter glass bottles under anaerobic conditions and incubated for 30 days at -10°C. After incubation, the brine ultra-small microbial assemblage (LVBrUMA) was filtered again through 0.22 µm pore-sized Sterivex filters (polyethersulfone membrane). Cells passing the 0.22 µm filters were collected for further DNA extraction and culturing. For DNA extraction, cells were

collected on 0.1 µm pore-sized Supor®-100 filters and stored in a sucrose lysis buffer at -80°C. The filtrate that passed through the 0.22 µm pore-sized Sterivex filters was preserved in 20% glycerol (v/v) and stored at -80°C for culturing.

### 2.2.2. Microscopy (Confocal, SEM, and STEM) and EDS analysis

Confocal microscopy using an Olympus FluoViewTM 1000 Confocal Microscope was used to determine cell abundance in Lake Vida brine. Brine samples were fixed with 3.7% anoxic formalin (v/v) or 0.5% anoxic glutaraldehyde (v/v) under anoxic conditions during sampling. Fixed cells in the brine were stained with SYBR Gold (Invitrogen, Carlsbad, CA, USA) for 10 minutes and transferred to 0.22 µm pore-sized black polycarbonate filters (Millipore, Inc. Boston, MA, USA) under a nitrogen atmosphere.

Electron microscopy was used to assess the morphology of the brine cells. For scanning electron microscopy (SEM), brine aliquots (0.2 - 1 mL) were fixed with 0.5% anoxic glutaraldehyde (v/v final concentration) under anoxic conditions and placed on 0.2 µm pore-sized polycarbonate and 0.02 µm pore-sized Anodisc filters (Whatman, Piscataway, NJ, USA). Once on the filters, the cells were dehydrated by an ethanol series (30, 50, 70, 90, and 100%) under ambient conditions, dried in air, and coated with 1-3 nm of iridium or platinum to prevent charging during microscopy observation. For image acquisition by scanning transmission electron microscopy (STEM), two protocols for cells preparation were developed. In one protocol, cells from a drop (20 µL) of anoxic

fixed brine were transferred to a LuxFilmTM 2 mm open area grid on a copper support, and subsequently negative stained with phosphotungstic acid (PTA, $H_3PW_{12}O_{40}$) at pH 0.4 for 30 seconds. In the second protocol, before preparing the cells for STEM observations, anoxic fixed brine was treated with 40 mM EDTA (ethylenediaminetetraacetic acid, sample ratio by volume of 9:1) pH 8.0 for 10 minutes at room temperature. The cells from the EDTA treated brine were then transferred to a carbon type B 300 mesh copper grid, negative stained with uranyl acetate (UA, $UO_2(CH_3COO)_2 \cdot 2H_2O$) at pH 4-5 for 5 minutes, and washed in deionized water. All buffers, solutions, and stains for EM analyses were filtered through 0.02 µm pore-sized filters immediately before use. Electron microscopy was performed with a Carl Zeiss Ultra55 field emission instrument. SEM images were acquired with an Everhart-Thornley or annular secondary electron detector at working distances of 4-5 mm and 2.0 keV accelerating voltage. Beam energies of 20-30 keV were used to acquire STEM images.

Energy dispersive x-ray spectroscopy (EDS) of brine sample was analyzed by an Oxford Instruments INCA 350 system. An XMax 80 $mm^2$ silicon drift detector was used to determine the elemental composition of the Lake Vida brine cells supported on filters and grids. EDS signal counts were collected for 240 seconds (live time) with an accelerating voltage of 4-5 keV and an analytical working distance of 8.5 mm.

Size measurement of cells and unidentified particles was performed using the program ImageJ (http://imagej.nih.gov/ij). The diameter from a total of 465 cells/particles from SEM and STEM micrographs were measured.

### 2.2.3. Grazing angle X-ray diffraction

Grazing angle X-ray diffraction patterns were acquired with a Bruker D8 Advance diffractometer using Cu(Kα) radiation (λ = 1.5405 Å) and a NaI scintillation detector. A 0.2 μm polycarbonate filter with a tan-colored mat of captured cells and brine material in suspension was used for the measurements. The diffractometer was equipped with parallel beam optics (Goebel mirror) and thin film receiving slit accessories to enable measurement at an angle of 1°. Diffraction patterns were acquired between 10 – 80°, 2θ with a data spacing of 0.02° at three seconds per step.

### 2.2.4. Cultivation of heterotrophic bacteria from brine fractions

Cells from the glycerol stocks of 0.22 μm filtered brine and from the unfiltered brine were inoculated under aerobic growth conditions at three different temperatures: 2°C, 10°C and 21°C on five different solid media: (1) R2A, (2) R2A + 5% NaCl, (3) Marine Agar + 5% NaCl, (4) Casein agar, and (5) Malt extract agar (for fungi). After 120 days, 27 colony-forming units (CFU) were counted on the plates inoculated with the 0.22 μm filtered brine, from which 15 CFU were selected based on distinct morphological characteristics, for isolation and identification by 16S rRNA gene phylogeny. After 110 days, a total of 335 CFU were counted from plating the unfiltered brine on agar plates, and 33 CFU were selected based on distinct morphological characteristics, for isolation and identification by 16S rRNA gene phylogeny.

### 2.2.5. Nucleic acid extractions, 16S rRNA gene and phylogenetic analyses

Genomic DNA from both cell-size assemblages, cells ≥ 0.2 µm and cells with sizes between 0.1 and 0.22 µm (filtrate), were extracted from 0.22 µm pore-sized Sterivex filters and 0.1 µm pore-sized Supor®-100 filters, respectively, according to Massana *et al.* (1997). The genomic DNA extracted was quantified using the Quant-iT PicoGreen dsDNA Assay kit (Invitrogen, Carlsbad, CA, USA).

Total RNA was extracted from cells collected on 0.22 µm pore-sized filters using a modified Totally RNA protocol (Applied Biosystems/Ambion, Austin, TX, USA) and RNeasy Cleanup kit (Qiagen, Valencia, CA, USA). In brief, for RNA extraction, acid phenol:chloroform was used as an initial separation step along with an additional phenol:chloroform extraction, followed by two chloroform extractions. Also, linear polyacrylimide was added to the first isopropanol precipitation step. Two DNase I treatments (Applied Biosystems/Ambion, Austin, TX, USA) were performed to ensure no genomic DNA contamination. As a final treatment, the extracted RNA was purified using an RNeasy kit (Qiagen, Valencia, CA, USA). Extracted RNA was quantified using the Quant-iT RiboGreen RNA Assay kit (Invitrogen, Carlsbad, CA, USA). Complementary DNA (cDNA) was generated by priming total RNA with random hexamers and reverse transcribed by reverse transcriptase using a SuperScript III kit (Invitrogen, Carlsbad, CA, USA). No RNA was extracted from the 0.1-0.22 µm cell-size fraction.

Molecular profiles of 16S rRNA gene fragments were conducted by Denaturing Gradient Gel Electrophoresis (DGGE), where (1) complementary DNA (cDNA) from

cells ≥ 0.2 µm, (2) genomic DNA from cells ≥ 0.2 µm and (3) genomic DNA from the 0.1-0.22 µm cell-size fraction were amplified with the primers targeting the V3 variable region of the 16S rRNA gene: BactGC358f and 517r (Murray *et al.*, 1996). The PCR reaction for DGGE consisted of initial denaturation at 94°C for 3 minutes, annealing at 65°C followed by 10 cycles of touchdown PCR decreasing the annealing temperature of the reaction 1°C per cycle. The touchdown PCR was tailed to 18 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds, and elongation at 72°C for 30 seconds. A final elongation step of 72°C for 7 minutes was added to the reaction after 28 amplification cycles. DGGE was performed according to Mosier *et al.* (2007) and ran for a total of 1000 volt-hours. A linear gradient of 30 to 65% denaturants (100% denaturants corresponded to 7 M urea and 40% (v/v) deionized formamide) was used. Gels were stained with SYBR Gold nucleic acid stain and photographed. Sequenced 16S rRNA genes of Lake Vida brine clones from 2005 samples (Murray *et al.,* 2012) were used as DGGE molecular markers to identify co-migrating phylotype bands in the brine microbial assemblages. Analyses of DGGE banding profiles were performed using the GelCompar II version 6.5 software package (Applied Maths, St-Martens-Latem, Belgium). DGGE banding patterns based on presence and absence were compared using Pearson's correlation coefficient (r) and cluster analyses were performed according to the unweighted pair-group method with average linkages (UPGMA) using the Statistica version 6.1 software package (StatSoft Inc., Tulsa, OK, USA).

A 16S rRNA gene clone library was constructed to phylogenetically identify the components of the 0.1-0.22 µm cell-size fraction. Products of four PCR reactions (24

cycles each) of genomic DNA amplified with the primers Bact27F and Univ1391R (Lane, 1991) were pooled for cloning. The PCR reactions were run on a 1.5% agarose gel electrophoresis and bands of approximately 1400 bp were extracted from the gel, combined and purified. The clone library was constructed using the pCR®4-TOPO® Vector (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's recommendation, and eletrotransformed into Top10 cells (Invitrogen, Carlsbad, CA, USA). Transformants were screened using blue-white screening. Plasmids were extracted by QIAprep® Spin Miniprep kit (Qiagen, Valencia, CA, USA) and sequenced with the BigDye V3.1 kit (Applied Biosystems, Foster City, CA, USA) using the vector's primers T7 and T3 as initiators. DNA extracted from the heterotrophic bacteria cultivated was amplified with the primers Bact27F and Univ1391R and amplified fragments were sequenced as well. DNA fragments were sequenced on a Prism 3730 DNA Analyzer (Applied Biosystems, Carlsbad, CA, USA) and phylogenetically analyzed.

The sequence reads from the clone library and cultures were trimmed and assembled. After a chimeras check using Bellepheron on-line tools (http://greengenes.lbl.gov) (Huber *et al.*, 2004), the sequences were edited using the software BioEdit (Hall, 1999) and aligned using the Nearest Alignment Space Termination (NAST) algorithm (DeSantis *et al.*, 2006). To identify nearest neighbors, sequences were compared to the GenBank database by BLASTn (Altschul *et al.*, 1990) against the nucleotide collection (nr/nt). A phylogenetic tree was constructed by Neighbor-Joining (NJ) analysis in Mega 5 software (Tamura *et al.*, 2011) with 1000

bootstrap replicates. Evolutionary distances were computed using the Jukes-Cantor model (Jukes and Cantor, 1969).

The nucleotide sequences of clone representatives from Lake Vida ultra-small microbial assemblage (0.1-0.22 µm cell-size fraction) and the cultivated isolate-group representatives were deposited at the GenBank database through the accession numbers JX136966 to JX136986 and KF384117 to KF384126.

Comparative phylogenetic analyses were performed among clone libraries from LVBrUMA, Lake Vida brine 2005 clone library (Murray *et al.*, 2012), the microbial assemblage within Lake Vida ice cover (Mosier *et al.*, 2007), and the Blood Falls (a subglacial outflow from the Taylor Glacier) microbial assemblage (Mikucki and Priscu, 2007). The rRNA gene sequence data sets were downloaded from GenBank and aligned using the NAST algorithm (DeSantis *et al.*, 2006). Distance matrices were generated individually for each library and the mothur software was run to generate nonredundant data sets for each library with all sequences represented at a distance of 0.01 (Schloss *et al.*, 2009). Overlapping regions between the five libraries provided an alignment with 1068 nucleotide positions. UniFrac (Lozupone *et al.*, 2005) was used to calculate statistical comparisons among the libraries to investigate their associations considering the richness and evenness of their microbial assemblages.

## 2.3. Results

### 2.3.1. Microscopy and elemental analyses

Confocal microscopy observations of the brine cells stained with DNA fluorescent stain indicated the presence of nucleic acid content in the "cell-like particles". Coccoid cells with consistent diameter of ~0.2 μm accounted for $6.13 \pm 1.65 \times 10^7$ cells $mL^{-1}$. Cells with sizes typical of aquatic bacteria (0.5–1.5 μm in length) accounted for $1.47 \pm 0.25$ X $10^5$ cells $mL^{-1}$ (Figure 2.3).

The two cell size populations were confirmed by electron microscopy observations (Figure 2.4). Coccoid cells with ~0.2 μm in diameter and rod-shaped bacteria with dimensions of ~0.4 μm in diameter and ~1.5 μm in length are shown in figures 2.4A and 2.4B. In addition to the observation of the two cell size populations, abundant smaller particles ranging in size from approximately 0.020 to 0.140 μm in diameter (nanoparticles) were observed (Figures 2.4A-C).

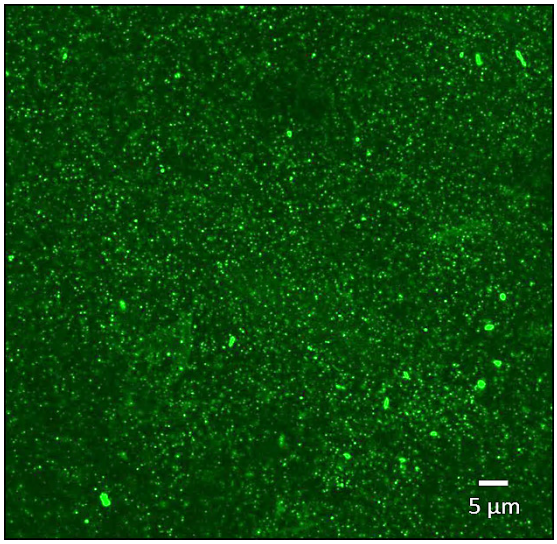



**Figure 2.3.** Confocal microscopy micrograph of brine cells stained with SYBR Gold.

When 1 mL of brine was prepared and dehydrated on the polycarbonate filter, a layer of mixed organic and inorganic material covered the filter surface (Figure 2.4A). Filtration of smaller volumes of brine, ≤ 500 µL, facilitated the visualization of the ultra-small cells and enabled the observation of structural details such as nanoparticles attached to the cell surfaces and fiber-like filaments connecting the cells (Figure 2.4C).

STEM micrographs of negative-stained ultra-small cells showed the presence of an electron-dense layer surrounding a less electron-dense region (Figure 2.4B and 2.4D). This observation is consistent with standard EM observations of stained microbial cells; the cell's membrane appears as an electron-dense layer and surrounds a less electron-dense region which is the cytoplasm. Larger bacterial cells in the brine presented the same characteristics different density structures (Figure 2.4B). Morphological variability such as single coccic and diplococcic was observed among the ultra-small cells (Figure 2.4D).

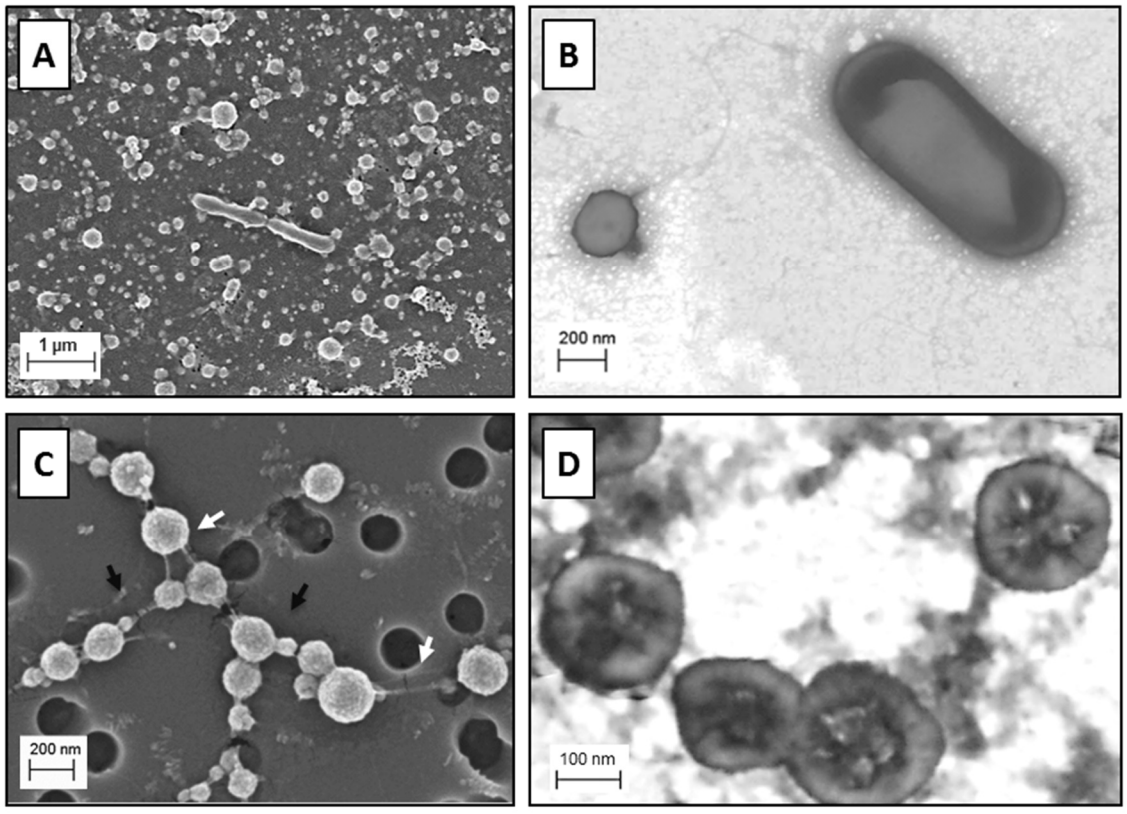



**Figure 2.4.** Electron microscopy micrographs of Lake Vida brine microbiota showing cell-representatives of the two cell size populations. A and C: SEM micrographs of brine cells prepared on filters. B and D: STEM micrographs of brine cells prepared on grids. (A): overview of brine cells surrounded by a dense layer of organic and inorganic material covering the filter (volume of brine filtered: 1 mL). (B): brine cells treated with EDTA and stained with UA pH 4-5. A cell with ~0.2 µm in diameter is shown on the left of the image and a rod-shaped bacterial cell is shown on the right. (C): Lake Vida ultra-small microbial cells prepared on a 0.2 µm polycarbonate filter (volume of brine filtered: 0.5 mL). Nanoparticles attached to the cells surface (indicated by black arrows) and uncharacterized filaments connecting the cells (indicated by white arrows) are shown. The filter's pores can be seen in the background (darker holes). (D) ultra-small brine cells stained with PTA pH0.4. Scale bars are shown in each micrograph.

Brine prepared on polycarbonate filters and air dried, without ethanol dehydration, presented a dense network of filamentous material co-located with cells and particles (Figures 2.5A and 2.5B). SEM micrographs in figure 2.5 show the dense filamentous network with a depth perception where brighter elements are in the surface and darker elements are located deeper in the network. A bacterial cell connected to the filamentous network is indicated with a white arrow in figure 2.5B.

Distinct EM preparations revealed different cell morphologies. Brine cells stained only with UA (Figure 2.6A) presented a rigid and granular appearance in their outer membrane which was not observed in cells treated with EDTA or stained with PTA at low pH (Figures 2.6B and 2.6C). Figure 2.6 shows the comparison of the different treatment observations; cells and nanoparticles stained with UA pH 4-5 presented a robust structure and distinct spherical morphologies, whereas the same cells and nanoparticles stained with UA pH 4-5 followed by PTA pH 0.4 did not present either the granular coat or the spherical shape. Indeed, with the PTA pH 0.4 treatment, the cells collapsed onto the filter as if structural support had been removed from them and nanoparticles were harder to identify (Figure 2.6B).

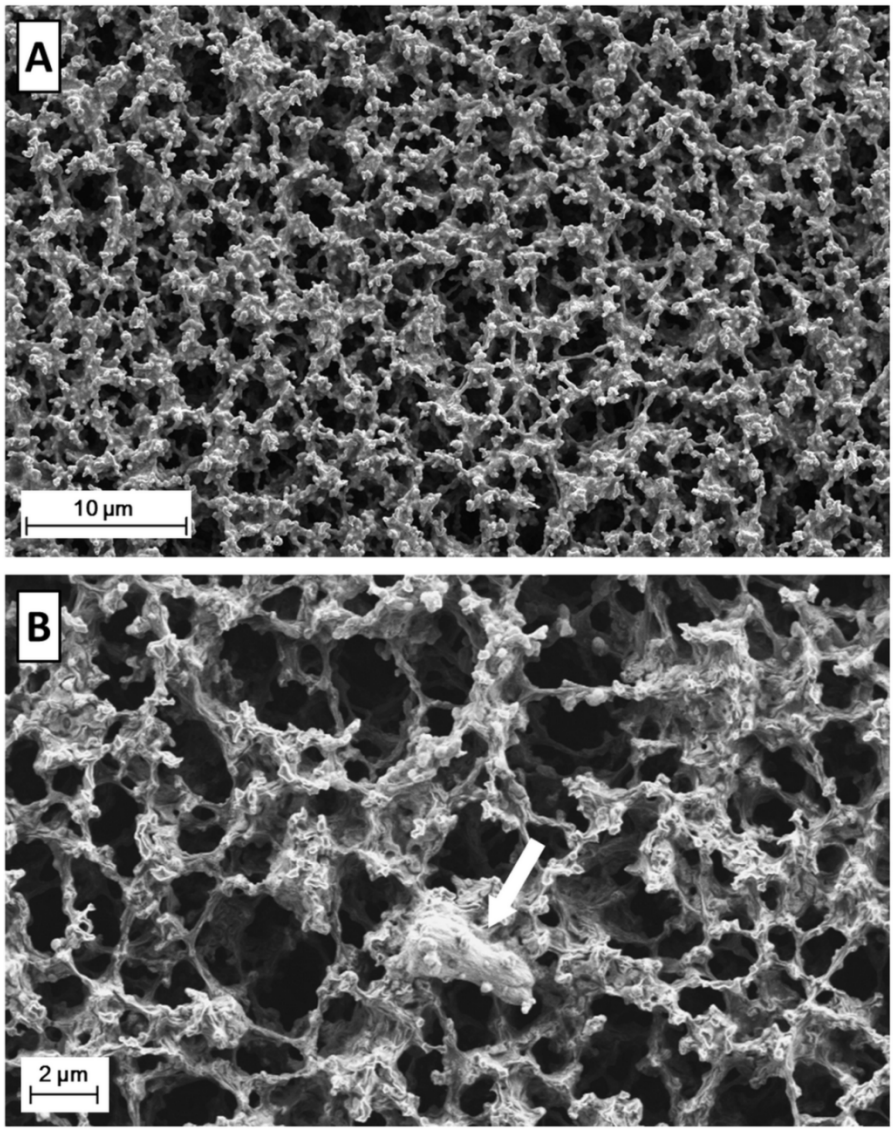



**Figure 2.5.** SEM micrographs of Lake Vida brine presenting the structural connection between the brine cells, filamentous network, and brine particles. Undehydrated-fixed brine (air-dried) was observed by SEM, giving a three dimensional perspective of the brine environment. (A) Overview of the filamentous network attached to cells and particles; (B) Close up of a microbial cell trapped on the filamentous network (indicated by white arrow). Scale bars are shown in each micrograph.

EDS measurements from the SEM observations showed that the granular coat and the spherical shape of the ultra-small cells and nanoparticles were associated with the presence of iron. The uranium signals observed in figure 2.6A indicated that the cells were stained by UA, which is co-located with the presence of iron in the same spectra. Grazing incidence X-ray diffraction analyses of a thick layer of brine insoluble components captured on a 0.2 μm filter suggested that the coat surrounding the cells is formed by precipitation of amorphous iron(III) oxide-hydroxide precipitated (a:Fe-O). Likewise, the absence of iron resulted in collapsed cells which stained well by PTA, as shown by the intense tungsten signal acquired (Figure 2.6B, spectra 2).

EDS spectra taken during STEM observations showed a prominent tungsten signal at 1.8 keV at the ultra-small cell edge and the cell center indicating the collapsed cell membrane and smashed cytoplasm (Figure 2.6C). Significantly, the U and W signals are absent or weak when the electron beam was placed on the filter or TEM grid surface (backgrounds). Intense C and O peaks accompanied by a weak N signal were also detected in the cells while the background yielded weaker W, C, and O signals and no N signal (spectra 1, Figure 2.6C).

The majority (96%) of the brine particles (cells and unidentified nanoparticles) observed by electron microscopy fell into diameter sizes ≤ 0.5 μm (Figure 2.7). Half of the particles measured (52%) ranged between 0.150-0.250 μm in diameter. Statistically, the coccoid cells with an average diameter of approximately 0.2 μm are cells with diameter of 0.192 ± 0.065 μm (pick center B in top graph, figure 2.7) and comprised 56-60% of the total particles measured. The unidentified nanoparticles have a diameter of

0.084 ± 0.063 μm (pick center A in top graph, figure 2.7) and comprised 31-36% of the total particles measured.

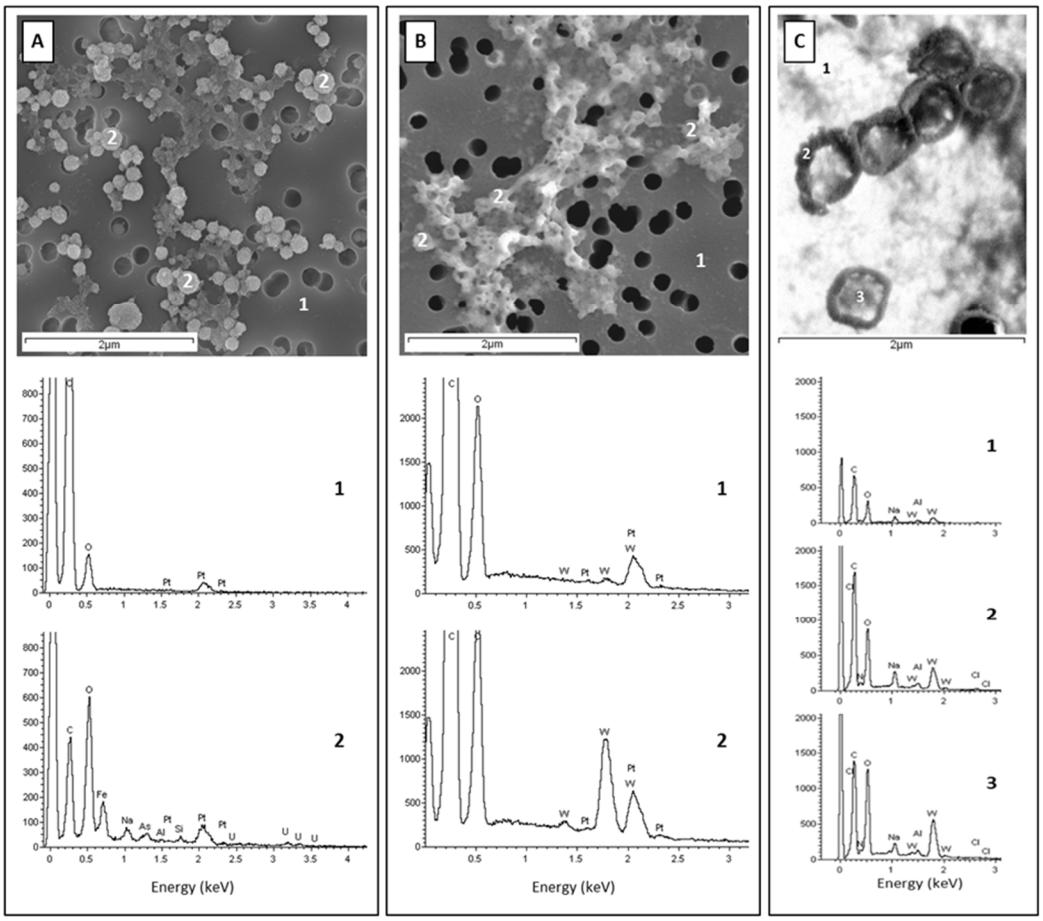



**Figure 2.6.** Electron microscopy micrograph and EDS analyses of the brine ultra-small cells. A and B: SEM micrograph and EDS analyses from cells on the presence (A) and absence (B) of iron precipitates. (A) cells stained with uranyl acetate (UA, $UO_2(CH_3COO)_2 \cdot 2H_2O$) pH 4-5. (B) cells stained with UA followed by phosphotungstic acid stain (PTA, $H_3PW_{12}O_{40}$) pH 0.4. Cells were treated, deposited on a polycarbonate filter, and coated with platinum. Numbers 1 and 2 mark the spots where EDS analyses were measured: (1) the background, (2) the cells. (C): STEM micrograph of cells negative stained with PTA pH 0.4 and deposited on a LuxFilm$^{TM}$ grid. Numbers 1, 2, and 3 mark the spots where EDS analyses were measured: (1) the background, (2) electron-dense edge of the ~0.2 μm size-cell, and (3) less electron-dense in the middle of the cell. Elements identified by EDS: C - Carbon, N - Nitrogen, O - Oxygen, Fe - Iron, Na - Sodium, As - Arsenic, Pt - Platinum, W - Tungsten, Al - Aluminum, Si - Silicon, Cl - Chlorine, and U - Uranium. Scale bar is shown in the micrograph.

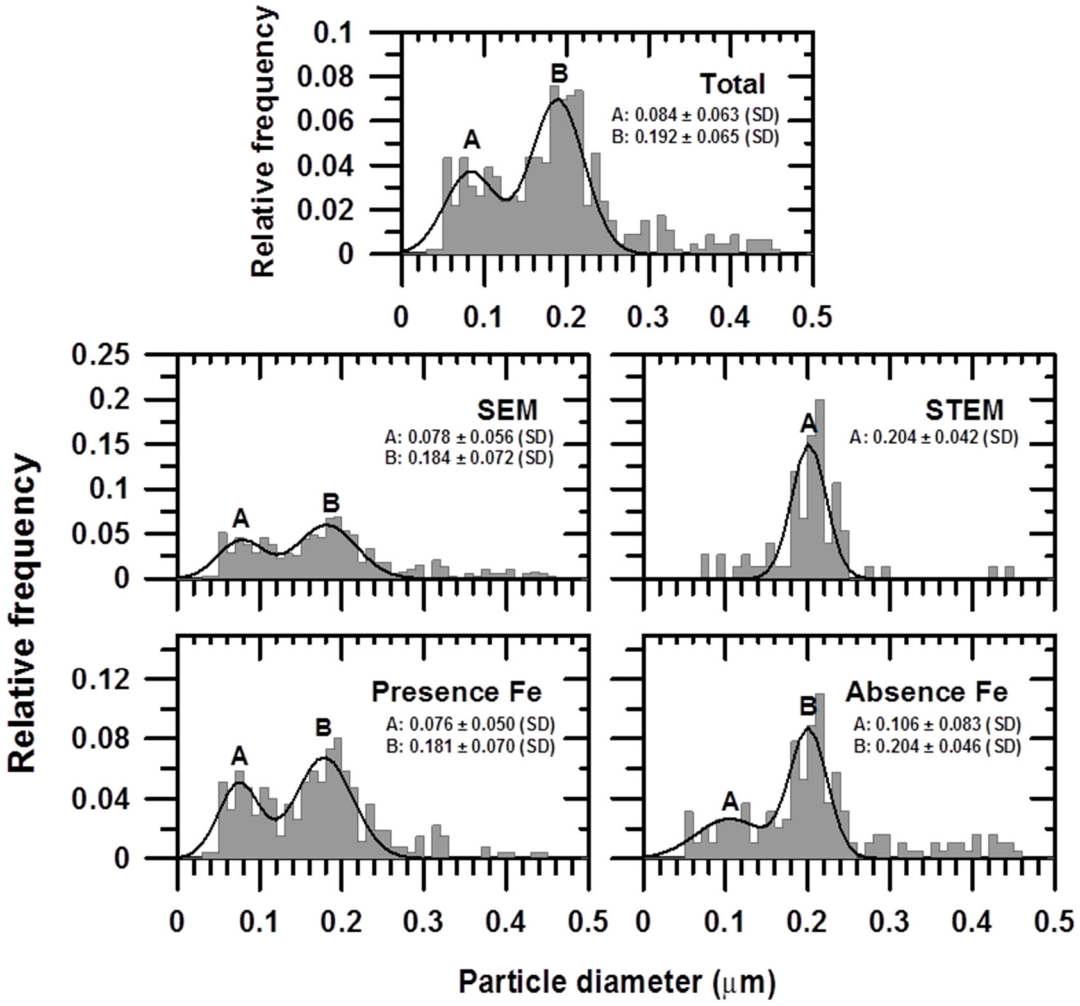



**Figure 2.7.** Size distribution of cells and unidentified nanoparticles in Lake Vida brine. Size distribution of 465 cells/particles measured from SEM and STEM observations and in the presence and the absence of amorphous iron(III) oxide-hydroxide are shown in five distinct graphs. Gaussian distribution fits, pick centers, and standard deviation are plotted in each graph.

Differences in size distributions were identified when comparing the measurements of particles observed by SEM *versus* STEM and when comparing the particles with the presence *versus* absence of a:Fe-O in the preparation (Figure 2.7). In the SEM observations, cells and nanoparticles appeared with similar abundance. The

nanoparticles, however, were not significantly present in the STEM observations and were less abundant in preparations with the absence of a:Fe-O.

### 2.3.2. Heterotrophic bacteria cultivation

Cultivation efforts with the ≤ 0.2μm brine fraction using three different temperatures, five different media, and aerobic conditions yielded a total of 27 CFU on the R2A+5% NaCl plates after 120 days. The other four media tested did not yield colonies. From these 27 CFU, 15 were isolated and classified as being related to three *Actinobacteria* genera: *Microbacterium lacus* strain R-43968 (13 isolates, 98.8% identity), *Microbacterium testaceum* strain S15-M13 (1 isolate, 99.8% identity), and *Kocuria marina* strain KMM 3905 (1 isolate, 99.8% identity) (Table 2.1).

Cultivation efforts with the unfiltered brine yielded a total of 335 CFU from which the 33 selected for isolation were phylogenetically affiliated with *Marinobacter* sp. strain LV10S (20 isolates, 99.5% identity), *Marinobacter* sp. strain ELB17 (8 isolates, 99.2% identity), *Marinobacter psychrophilus* strain R-36953 (1 isolate, 98.6% identity), *Psychrobacter* sp. strain LV414 (3 isolates, 99.5% identity), and *Arthrobacter halodurans* strain JSM 078085 (1 isolate, 97.2% identity) (Table 2.1). Fungi were not detected in any culturing plate and *Marinobacter* and *Psychrobacter* affiliated bacteria - the major isolates identified in the unfiltered brine cultivars – were not isolated from the ≤ 0.2μm brine fraction.

**Table 2.1.** Summary of the heterotrophic bacteria strains isolated from Lake Vida brine collected in 2010. Isolate ID: Brine with (LV10f) and without (LV10) size fractionation. Isolates identification was based on 16S rRNA gene sequences amplified with the set of primers Bact27F and Univ1391R and sequenced bidirectionally.

| Lake Vida brine isolate-group representatives* | Number of isolates | Media and temperatures | Closest cultivated relative (GenBank ID) | % identity (nucleotides) | Comments about closest-affiliate cultivated |
|---|---|---|---|---|---|
| LV10fR510-15 (KF384117) | 13 | R2A+5% NaCl 10 and 20°C | *Microbacterium lacus* strain R-43968 (FR691402) | 98.8 (1185/1199) | Psychrophilic irregular rod-shaped actinobacterium isolated from Lundström Lake, Shackleton Range, Antarctica (Peeters et al., 2011) |
| LV10fR510-14 (KF384118) | 1 | R2A+5% NaCl 10°C | *Microbacterium testaceum* strain S15-M13 (AM234161) | 99.8 (1197/1199) | Actinobacterium isolated from radioactive waste depository Tomsk-7, Siberia, Russia (Nedelkova et al., 2007) |
| LV10fR520-8 (KF384119) | 1 | R2A+5% NaCl 20°C | *Kocuria marina* strain KMM 3905 (AY211385) | 99.8 (1203/1205) | Halotolerant denitrifier coccoid actinobacterium isolated from marine sediment from East Siberian Sea (Kim et al., 2004) |
| LV10R510-5 (KF384120) LV10R510-8 (KF384121) | 20 | R2A+5% NaCl and Marine agar+5% NaCl 10 and 20°C | *Marinobacter* sp. strain LV10S (EU908281) | 99.5 (1211/1217) | Psychrotolerant/ halophilic rod-shaped gammaproteobacterium isolated from Lake Vida brine collected in 2005 (Mondino et al., 2009) |
| LV10R510-11A (KF384122) LV10MA510-1 (KF384123) | 8 | R2A+5% NaCl and Marine agar+5% NaCl 10 and 20°C | *Marinobacter* sp. strain ELB17 (AY518678) | 99.2 (1221/1230) | Psychrophilic denitrifier rod-shaped gammaproteobacterium isolated from East lobe of Lake Bonney, Antarctica, at 17 m depth (Ward and Priscu, 1997) |
| LV10R510-2 (KF384124) | 1 | R2A+5% NaCl 10°C | *Marinobacter psychrophilus* strain R-36953 (FR691434) | 98.6 (1214/1230) | Psychrophilic rod-shaped gammaproteobacterium isolated from Forlidas Pond, Pensacola Mountains, Antarctica (Peeters et al., 2011) |
| LV10R520-6 (KF384125) | 3 | R2A+5% NaCl and Casein agar 10 and 20°C | *Psychrobacter* sp. strain LV414 (EU908288) | 99.5 (1209/1214) | Psychrophilic/ halophilic coccoid gammaproteobacterium isolated from Lake Vida brine collected in 2005 (Mondino et al., 2009) |
| LV10R520-3 (KF384126) | 1 | R2A+5% NaCl 20°C | *Arthrobacter halodurans* strain JSM 078085 (EU583729) | 97.2 (1170/1203) | Halotolerant denitrifier actinobacterium with rod-coccus growth cycle isolated from South China Sea seawater (Chen et al., 2009) |

*Isolate nomenclature: Sample, media, temperature – isolate number. Examples: LV10fR520-8 (LV10f/R5/20-/8: Lake Vida brine collected in 2010 and fractionated, isolate cultivated on R2A plus 5% NaCl, at 20 ºC, isolate number 8); LV10MA510-1 (LV10/MA5/10-/1: Lake Vida brine collected in 2010, isolate cultivated on Marine agar plus 5% NaCl, at 10ºC, isolate number 1). Number in parentheses under ID indicates the GenBank accession number of the 16S rRNA gene sequence of the isolate.

### 2.3.3. Molecular analyses

The microbial assemblage of the unfiltered brine and LVBrUMA were initially compared through DGGE patterns of the amplified 16S rRNA gene (Figure 2.8). Cluster analysis showed that the DNA and cDNA profiles from the unfiltered brine microbial assemblage were similar, with a Linkage Distance of 0.8 (Figure 2.8, lanes 1 and 2, respectively), and distinctly separated from the LVBrUMA DNA profile (lane 3), with a Linkage Distance of 1.2. The LVBrUMA 16S rRNA gene banding profile shared only six phylotypes out of 30 with the unfiltered brine microbial assemblage cDNA banding profile (Figure 2.8, A-F). Based on co-migration, phylotypes B, D, E, and F were related to *Psychrobacter* sp., uncultured *Epsilonproteobacterium*, uncultured *Epsilonproteobacterium*, and *Marinobacter* sp., respectively.

The analyses of 100 16S rRNA gene sequences from LVBrUMA clone library showed that the 0.1-0.2 μm size fractionated cells were affiliated with five bacterial phyla: *Proteobacteria* (*Beta-*, *Gamma-*, *Epsilon-)*, *Actinobacteria, Firmicutes, Bacteroidetes,* and Candidate Division TM7 (Figure 2.9). The majority of the sequences, 80%, were affiliated with the phylum *Proteobacteria*. Comparison between the 2005 16S rRNA gene clone library data of the brine microbial assemblage collected on a 0.2 μm pore-sized filter (LVBr) (Murray *et al.*, 2012) and the 2010 ultra-small microbial assemblage 16S rRNA gene clone library data from the 0.1–0.2 μm fraction (LVBrUMA) revealed that the phylogenetic composition of the two cell-size fractions were not identical. *Gammaproteobacteria* was the main class detected in both size fractions (38% of both assemblages). Sequences closely affiliated with the genera *Psychrobacter,*

*Marinobacter*, and the strain *Pseudoalteromonas haloplanktis* TAC125 (98.6 to 99.1%, 98.6 to 99%, and 99.3% of identity, respectively) prevailed in the *Gammaproteobacteria* class of LVBrUMA fraction. The second most abundant group detected in the LVBrUMA fraction (35% of sequences) was affiliated with *Betaproteobacteria*. Sequences closely affiliated with the genus *Herbaspirillum* (99.7% of identity) prevailed in this class. *Betaproteobacteria* affiliated sequences were not detected in LVBr. Additionally, *Deltaproteobacteria, Spirochaetea*, and *Lentisphaerae* sequences (comprising 1%, 7%, and 16% of LVBr, respectively), were not detected in LVBrUMA.

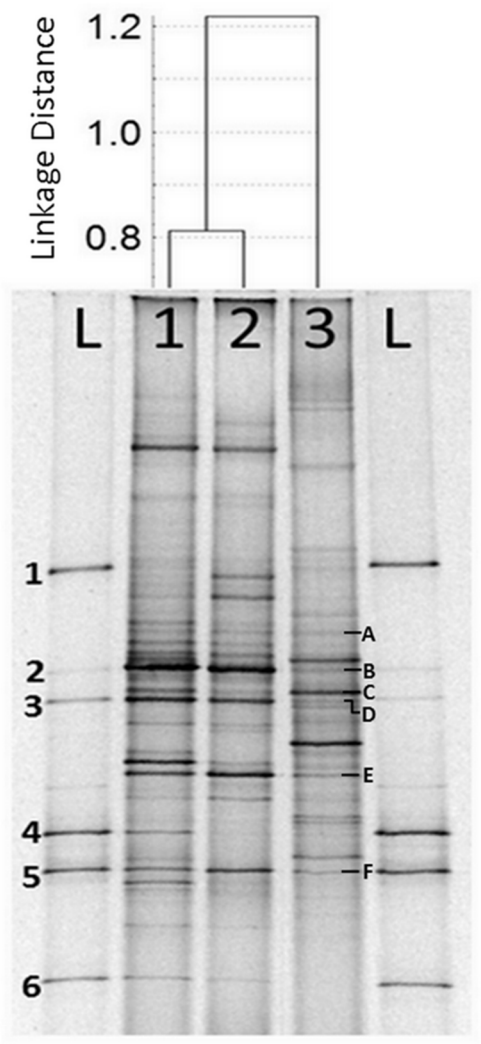



**Figure 2.8.** DGGE profile and cluster analysis of 16S rRNA gene and 16S cRNA from Lake Vida microbial assemblages. Samples were loaded as follows: lane 1: amplified DNA from the unfiltered brine; lane 2: amplified cDNA from the unfiltered brine; lane 3: amplified DNA from LVBrUMA. (L) Ladder: Lake Vida 16S rRNA gene clones phylogenetically related to (1) uncultured Verrucomicrobia bacterium (GQ167324); (2) *Psychrobacter* sp. (GQ167313); (3) uncultured epsilonproteobacterium (GQ167319); (4) uncultured flavobacterium (GQ167312); (5) *Marinobacter* sp. (GQ167320); (6) *Sphaerochaeta* sp. (GQ167322). The dendrogram, calculated based on the presence and absence of the amplified 16S rRNA gene partial fragments, shows the linkage distance between the samples. Letters A to F indicate the six phylotypes detected in LVBrUMA DNA and the unfiltered brine cDNA.

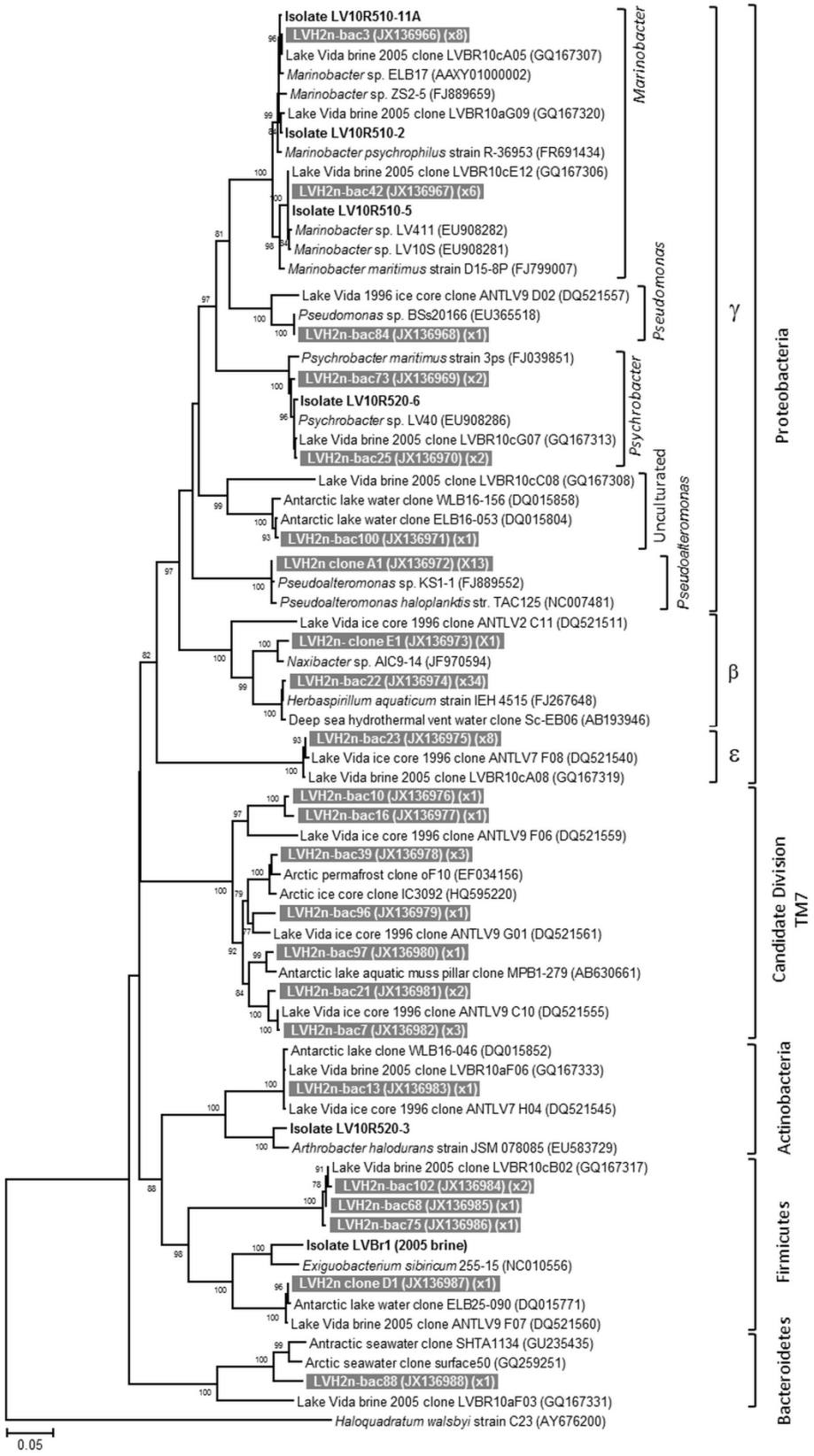

Isolate LV10R510-11A
LVH2n-bac3 (JX136966) (x8)
Lake Vida brine 2005 clone LVBR10cA05 (GQ167307)
Marinobacter sp. ELB17 (AAXY01000002)
Marinobacter sp. ZS2-5 (FJ889659)
Lake Vida brine 2005 clone LVBR10aG09 (GQ167320)
Isolate LV10R510-2
Marinobacter psychrophilus strain R-36953 (FR691434)
Lake Vida brine 2005 clone LVBR10cE12 (GQ167306)
LVH2n-bac42 (JX136967) (x6)
Isolate LV10R510-5
Marinobacter sp. LV411 (EU908282)
Marinobacter sp. LV10S (EU908281)
Marinobacter maritimus strain D15-8P (FJ799007)
Marinobacter
Lake Vida 1996 ice core clone ANTLV9 D02 (DQ521557)
Pseudomonas sp. BSs20166 (EU365518)
LVH2n-bac84 (JX136968) (x1)
Pseudomonas
Psychrobacter maritimus strain 3ps (FJ039851)
LVH2n-bac73 (JX136969) (x2)
Isolate LV10R520-6
Psychrobacter sp. LV40 (EU908286)
Lake Vida brine 2005 clone LVBR10cG07 (GQ167313)
LVH2n-bac25 (JX136970) (x2)
Psychrobacter
Lake Vida brine 2005 clone LVBR10cC08 (GQ167308)
Antarctic lake water clone WLB16-156 (DQ015858)
Antarctic lake water clone ELB16-053 (DQ015804)
LVH2n-bac100 (JX136971) (x1)
Uncultured
LVH2n clone A1 (JX136972) (X13)
Pseudoalteromonas sp. KS1-1 (FJ889552)
Pseudoalteromonas haloplanktis str. TAC125 (NC007481)
Pseudoalteromonas
γ
Lake Vida ice core 1996 clone ANTLV2 C11 (DQ521511)
LVH2n- clone E1 (JX136973) (X1)
Naxibacter sp. AIC9-14 (JF970594)
LVH2n-bac22 (JX136974) (x34)
Herbaspirillum aquaticum strain IEH 4515 (FJ267648)
Deep sea hydrothermal vent water clone Sc-EB06 (AB193946)
β
LVH2n-bac23 (JX136975) (x8)
Lake Vida ice core 1996 clone ANTLV7 F08 (DQ521540)
Lake Vida brine 2005 clone LVBR10cA08 (GQ167319)
ε
Proteobacteria
LVH2n-bac10 (JX136976) (x1)
LVH2n-bac16 (JX136977) (x1)
Lake Vida ice core 1996 clone ANTLV9 F06 (DQ521559)
LVH2n-bac39 (JX136978) (x3)
Arctic permafrost clone oF10 (EF034156)
Arctic ice core clone IC3092 (HQ595220)
LVH2n-bac96 (JX136979) (x1)
Lake Vida ice core 1996 clone ANTLV9 G01 (DQ521561)
LVH2n-bac97 (JX136980) (x1)
Antarctic lake aquatic muss pillar clone MPB1-279 (AB630661)
LVH2n-bac21 (JX136981) (x2)
Lake Vida ice core 1996 clone ANTLV9 C10 (DQ521555)
LVH2n-bac7 (JX136982) (x3)
Candidate Division TM7
Antarctic lake clone WLB16-046 (DQ015852)
Lake Vida brine 2005 clone LVBR10aF06 (GQ167333)
LVH2n-bac13 (JX136983) (x1)
Lake Vida ice core 1996 clone ANTLV7 H04 (DQ521545)
Isolate LV10R520-3
Arthrobacter halodurans strain JSM 078085 (EU583729)
Actinobacteria
Lake Vida brine 2005 clone LVBR10cB02 (GQ167317)
LVH2n-bac102 (JX136984) (x2)
LVH2n-bac68 (JX136985) (x1)
LVH2n-bac75 (JX136986) (x1)
Isolate LVBr1 (2005 brine)
Exiguobacterium sibiricum 255-15 (NC010556)
LVH2n clone D1 (JX136987) (x1)
Antarctic lake water clone ELB25-090 (DQ015771)
Lake Vida brine 2005 clone ANTLV9 F07 (DQ521560)
Firmicutes
Antractic seawater clone SHTA1134 (GU235435)
Arctic seawater clone surface50 (GQ259251)
LVH2n-bac88 (JX136988) (x1)
Lake Vida brine 2005 clone LVBR10aF03 (GQ167331)
Bacteroidetes
Haloquadratum walsbyi strain C23 (AY676200)
0.05

**Figure 2.9.** Neighbor-Joining phylogenetic tree based on near-complete 16S rRNA gene sequences from LVBrUMA and distance-affiliated sequences. The alignment utilized fragment sizes of 1458 nucleotide positions. The bootstrap of 1000 repetitions with support values above 75 are represented at nodes. The scale bar denotes the divergence percentage between sequences. LVBrUMA environmental clones are represented by gray boxes with white text. GenBank accession numbers for the nucleotide sequences are in the first set of parentheses, following the sequence name. The number of clones related to the same sequence in the clone library (0.01 distance) is in the second set of parentheses. Lake Vida brine heterotrophic strains isolated and characterized in this study are in bold (exception: LVBr1 was isolated from brine collected in 2005).

None of the sequences affiliated with Candidate Division TM7 identified in the LVBrUMA 16S rRNA gene clone library were closely affiliated with the sequences from the same Candidate Division TM7 group found in the LVBr clone library. However, the LVBrUMA sequences were affiliated with environmental sequences identified in an ice core recovered at 15.9 m from Lake Vida in 1996 with distance of 0.06 (Figure 2.9).

Comparison of LVBrUMA with other environmental 16S rRNA gene clone libraries [LVBr (Murray *et al.*, 2012), Lake Vida ice core microbial assemblages from the depth sections 4.8 m and 15.9 m (Mosier *et al.*, 2007), and Blood Falls microbial assemblage (Mikucki and Priscu, 2007)] showed different evenness and richness clustering results. Evenness analyses showed LVBrUMA forming an external cluster with LVBr and the 15.9 m deep portion of Lake Vida ice cover (Figure 2.10A). Richness analyses showed LVBrUMA clustering with the 15.9 m deep ice core assemblage of the lake (Figure 2.10B).

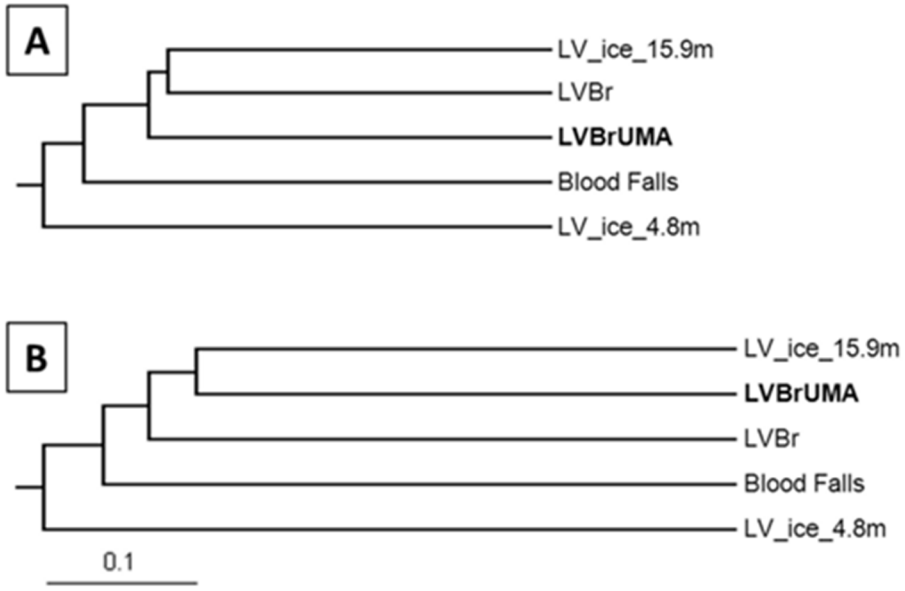



**Figure 2.10.** Comparison of LVBrUMA with previously characterized microbial assemblages from Lake Vida and Blood Falls. (A) UniFrac cluster dendrogram considering the normalized abundance weight (evenness) of the 16S rRNA gene sequences of the microbial assemblages compared. (B) UniFrac cluster dendrogram considering only the presence and absence (richness) of the 16S rRNA gene sequences of the microbial assemblages compared. LVBrUMA: Lake Vida brine ultra-small microbial assemblage; LVBr: Lake Vida brine microbial assemblage collected in 2005 (Murray *et al.*, 2012); LV_ice_4.8 m and LV_ice_15.9 m: Lake Vida ice core microbial assemblages from the depth sections 4.8 m and 15.9 m (Mosier *et al.*, 2007); Blood Falls: microbial assemblage from Taylor Glacier's subglacial outflow (Mikucki and Priscu, 2007).

## 2.4. Discussion

### 2.4.1. Electron microscopy and elemental analyses

The detection of membrane boundaries and stain signals above background by electron microscopy and EDS indicates that Lake Vida brine ~0.2 μm "cell-like particles" are cellular entities. The observation of cellular morphological variability such as

diplococci, cells connected by membrane-septa, presence of a capsular structure, and formation of aggregates support the conclusion that they are cells.

EDS measurements and grazing angle X-ray diffraction show that the cells are surrounded by an amorphous iron oxy-hydroxide (a:Fe-O) layer with unknown genesis. We are describing the layer as a:Fe-O because no peaks due to minerals such as ferrihydrite, goethite, or lepidocrocite were observed by X-ray diffraction. Different treatments of the brine sample suggested that the a:Fe-O layer surrounding the cells may be a consequence of the contact of the abundant dissolved iron in the brine with the air during sample preparation. Aqueous $Fe^{2+}$ intrinsic to Lake Vida brine could oxidize to $Fe^{3+}$ in air, which then forms insoluble hydrous ferric oxide at near neutral pH, and precipitates on the cell and particles surfaces (Cotton *et al.*, 1972). To overcome this side effect and be able to observe the cells without iron precipitates, PTA pH 0.4 and EDTA were used to keep the iron in solution. Another explanation for the a:Fe-O coat is that the cells could be covered by an oxidized a:Fe-O layer in the natural brine environment. Microbial cells enriched with exopolysaccharides (which carry a net negative electric charge and have the capacity to bind $Fe^{2+}$ ions) are known to serve as a nucleation site for Fe(III) encrustation formed from oxidation reactions (Frankel, 2003; Miot *et al.,* 2009).

The nature of the nanoparticles of $0.084 \pm 0.063$ μm in diameter, observed mostly by SEM, remains under investigation. As the nanoparticles were more frequently observed in the presence of a:Fe-O, they could be a consequence from the nucleation of a:Fe-O on cellular debris, on viral particles, on organic matter, or on inorganic particles present in the brine.

Evidences including size, morphology, and attachment to the cell surfaces suggest that some of the nanoparticles might be extracellular membrane vesicles (MV, Figure 2.11). The production of MV is a biochemically controlled process frequently described in Gram-negative bacteria and recently observed in Gram-positive bacteria as well (Duda *et al.*, 2009; Kulp and Kuehn, 2010; Deatherage and Cookson, 2012). Vesiculation was described as a temperature stress response used by mesophilic cells to remove undesirable soluble and insoluble components from the bacterial envelope (McBroom and Kuehn, 2007). The authors observed that the production of MV correlated positively with survival by eliminating misfolded proteins from the cell. Protein misfolding in cold and freezing environments is a crucial obstacle that cells need to overcome for survival. Cold adapted bacteria from the genera *Marinobacter, Psychrobacter,* and *Pseudoalteromonas* have been described as producing large numbers of vesicles in cultures grown *in vitro* from 4ºC to 15ºC (Nevot *et al.*, 2006a; Nevot *et al.*, 2006b; Frias *et al.*, 2010). The vesicles produced by these microorganisms were spherical, ranging in size from 20 to 100 nm, and were observed to be connected to cells and to extracellular polymeric substances (EPS).

The nanoparticles could also be representative of inclusion bodies (aggregated proteins) or viral particles. One of the prominent markers of cell aging associated with cellular deterioration is the formation of inclusion bodies (Nyström, 2007; Lindner *et al.,* 2008). Inclusion bodies released by disrupted microbial cells could serve as nucleation sites for precipitating a:Fe-O thereby forming the nanoparticles. Similarly, viral particles, ubiquitous in worldwide and in Antarctic aquatic ecosystems (Kepner *et al.,* 1998), could

serve as nucleation sites. Abundance of $10^5$ to $10^7$ virus-like particles per mL were reported in the water column of perennially ice-covered (3-6 m thick) lakes located in Taylor and Wright Valleys, Antarctica (Lisle and Priscu, 2004). For Lake Vida ice-encapsulated brine, the presence and dynamic of viral particles is currently unknown. We can state that recognizable tailed bacteriophage particles were not observed by electron microscopy. However, preliminary analysis of viral nucleic acid sequences from the 0.1-0.22 μm cell-size brine fraction metagenome (draft) indicate the presence of sequences affiliated to the order *Caudovirales*, tailed bacteriophage particles which size range between 30 to 140 nm.

Further investigation of the brine metagenome data in addition to experimentation with the brine isolates could assist in characterizing the nanoparticles and reach an accurate conclusion about their nature, *e.g.* if they are MV, viral particles, inclusion bodies, organic-inorganic particles, or a combination.

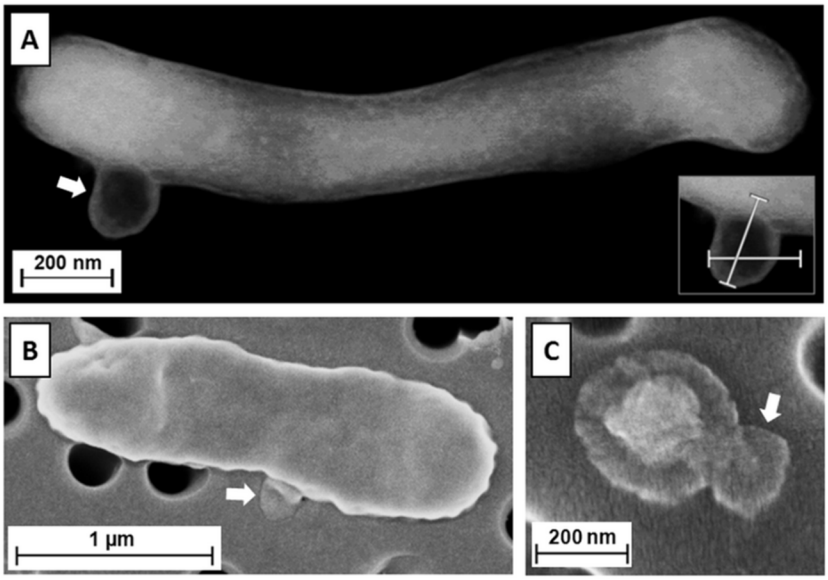



**Figure 2.11.** External cellular structures observed in Lake Vida brine cells that could be representative of membrane vesicles (indicated by arrows). (A) STEM dark field micrograph of a brine cell stained with UA pH4-5. Detail on the right shows the scale bar (200 nm) on the external cellular structure; (B) SEM micrograph of a brine cell presenting a similar structure as (A); (C) SEM micrograph of an ultra-small cell from the brine presenting a capsular structure surrounding the cell and an extracellular structure that could also be representative of a membrane vesicle. Scale bars are shown in each micrograph.

Lake Vida brine contains a network of uncharacterized fiber-like filaments connecting cells and particles (Figures 2.4C and 2.5). The fiber-like strands observed from our preparations are similar to collapsed and dehydrated EPS observed in a cryo-electron microscopy study of *Shewanella oneidensis* (Dohnalkova *et al.*, 2011). In Dohnalkova *et al.* (2011) study, controlled loss of water from the hydrated specimen showed that the diffuse low-contrast EPS collapsed into long branching filaments. Based on Dohnalkova *et al*. (2011), the fiber-like filaments observed in Lake Vida brine are

most likely EPS that has undergone severe contraction during water loss. Many of the cultivated microorganisms that are phylogenetically associated with the Lake Vida microbial community have been described producing complex EPS matrices *in vitro* (*e.g. P. antarctica* $NF_3$, Nevot *et al.*, 2006a). The production of EPS matrices by the cells in cold and freezing environments is a strategy used for cryo-protection (Krembs and Deming, 2008). Further investigation is needed to characterize the brine filamentous network. Stabilizing the original brine with vitreous ice and observing it by STEM through ultrathin sections could potentially allow the observation of undisturbed filamentous network, cells and nanoparticles.

### 2.4.2. Heterotrophic bacteria cultivation

Cultivation of the ≤ 0.2µm brine fraction resulted in isolation of Gram-positive bacteria. The 15 isolates cultivated and identified from the ≤ 0.2µm brine fraction belong to the family *Micrococcaceae* (class *Actinobacteria*). Strain LV10fR510-15, which represented 13 of the 15 isolates, was phylogenetically identified as *Microbacterium lacus*, with 98.8% identity to the strain R-43968 isolated from Lundström Lake, West Antarctica (Peeters *et al.*, 2011). Until now, the only type strain of this species characterized, strain A5E-52, was isolated from the sediment of a river in Japan (Kageyama *et al.,* 2007). Strain A5E-52 is affiliated with Lake Vida strain at a level of 98.6% 16S rRNA gene sequence identity. Strain LV10fR510-14 was phylogenetically affiliated (99.8% identity) with a *Microbacterium testaceum* isolated from the radioactive waste depository Tomsk-7, Siberia, Russia (Nedelkova *et al.*, 2007). Strain LV10fR520-8

was phylogenetically affiliated (99.8% identity) with the halotolerant *Kocuria marina* isolated from East Siberian Sea marine sediment (Kim *et al.*, 2004) (Table 2.1). None of these *Micrococcaceae*-related isolates were cultivated from the unfiltered brine.

Cultivation of the unfiltered brine resulted in isolation of 32 Gram-negative bacteria and one Gram-positive bacterium. *Marinobacter* strains were the most abundant isolates identified among the 33 isolates. The *Marinobacter* sp. isolates were clustered (distance of 0.01) into three 16S rRNA sequence-based groups (Table 2.1). The major group (represented by the strains LV10R510-5 and LV10R510-8) presented 99.5% nucleotide identity to one of the psychrotolerant/halophilic rod-shaped strains isolated from Lake Vida brine collected in 2005, strain LV10S (Mondino *et al.*, 2009). Two out of the three *Marinobacter* sp. identified in this study were not isolated from the Lake Vida system before, but were isolated from other Antarctic lake systems (Ward and Priscu, 1997; Peeters *et al.*, 2011; Table 2.1). None of these *Micrococcaceae* related isolates were cultivated from the unfiltered brine.

Three *Psychrobacter* and one *Arthrobacter* related isolates were identified among the unfiltered brine cultures. The *Psychrobacter* isolates were phylogenetically affiliated with the strain LV414 isolated from the lake brine collected in 2005 (Mondino *et al.*, 2009). The Gram-positive bacterium was phylogenetically affiliated with *Arthrobacter* sp. strain LV10R520-3, a halotolerant actinobacterium cultivated from seawater (Chen *et al.*, 2011). Members of the genus *Arthrobacter* are widespread coryneform facultative aerobic bacteria abundant in soils and sediments (Sheridan *et al.*, 2003; Peeters *et al.*, 2011), with a rod-coccus morphological growth cycle and known to

inhabit the dry valley soils (Cary *et al.*, 2010). The genus was cultivated from the unfiltered brine, but was not detected in the LVBrUMA clone library and was not detected among the cultures obtained from the ≤ 0.2µm brine fraction.

*Psychrobacter* and *Marinobacter* are genera well known for inhabiting cold and salty environments (Gilichinsky *et al.*, 2005; Zhang *et al.*, 2008; Ayala-del-Río *et al.*, 2010). In the Lake Vida system, in addition to cultivated *Psychrobacter* and *Marinobacter* cells, *Marinobacter* associated 16S rRNA gene sequences were detected in the brine and in the lake ice cover collected at the depth of 15.9 m (Mosier *et al.*, 2007; Mondino *et al.*, 2009; Murray *et al.*, 2012). It was expected that *Marinobacter* sp. strains would be recovered from both brine fraction culturing efforts. However, the major biological component of Lake Vida known to be active, viable, and cultivated, was not recovered from the ≤ 0.2µm brine fraction.

### 2.4.3. Molecular analyses

According to our results, size fractionation allowed the detection of phylotypes that were not previously detected in the brine assemblage collected on 0.2 µm pore-sized filters. Size fractionation has been applied to characterize microbial communities (Miyoshi *et al.*, 2005; Nakai *et al.*, 2011) and to isolate and enrich new and not readily cultivated bacteria (Hahn *et al.*, 2004; Miteva *et al.*, 2005). The predominant taxon detected in LVBrUMA, which was a group not detected in the LVBr clone library, belongs to the *Betaproteobacteria* class. The enrichment of the *Betaproteobacteria* class

taxa by size fractionation methodologies was previously reported in microbial assemble composition studies from glacier ice core samples from Greenland, deep-ground water, and deep-sea hydrothermal vents (Miteva *et al.* 2005; Miyoshi *et al.*, 2005).

Sequences affiliated with Candidate Division TM7 group and *Pseudoalteromonas haloplanktis* strain TAC125 were also enriched in LVBrUMA clone library by size fractionation when comparing LVBrUMA and LVBr clone libraries. The genus *Pseudoalteromonas* is ubiquitous in marine environments (Lovejoy *et al.*, 1998) and *P. haloplanktis,* a psychrophilic moderate halophilic, Gram-negative bacterium, isolated from Antarctic seawater (Médigue *et al.*, 2005), has the capacity to grow at low temperatures, even at 0$^{o}$C, and is well adapted to high salinities with optimal growth between 1.5% and 3.5% of NaCl. The strain is also able to deal with membrane fluidity challenges at low temperatures by expressing lipid desaturases and has high expression of genes involved in the syntheses of biofilm (Médigue *et al.*, 2005).

### 2.4.4. Possible nature of the ultra-small cells

The morphology observed in the brine ultra-small microbial cells is comparable to cells with pleomorphic life cycles (emerging to ultra-small cells in late life cycle stage), cyst-like resting cells, and cells exposed to stressful osmotic conditions (Soina *et al.*, 2004; Joshi & Toleti 2009). None of the bacteria groups detected in the LVBrUMA have been reported as ultramicrobacteria (small cells that maintain their size and volume regardless of the *in vitro* growth condition). These observations guide us to infer that LVBrUMA are ultramicrocells.

Pleomorphism could be an explanation for the ultra-small size of LVBrUMA (Figure 2.1C). Morphological transition is a cellular controlled phenomenon prevalent in bacterial life cycles and correlated to aging (Lange and Hengge-Aronis, 1991; Young, 2006; Kageyama *et al.*, 2008). Bacterial cytokinesis processes are functionally asymmetric. After cellular division, one of the daughter cells inherits a cellular pole that is older than the one inherited by its sibling (Stewart *et al.,* 2005; Ksiazek, 2010). As a consequence, the cell that inherited the older pole drifts in morphology and decreases in size during the end of the log phase and during the stationary phase (Lange and Hengge-Aronis, 1991; Ackermann *et al.,* 2003; Ksiazek, 2010). *Actinobacteria* are relatively well known for developing pleomorphism emerging to ultra-small cells during their life cycle. Cells of the genus *Arthrobacter* were reported to become smaller and spheroidal in the stationary phase due to continuing cell division in the absence of an increase in cell mass (Demkina *et al.*, 2000; Soina *et al.*, 2004). On the other hand, the genus *Microbacterium* does not present a clear rod-coccus life cycle, but old cells do become shorter or spherical (Takeuchi and Hatano, 1998).

Another morphological transition that could be an explanation for the ultra-small size of LVBrUMA is the development of cyst-like resting cells. Non-spore forming bacteria such as *Pseudomonas*, *Micrococcus*, and *Arthrobacter* were described with the capacity to reduce cellular activity, decrease in size, and synthesize a thick cellular envelope in response to unfavorable environmental conditions (Suzina *et al.*, 2006; Mulyukin *et al.*, 2008). Cyst-like resting cells from *Arthrobacter globiformis* isolated from Siberian permafrost were characterized as dwarf forms with 0.2 μm in diameter,

exhibiting a thick cell wall with an external capsular layer and low-electron-dense cytoplasm (Soina *et al.*, 2004). Lake Vida ultra-small microbial cells showed the same morphological characteristics: diameter of ~0.2 μm, low-electron-dense cytoplasm, and external capsular layer surrounding the cell.

Hyperosmotic stress could also be an explanation for the ultra-small size of LVBrUMA coccoid cells due to cell shrinkage in consequence of water loss (Figure 2.1B). The event has been described in Bacteria and Archaea microorganisms. Bacteria with thick cell walls (Gram-positive bacteria), such as *Actinobacteria* and *Firmicutes*, typically do not change morphology in response to hyperosmotic stress because of the rigidity of the cell wall. However, cells with a thin cell wall (Gram-negative bacteria), such as *Proteobacteria*, can shrink and become distorted. The morphological changes of a *Gammaproteobacteria* in response to hyperosmotic stress showed that a culture with 67% of rod-shape cells and 33% of elongated cells became predominated by 80% of spherical cells (Pianetti *et al.*, 2009).

In a study with a halophilic archaea, Fendrihan *et al.* (2012) reported the correlation of the formation of small spheres from rod-shaped cells in response to reduction in external water activity. Rod-shaped cells of *Halobacterium* sp., isolated from halite brine inclusion, turned into ~0.4 μm spheroidal cells after being exposed to high concentrations of NaCl or LiCl. The small spherical cells contained a complete genome and were viable. The authors suggested that they might be in a dormant resting state which, in favorable conditions of proliferation, is reversed to growth as rod-shaped or pleomorphic cells (Fendrihan *et al.,* 2012). Observing morphological changes of Lake

Vida cell cultures in response to osmotic stress and applying TEM in cellular ultrathin sections will help in answering whether the brine isolates are capable of the same kind of transition described above. Osmotic stress studies with *Pseudoalteromonas* and *Herbaspirillum* isolates could also help in understand their high abundance in the brine. Our 16S rRNA clone libraries identified them as abundant in the brine, but no research on morphological variation related to aging or response to osmotic stress have been done in these genera.

Change in morphology is an induced and controlled process. Each of the phylotypes identified in LVBrUMA has its own way to response to the environmental condition imposed by the brine. Therefore, to understand the ultra-small cell size system of Lake Vida brine, firstly we need to understand the cellular machinery capacity of each of its members.

## 2.5. Summary and conclusions

The focus of this study was to describe the nature of the abundant ~ 0.2 μm cell-like particles from the ice-sealed, hypersaline, cold brine from Lake Vida. Electron microscopy observations revealed that the particles are a population of cells with 0.192 ± 0.065 μm in diameter. The presence of an a:Fe-O coat adds structural stability to the cells and in the absence of the inorganic coat, the cells lose their spherical morphology and collapse. Molecular analyses indicated that the ultra-small spherical cells are affiliated with the genera *Pseudoalteromonas* and *Herbaspirillum* in addition to other several

bacterial groups such as Candidate Division TM7, *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Epsilonproteobacteria*, and other genera from *Beta-* and *Gammaproteobacteria.* SEM observations allowed us to observe the morphology of the ultra-small microbial cells and structures such as capsules, external filamentous network connecting the cells, and the presence of a third smaller nanoparticle in the brine with diameter of 0.084 ± 0.063 μm. The use of negative stains such as UA and PTA improved the identification of the cells in the brine by the ability of the cellular membrane to accumulate heavy metal stains. STEM observations of the ultra-small cells confirmed the presence of an electron-dense layer surrounding a less electron-dense region representative of membrane and cytoplasm structures. Epifluorescence microscopy observations combined with the achievement of high molecular weight DNA extracted from the 0.1 to 0.2 μm brine fraction indicate the presence of intact nucleic acid in Lake Vida brine ultra-small microbial cells. Phylogeny investigation based on 16S rRNA gene and culturing efforts did not detect ultramicrobacteria in the brine leading to the conclusion that LVBrUMA are ultramicrocells.

Lake Vida brine ultra-small microbial cells are in accordance with the lowest theoretical size for a free-living prokaryotic cell, which is estimated to be a sphere of 250 to 300 nm in diameter (NRCSSB, 1999). However, their means of survival in the natural brine environment remains under investigation. Studies of extreme environments such as Lake Vida brine and its ultra-small microbial community provide additional impetus to understand the existence of life in freezing brines (potential exohabitats such as in icy

worlds) and to help determine the minimum size of a biological cell expected to be found in these subzero and hypersaline environments.

## 2.6. Acknowledgments

The development of this chapter was supported by Fulbright/CAPES-BRAZIL grant 2163-08-8. Funding for the Lake Vida project was provided by the NSF grant ANT-0739681. SEM work at San Francisco State University was supported by NSF grants 0821619 and 0949176. Thank you to Dr. Andrew Ichimura for the guidance during EM samples preparation and observations. Thank you to Vivian Peng for helping in the RNA extractions. Thank you to Gareth Trubl for continuing the studies on the isolates. Thanks are extended to Dr. Clive Hayzelden for assistance on the SEM and STEM, to Dr. Christiano Garnett Marques Brum, from the Arecibo Observatory, for supervision during data analyses, to K.C. King, from Desert Research Institute, and Tara N. Thomas, from the University of Nevada, for reviewing the chapter and providing extremely valuable feedback. Special thanks to members of the 2010 Lake Vida Expedition team: Dr. Peter Doran, Dr. Alison Murray, Dr. Chris Fritsen, Hilary Dugan, Dr. Bernd Wagner, Dr. Fabien Kenig, Dr. Brian Glazer, Dr. Seth Young, Peter Glenday, and Jay Kyne. Thank you to Raytheon Polar Services for logistic support.

## 2.7. References

Ackermann, M., Stearns, S. C., & Jenal, U. (2003). Senescence in a bacterium with asymmetric division. *Science*, *300*(5627), 1920.

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), 403-410.

Ayala-del-Río, H. L., Chain, P. S., Grzymski, J. J., Ponder, M. A, Ivanova, N., Bergholz, P. W., et al. (2010). The genome sequence of *Psychrobacter arcticus* 273-4, a psychroactive Siberian permafrost bacterium, reveals mechanisms for adaptation to low-temperature growth. *Applied and Environmental Microbiology*, *76*(7), 2304–2312.

Bae, H. C., Cota-Robles, E. H., & Casida, L. E. (1972). Microflora of soil as viewed by transmission electron microscopy. *Applied Microbiology*, *23*(3), 637–648.

Baker, B. J., Comolli, L. R., Dick, G. J., Hauser, L. J., Hyatt, D., Dill, B. D., et al. (2010). Enigmatic, ultrasmall, uncultivated Archaea. *Proceedings of the National Academy of Sciences, 107*(19), 8806–8811.

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 in Microbiology, 8,* 129-138.

Cavicchioli, R., & Ostrowski, M. (2003). Ultramicrobacteria. *In* Battista, J. (Ed.), Encyclopedia of Life Sciences. John Wiley & Sons, West Sussex, UK. pp 1-8.

Chen, Y. G, Tang, S. K., Zhang, Y. Q., Li, Z. Y., Yi, L. B., Wang, Y. X., et al. (2009). *Arthrobacter halodurans* sp. nov., a new halotolerant bacterium isolated from sea water. *Antonie Van Leeuwenhoek, 96*(1), 63-70.

Comolli, L. R., Baker, B. J., Downing, K. H., Siegerist, C. E., & Banfield, J. F. (2009). Three-dimensional analysis of the structure and ecology of a novel, ultra-small archaeon. *The ISME Journal*, *3*(2), 159–167.

Cotton, F. A., Wilkinson, G., Murillo, C. A., & Bochmann, M. (1972). *Advanced Inorganic Chemistry,* 6th ed. John Wiley & Sons Ltd: New York. pp 855-874.

Deatherage, B. L., & Cookson, B. T. (2012). Membrane vesicle release in bacteria, eukaryotes, and archaea: a conserved yet underappreciated aspect of microbial life. *Infection and Immunity*, *80*(6), 1948–1957.

Demkina, E. V, Soina, V. S., El Registan, G. I., & Zviagintsev, D. G. (2000). Reproductive resting forms of *Arthrobacter globiformis*. *Mikrobiologiia*, *69*(3), 377–382.

DeSantis, T. Z., Hugenholtz, P., Keller, K., Brodie, E. L., Larsen, N., Piceno, Y. M. et al. (2006). NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* 34 (Web Server issue), W394-399.

Dmitriev, V. V, Suzina, N. E., Barinova, E. S., Duda, V. I., & Boronin, A. M. (2004). An electron microscopic study of the ultrastructure of microbial cells in extreme biotopes in situ. *Mikrobiologiia*, *73*(6), 832–840.

Dmitriev, V. V, Suzina, N. E., Rusakova, T. G., Gilichinskii, D. A, & Duda, V. I. (2001). Ultrastructural characteristics of natural forms of microorganisms isolated from permafrost grounds of eastern Siberia by the method of low-temperature fractionation. *Doklady Biological Sciences: Proceedings of the Academy of Sciences of the USSR, Biological Sciences Sections*, *378*(6), 304–306.

Dmitriev, V. V., Suzina, N. E., Rusakova, T. G., Petrov, P. Y., Oleinikov, R. R., Esikova, T. Z., et al. (2011). Electron microscopic detection and in situ characterization of bacterial nanoforms in extreme biotopes. *Microbiology*, *77*(1), 39–46.

Dohnalkova, A. C., Marshall, M. J., Arey, B. W., Williams, K. H., Buck, E. C., & Fredrickson, J. K. (2011). Imaging hydrated microbial extracellular polymers: comparative analysis by electron microscopy. *Applied and Environmental Microbiology*, *77*(4), 1254-1262.

Doran, P. T., Fritsen, C. H., McKay, C. P., Priscu, J. C., & Adams, E. E. (2003). Formation and character of an ancient 19-m ice cover and underlying trapped brine in an “ice-sealed” east Antarctic lake. *Proceedings of the National Academy of Sciences, 100*(1), 26–31.

Doran, P. T., Fritsen, C. H., Murray, A. E., Kenig, F., McKay, C. P., & Kyne, J. D. (2008). Entry approach into pristine ice-sealed lakes-Lake Vida, East Antarctica, a model ecosystem. *Limnology and Oceanography Methods, 6,* 542–547.

Duda, V. I., Suzina, N. E., Esikova, T. Z., Akimov, V. N., Oleinikov, R. R., Polivtseva, V. N., et al. (2009). A cytological characterization of the parasitic action of ultramicrobacteria NF1 and NF3 of the genus *Kaistia* on chemoorganotrophic and phototrophic bacteria. *FEMS Microbiology Ecology*, *69*(2), 180–193.

Fendrihan, S., Dornmayr-Pfaffenhuemer, M., Gerbl, F. W., Holzinger, A., Grösbacher, M., Briza, P. et al (2012). Spherical particles of halophilic archaea correlate with exposure to low water activity - implications for microbial survival in fluid inclusions of ancient halite. *Geobiology*, *10*(5), 424-433.

Frankel, R. B. (2003). Biologically induced mineralization by bacteria. *Reviews in Mineralogy and Geochemistry*, *54*(1), 95–114.

Frias, A., Manresa, A., de Oliveira, E., López-Iglesias, C., & Mercade, E. (2010). Membrane vesicles: a common feature in the extracellular matter of cold-adapted antarctic bacteria. *Microbial Ecology*, *59*(3), 476–486.

Geissinger, O., Herlemann, D. P. R., Mörschel, E., Maier, U. G., & Brune, A. (2009). The ultramicrobacterium "*Elusimicrobium minutum*" gen. nov., sp. nov., the first cultivated representative of the termite group 1 phylum. *Applied and Environmental Microbiology*, *75*(9), 2831–2840.

Gilichinsky, D. A. (2001). Permafrost model of extraterrestrial habitat. *In* Astrobiology: The Quest for the Conditions of Life. Horneck, G. (Ed.), Berlin: Springer-Verlag. pp 271–295.

Gilichinsky, D., Rivkina, E., Bakermans, C., Shcherbakova, V., Petrovskaya, L., Ozerskaya, S., et al. (2005). Biodiversity of cryopegs in permafrost. *FEMS Microbiology Ecology*, *53*(1), 117–128.

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., et al. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science, 309*(5738), 1242–1245.

Hahn, M. W., Lünsdorf, H., Wu, Q., Höfle, M. G., Boenigk, J., Stadler, P., et al. (2003). Isolation of novel ultramicrobacteria classified as *Actinobacteria* from five freshwater habitats in Europe and Asia. *Applied and Environmental Microbiology, 69*(3), 1442-1451.

Hahn, M. W., Stadler, P., Wu, Q. L., & Pöckl, M. (2004). The filtration-acclimatization method for isolation of an important fraction of the not readily cultivable bacteria. *Journal of Microbiological Methods*, *57*(3), 379–390.

Hall, A. T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symposium Series*, *41*, 95–98.

Haller, C., Rölleke, S., Vybiral, D., Witte, A., & Velimirov, B. (2000). Investigation of 0.2 µm filterable bacteria from the Western Mediterranean Sea using a molecular approach: dominance of potential starvation forms. *FEMS Microbiology Ecology*, *31*(2), 153–161.

Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., & Stetter, K. O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature*, *417*(6884), 63–67.

Huber, T., Faulkner, G., & Hugenholtz, P. (2004). Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics, 20*(14), 2317-2319.

Iizuka, T., Yamanaka, S., Nishiyama, T., & Hiraishi, A. (1998). Isolation and phylogenetic analysis of aerobic copiotrophic ultramicrobacteria from urban soil. *The Journal of General and Applied Microbiology*, *44*(1), 75–84.

Joshi, H. M., & Toleti, R. S. (2009). Nutrition induced pleomorphism and budding mode of reproduction in *Deinococcus radiodurans*. *BMC Research Notes*, *2*, 123.

Jukes, T. H., & Cantor, C. R. (1969). Evolution of protein molecules. In: *Mammalian protein metabolism*. Academic Press: New York, 21-132.

Kageyama, A., Ã, Y. T., Matsuo, Y., & Adachi, K. (2007). *Microbacterium flavum* sp. nov. and *Microbacterium lacus* sp. nov., isolated from marine environments . *Actinomycetologica*, *21*(2), 53–58.

Kageyama, A., Morisaki, K., Omura, S., & Takahashi, Y. (2008). *Arthrobacter oryzae* sp. nov. and *Arthrobacter humicola* sp. nov. *International Journal of Systematic and Evolutionary Microbiology, 58*(1), 53-56.

Karl, D. M. (1999). Microorganisms in the accreted ice of Lake Vostok, Antarctica. *Science*, *286*(5447), 2144–2147.

Kepner, R. L. Jr., Warthon, R. A. Jr, & Suttle, C. A. (1998). Viruses in Antarctic lakes. *Limnology and Oceanography, 43*(7), 1754-1761.

Kim, S. B., Nedashkovskaya, O. I., Mikhailov, V. V., Han, S. K., Kim, K. O., Rhee, M. S. et al. (2004). *Kocuria marina* sp. nov., a novel actinobacterium isolated from marine sediment. *International Journal of Systematic and Evolutionary Microbiology, 54*(5), 1617-1620.

Krembs, C., & Deming, J. W. (2008). The role of exopolymers in microbial adaptation to sea ice. *In* Psychrophiles: from biodiversity to biotechnology. Margesin, R., Schinner, F., Marx, J. C., Gerday, G. (ed) Springer Verlag: Berlin, Heidelberg, 247-264.

Kulp, A., & Kuehn, M. J. (2010). Biological functions and biogenesis of secreted bacterial outer membrane vesicles. *Annual Review of Microbiology*, *64*, 163–84.

Ksiazek, K. (2010). Bacterial aging: from mechanistic basis to evolutionary perspectives. *Cellular and Molecular Life Sciences, 67*(18), 163-184.

Lane, D. J. (1991). 16S/23S rRNA sequencing. *In* Nucleic Acid Techniques in Bacterial Systematic. Stackebrandt, E., Goodfellow, M. (eds) John Wiley & Sons Ltd: New York, 115–175.

Lange, R., & Hengge-Aronis, R. (1991). Growth phase-regulated expression of *bol*A and morphology of stationary-phase *Escherichia coli* cells are controlled by the novel sigma factor sigma $\sigma^S$. *Journal of Bacteriology*, *173*(14), 4474–4481.

Lindner, A. B., Madden, R., Demarez, A., Stewart, E. J., & Taddei, F. (2008). Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proceedings of the National Academy of Sciences, 105*(8), 3076–3081.

Lisle, J. T., & Priscu, J. C. (2004). The occurrence of lysogenic bacteria and microbial aggregates in the lakes of the McMurdo Dry Valleys, Antarctica. *Microbial Ecology, 47*(4), 427-439.

Litchfield, C. D. (1998). Survival strategies for microorganisms in hypersaline environments and their relevance to life on early Mars. *Meteoritics & Planetary Science*, *33*(4), 813–819.

Lovejoy, C., Bowman, J. P., & Hallegraeff, G. M. (1998). Algicidal effects of a novel marine *Pseudoalteromonas* isolate (Class *Proteobacteria*, Gamma Subdivision) on harmful algal bloom species of the genera *Chattonella*, *Gymnodinium*, and *Heterosigma. Applied and Environmental Microbiology, 64*(8), 2806-2813.

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., & Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *The ISME Journal, 5*(2), 169-172.

Massana, R., Murray, A. E., Preston, C. M., & DeLong, E. F. (1997). Vertical distribution and phylogenetic characterization of marine planktonic *Archaea* in the Santa Barbara Channel. *Applied and Environmental Microbiology, 63*(1), 50-56.

McBroom, A. J., & Kuehn, M. J. (2007). Release of outer membrane vesicles by Gram-negative bacteria is a novel envelope stress response. *Molecular Microbiology*, *63*(2), 545–58.

Médigue, C., Krin, E., Pascal, G., Barbe, V., Bernsel, A., Bertin, P. N., Danchin, A., et al. (2005). Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Research*, *15*(10), 1325–35.

Mikucki, J. A., & Priscu, J. C. (2007). Bacterial diversity associated with Blood Falls, a subglacial outflow from the Taylor Glacier, Antarctica. *Applied and Environmental Microbiology, 73*(12), 4029-4039.

Miot, J., Benzerara, K., Morin, G., Kappler, A., Bernard, S., Obst, M., et al. (2009). Iron biomineralization by anaerobic neutrophilic iron-oxidizing bacteria. *Geochimica et Cosmochimica Acta*, *73*(3), 696–711.

Miteva, V. I., Brenchley, J. E., & Core, I. (2005). Detection and Isolation of ultrasmall microorganisms from a 120,000-Year-Old Greenland glacier ice core. *Applied and Environmental Microbiology*, *71*(12), 7806-7818.

Miyoshi, T., Iwatsuki, T., & Naganuma, T. (2005). Phylogenetic characterization of 16S rRNA gene clones from deep-groundwater microorganisms that pass through 0.2- micrometer-pore-size filters. *Applied and Environmental Microbiology, 71*(2), 1084-1088.

Mondino, L. J., Asao, M., & Madigan, M. T. (2009). Cold-active halophilic bacteria from the ice-sealed Lake Vida, Antarctica. *Archives of Microbiology, 191*(10), 785-790.

Mosier, A. C., Murray, A. E., & Fritsen, C. H. (2007). Microbiota within the perennial ice cover of Lake Vida, Antarctica. *FEMS Microbiology Ecology, 59*(2), 274-288.

Mulyukin, A. L., Suzina, N. E., Duda, V. I., & El'-Registan, G. I. (2008). Structural and physiological diversity among cystlike resting cells of bacteria of the genus *Pseudomonas*. *Mikrobiologiia*, *77*(4), 512-523.

Murray, A. E., Hollibaugh, J. T., & Orrego, C. (1996). Phylogenetic compositions of bacterioplankton from two California estuaries compared by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Applied and Environmental Microbiology, 62*(7), 2676-2680.

Murray, A. E., Kenig, F., Fritsen, C. H., McKay, C. P., Cawley, K. M., Edwards, R., Kuhn, E., et al. (2012). Microbial life at -13ºC in the brine of an ice-sealed Antarctic lake. *Proceedings of the National Academy of Sciences, 109*(50), 20626-20631.

Knoll, A., Osborn, M. J., Baross, J., Berg, H. C., Pace, N. R., & Sogin, M. (1999). Size limits of very small microorganisms: proceedings of a workshop. National Research Council Space Studies Board. National Academy Press, Washington, D.C.

Naganuma, T., Miyoshi, T., & Kimura, H. (2007). Phylotype diversity of deep-sea hydrothermal vent prokaryotes trapped by 0.2- and 0.1-microm-pore-size filters. *Extremophiles: Life under Extreme Conditions*, *11*(4), 637–46.

Nakai, R., Abe, T., Takeyama, H., & Naganuma, T. (2011). Metagenomic analysis of 0.2-μm-passable microorganisms in deep-sea hydrothermal fluid. *Marine Biotechnology, 13*, 900-908.

Nedelkova, M., Merroun, M. L., Rossberg, A., Hennig, C., & Selenska-Pobell, S. (2007). *Microbacterium* isolates from the vicinity of a radioactive waste depository and their interactions with uranium. *FEMS Microbiology Ecology, 59*(3), 694-705.

Nevot, M., Deroncele, V., López-Iglesias, C., Bozal, N., Guinea, J., & Mercade, E. (2006a). Ultrastructural analysis of the extracellular matter secreted by the psychrotolerant bacterium *Pseudoalteromonas antarctica* NF3. *Microbial Ecology*, *51*(4), 501–507.

Nevot, M., Deroncelé, V., Messner, P., Guinea, J., & Mercadé, E. (2006b). Characterization of outer membrane vesicles released by the psychrotolerant bacterium *Pseudoalteromonas antarctica* NF3. *Environmental Microbiology*, *8*(9), 1523–1533.

Nyström, T. (2007) A bacterial kind of aging. *PLoS Genetics, 3*(12), 2355-2357.

Peeters, K., Ertz, D., & Willems, A. (2011). Culturable bacterial diversity at the Princess Elisabeth Station (Utsteinen, Sør Rondane Mountains, East Antarctica) harbours many new taxa. *Systematic and Applied Microbiology*, *34*(5), 360–7.

Pianetti, A., Battistelli, M., Citterio, B., Parlani, C., Falcieri, E., & Bruscolini, F. (2009). Morphological changes of *Aeromonas hydrophila* in response to osmotic stress. *Micron, 40*(4), 426-433.

Ponder, M. A., Thomashow, M. F., & Tiedje, J. M. (2008). Metabolic activity of Siberian permafrost isolates, *Psychrobacter arcticus* and *Exiguobacterium sibiricum*, at low water activities. *Extremophiles: Life under Extreme Conditions*, *12*(4), 481–490.

Priscu, J. C. (1999). Geomicrobiology of subglacial ice above Lake Vostok, Antarctica. *Science*, *286*(5447), 2141–2144.

Rappé, M. S., Connon, S. A, Vergin, K. L., & Giovannoni, S. J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature*, *418*(6898), 630–633.

Rutz, B. A., & Kieft, T. L. (2004). Phylogenetic characterization of dwarf archaea and bacteria from a semiarid soil. *Soil Biology and Biochemistry*, *36*(5), 825–833.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B. et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology, 75*(23), 7537-7541.

Sheridan, P. P. (2003). *Rhodoglobus vestalii* gen. nov., sp. nov., a novel psychrophilic organism isolated from an Antarctic Dry Valley lake. *International Journal of Systematic and Evolutionary Microbiology*, *53*(4), 985–994.

Soina, V. S., Mulyukin, A. L., Demkina, E. V, Vorobyova, E. A., & El-Registan, G. I. (2004). The structure of resting bacterial populations in soil and subsoil permafrost. *Astrobiology*, *4*(3), 345–358.

Stewart, E. J., Madden, R., Paul, G., & Taddei, F. (2005). Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biology, 3*(2), e45.

Suzina, N. E., Mulyukin, A. L., Dmitriev, V. V., Nikolaev, Y. A., Shorokhova, A. P., Bobkova, Y. S., Barinova, E. S., et al. (2006). The structural bases of long-term anabiosis in non-spore-forming bacteria. *Advances in Space Research, 38*(6), 1209–1219.

Takeuchi, T., & Hatano, K. (1998). Union of the genera *Microbacterium* Orla-Jensen and *Aureobacterium* Collins *et al*. in a redefined genus *Microbacterium*. *International Journal of Systematic Bacteriology, 48*(3), 739-747.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution, 28*(10), 2731-2739.

Torrella, F., & Morita, R. Y. (1981). Microcultural study of bacterial size changes and microcolony and ultramicrocolony formation by heterotrophic bacteria in seawater. *Applied and Environmental Microbiology*, *41*(2), 518–527.

Ward, B. B., & Priscu, J. C. (1997). Detection and characterization of denitrifying bacteria from a permanently ice-covered Antarctic Lake. *Hydrobiologia, 347*, 57-68.

Young, K. D. (2006). The selective value of bacterial shape. *Microbiology and Molecular Biology Reviews, 70*(3), 660-703.

Zhang, D. C., Li, H. R., Xin, Y. H., Chi, Z. M., Zhou, P. J., & Yu, Y. (2008). *Marinobacter psychrophilus* sp. nov., a psychrophilic bacterium isolated from the Arctic. *International Journal of Systematic and Evolutionary Microbiology*, *58*(6), 1463–1466.

# Chapter III

## Microbial assemblage within deep ice and deep sediment layers from a frozen lake: Lake Vida, Antarctica


### Abstract

Lake Vida – located in the McMurdo Dry Valleys, Antarctica – has a thick (27+ m) ice cover from the top to the bottom of the lake that seals a cryogenic brine reservoir within the lake ice below 16 m. A core retrieved from the lake in 2010 reached a depth of nearly 27 m, 10.5 m deeper than previous coring efforts, revealing the existence of thick sediment layers banding the ice below 21 m. By applying Illumina tag sequencing (iTag) targeting the 16S rRNA gene and RNA (as cDNA) in ice and sediment samples retrieved from below 18 m depth, we discovered that the lower portion of the lake ice harbors a diverse and cell-rich microbial assemblage with cell counts ranging from $3.73 \times 10^4$ to $8.58 \times 10^6$ cells $mL^{-1}$. This assemblage is composed of bacteria that are affiliated with

organisms known to be specialized in the reduction and oxidation of sulfur compounds, and known to be able to degrade complex polymers. *Proteobacteria, Bacteroidetes*, *Actinobacteria*, and *Firmicutes* were the most abundant phyla detected among the DNA iTag libraries. We observed that composition of the microbial assemblage differed depending on the presence of sediment particles, total dissolved solids (TDS), total carbon, $SO_4^{2-}$, and $Na^+$ concentration. Comparison among the DNA iTag libraries revealed that the ice segments were dominated by heterotrophic (such as *Lutibacter* and *Desulforomonas*) and chemolithoautotrophic microorganisms (such as *Sulfurovum* and *Desulfocapsa*), and the sediment segments were dominated by heterotrophic microorganisms (such as *Cellulomonas* and *Conexibacter*). cDNA iTag libraries from the microbial assemblages within the sediment layers indicated that the major components carrying metabolic capacity within the sediment layers are mostly bacteria affiliated with the genera *Pseudoalteromonas* and *Vibrio,* often identified in Antarctic marine environments and adapted to low temperatures and high salinity. Diversity index analyses throughout the core from 18 to 27 m depth indicated that there is a high diversity and high variability of microbial assemblages that roughly cluster with the sample type. Therefore, the concentration of sediment particles, $Cl^-$, and $SO_4^{2-}$ appear to be an important factor controlling distribution of microbial assemblages within Lake Vida's lower ice and sediment layers within the ice.

## 3.1. Introduction

The McMurdo Dry Valleys, the largest ice-free area in Antarctica, contain more than 20 permanent lakes and ponds, mostly remnants of larger and deeper glacial lakes that occupied the valleys thousands of years ago (Doran *et al.,*1994; Hall *et al.,* 2002; 2010). One of them, Lake Vida, located in Victoria Valley, is thought to be a remnant of Glacial Lake Victoria, which was at least 200 m deep and covered more than 100 $km^2$ of the valley between 20,000 and 8,600 $^{14}C$ yr BP (Hall *et al.*, 2002). Although Calkin and Bull (1967) theorized that the lake was frozen to its base, Doran *et al.* (2003) postulated that Lake Vida was an ice-covered lake with a saline water basin under 20 m of perennial ice. Ice core drilling in 2010 reached nearly 27 m, and no liquid water column was found. However, the thick 27+ m ice cover contained a brine reservoir within the ice below 16 m depth (Murray *et al.*, 2012) and seven sediment layers banding the ice below 21 m. The drilling process stopped at nearly 27 m depth, and it is not certain if the bottom of the lake were reached since ice was present at the bottom of the core.

Life in Lake Vida is restricted to microorganisms. Microbial assemblages from the upper portion the lake ice (from 4.8 m to 15.9 m depth) and from the sealed brine collected below 16 m were described through microscopic and molecular approaches including 16S rRNA gene clone libraries (Mosier *et al.*, 2007; Murray *et al.*, 2012; chapter II in this dissertation). Those studies revealed that the ice and brine harbored a diverse bacterial assemblage with at least 14 phyla detected. Lake Vida surface ice (4.8 m), which contained high amounts of sediment particles, was dominated by *Actinobacteria* (42%) and *Bacteroidetes* (13%; Mosier *et al*., 2007), which are bacteria

specializing in carbon mineralization, often found in microbial mats and soils (Niederberger *et al*., 2008; Tang *et al.*, 2013). The 15.9 m deep clear ice was dominated by *Gammaproteobacteria* (52%; *Marinobacter* sp. as the major genus) and *Bacteroidetes* (15%; Mosier *et al*., 2007). *Gammaproteobacteria* comprise a class of heterotroph, chemoautotroph, and fermentation-capable microorganisms that are physiologically diverse. Similarly, Lake Vida brine is inhabited by a reasonably diverse microbial assemblage dominated by *Gammaproteobacteria* (39% affiliated with the genus *Marinobacter* sp.), *Lentisphaera* (15%), and *Epsilonproteobacteria* (15%; Murray *et al.,* 2012). The brine ultra-small microbial assemblage (cells ~0.2 μm diameter) is mainly composed of *Gammaproteobacteria* (38% affiliated with *Pseudoalteromonas* sp.) and *Betaproteobacteria* (35% affiliated with *Herbaspirillum* sp.; chapter II). *Betaproteobacteria* and *Bacteroidetes* are commonly found in aquatic environments and soils enriched with organic compounds (Kirchman, 2002). Eukarya rRNA gene sequences were detected in the upper portion of the ice (4.8 m and 15.9 m) but not in the lake brine. Archaeal rRNA gene sequences were neither detected in the upper portion of the ice nor in the brine (Mosier *et al.*, 2007; Murray *et al.*, 2012).

Microbial life in the Dry Valleys has been described in glaciers (Mikucki and Priscu, 2007), lakes, lake ice, streams (Laybourn-Parry and Pearce, 2007; Paerl and Priscu, 1998), soils, permafrost, and rocks (Gilichinsky *et al.*, 2007; Pointing *et al.*, 2009; Antibus *et al*., 2012). The most abundant taxa vary depending on the environment. For example, dry valley soils are dominated by *Acidobacteria*, *Actinobacteria*, and *Bacteroidetes* (Cary *et al.*, 2010); surface layers of streams and perennially ice-covered

lakes are dominated by *Cyanobacteria* (Alger *et al*.; 1997; Priscu *et al.*, 2005); and the basal ice from Taylor Glacier is dominated by *Firmicutes* (Christner *et al*., 2010). Blood Falls – a hypersaline, iron-rich, liquid brine that drains from under Taylor Glacier – is dominated by *Gammaproteobacteria, Betaproteobacteria*, *Deltaproteobacteria*, and *Bacteroidetes* (Mikucki and Priscu, 2007).

The Lake Vida deep ice cover intercalated by frozen sandy layers is a recently discovered environment that has not yet been biologically characterized. Sediment layers exist in other dry valley lake ice covers as a result of aeolian deposition followed by summer melting (Squyres *et al.*, 1991; Priscu, 1998). However, deep sediment layers in this region have been described only in glacier basal ice and permafrost (Doyle *et al.,* 2013). To advance our knowledge in life within Lake Vida, we characterized the composition and potential metabolic function of microbial assemblages through the lower portion of the lake ice, from 18 to 27 m depth, in relation to physiochemical properties of the ice core. Illumina parallel tag sequencing (iTag; Bartram *et al.*, 2011; Degnan and Ochman, 2012) was used to generate millions of short (250 bp) reads from the variable region V4 of the 16S rRNA gene and rRNA (through cDNA). Our goal was to provide assessments of microbial diversity, community structure, and potential metabolic function in an unexplored environment at a resolution two orders of magnitude greater than results determined by clone libraries.

## 3.2. Materials and methods

### 3.2.1. Ice core sampling and processing

Ice core retrieval from Lake Vida was conducted according to Doran *et al*. (2008). An electric 15 cm diameter SideWinder drill was used to retrieve a nearly 27 m ice core in the center of Lake Vida (77.39ºS 161.93ºE) in November 2010. Sections of the 27 m core (16–20 m and 23–25 m) were lost during core retrieval and were not used in this investigation, since depths could not be ascertained ("missing" sections in Figure 3.1). The core was stored at -20°C, photographed, and stored at -80°C until it was split for archiving, chemical analyses, and biological analyses. Eight core sections of the lake ice between 18.5 and 27 m were sampled for biological analyses and are the core sections under investigation here (Figure 3.1 A). The core contained a total of seven sediment layers banding the ice below 21 m. Four of them presented lengths ranging from 11 to 15 cm. The eight core sections analyzed in this study were subdivided into 18 segments and analyzed as individual samples: eight ice segments, five sediment segments, and five ice/sediment transition segments (Figure 3.1 B).

Ice and sediment sectioning and decontamination processes were conducted in a sterilized laminar flow hood housed inside a 2°C cold room. The laminar flow hood was cleaned with 5% sodium hypochlorite, then 3% hydrogen peroxide, and sterilized with UV radiation. Facemasks and sterile gloves were worn during the process. To prevent carryover contamination, gloves were changed, the hood was wiped with 3% hydrogen peroxide and ethanol, and the set of tools used was replaced by sterilized clean tools

between handling of each sample. Sterilized aluminum foil also was used to cover the surface inside the hood where the samples were deposited and was changed before handling the next sample.

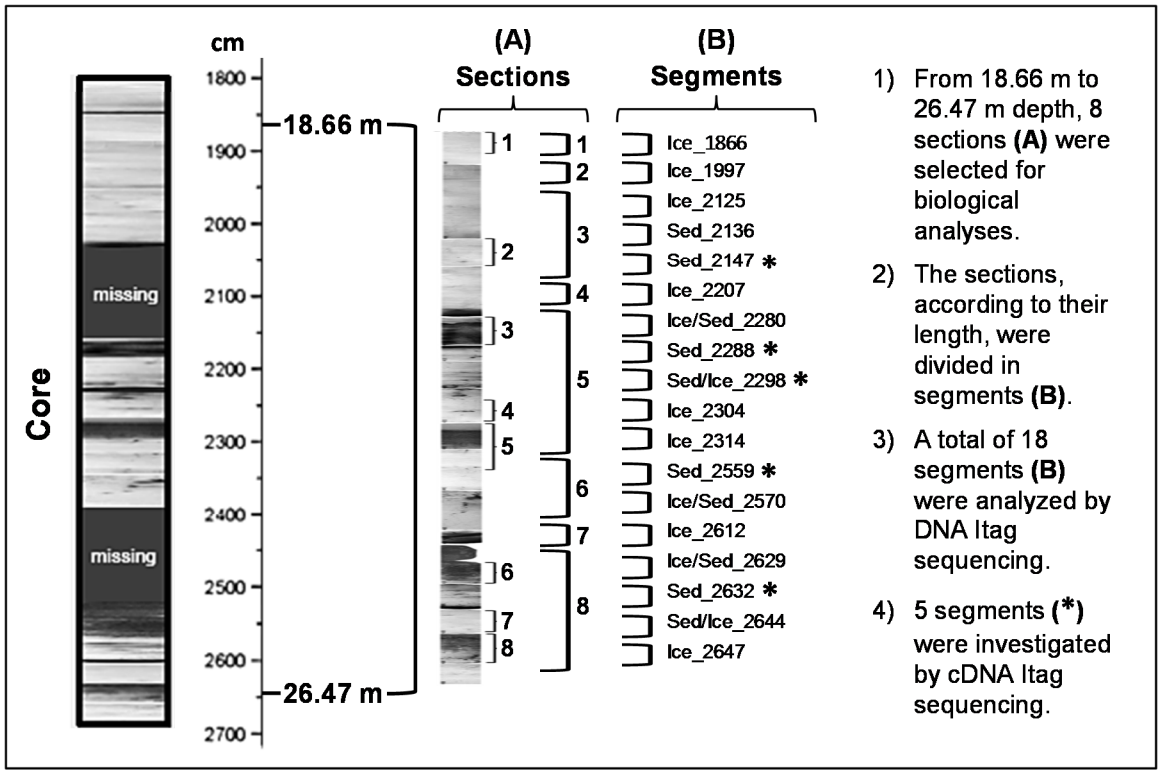



**Figure 3.1.** Samples and depths of Lake Vida core analyzed in this study. On the left, representation of the lower core portion (from 1800 to 2681 cm) recovered in November 2010. The dark bands and dark areas in the core are indicative of the presence of solid particles within the ice. The "missing" boxes indicate sections of unknown precise depths that were not added to the core mosaic. Eight sections of the core were selected for biological characterization. (A) shows location of the sections in the core and (B) shows the 18 segments originating from the sections and their IDs (Ice, Sed, Ice/Sed, or Sed/Ice indicate the nature of sample in the segment and the number indicates the depth in cm from which the sample was collected). The numeric list at the right explains the process illustrated in the figure. iTag: Illumina tag sequencing technology.

Decontamination of the ice and sediment samples was conducted independently. Ice segments were decontaminated by rinsing the sample (using wash bottles) with 3% hydrogen peroxide, followed by 70% ethanol solution, and a final rinse with 100% ethanol. Sediment segments were decontaminated through mechanical ablation using sterilized razor blades to scrape 5 mm from the outer surfaces of each segment. Each of the six sides of the segment was scraped with a clean, sterilized blade. After the decontamination process, cleaned samples were stored separately, inside sterile Whirl-Pak® bags.

### 3.2.2. Geochemical analyses

Nutrients within the Lake Vida ice core were analyzed from 5 cm subsamples. The ice subsamples were rinsed with deionized MilliQ water to remove external contamination and melted at room temperature. An ICS-1500 Dionex ion chromatograph was used to determine the concentration of anions ($F^-$, $Cl^-$, $NO_2^-$, $Br^-$, $NO_3^-$, $PO_4^{3-}$, and $SO_4^{2-}$) and cations ($Li^+$, $Na^+$, $NH_4^+$, $K^+$, $Mg^{2+}$, and $Ca^{2+}$). Anions were determined using an AS14A column, and cations were determined using a CS12A column. Ions were eluted in 20 mM of methanesulfonic acid (MSA). Total dissolved solids (TDS) concentration was calculated by summing the mass contribution of individual ions (Mora *et al.,* 2004)

Sediment subsamples of 1 cm length were analyzed for grain size, water content, and organic and inorganic carbon content. The biological analyses were conducted with

longer fragments (max of 20 cm in length) to secure enough material for further analyses after the decontamination procedure. Therefore, the biological segments had more than one chemical result (*e.g.,* sample Ice_1866, which is 15 cm long, had three ice subsamples of 5 cm each for chemical analysis).

### 3.2.3. Cell enumeration and scanning electron microscopy

Cells from ice segments were counted from 5, 10, or 15 mL of ice melted at 4ºC in sterile Whirl-Pak® bags. Cells were fixed with glutaraldehyde (1% final concentration) for five hours, stained with SYBR Gold DNA stain (Invitrogen, Carlsbad, CA, USA) for 20 minutes, and filtered onto 0.22 µm pore-sized black polycarbonate filters (Millipore, Inc. Boston, MA, USA). The cells were observed and counted using an epifluorescence microscope.

Cell enumeration of sediment segments was performed by counting the cells mechanically removed from 1 g of wet sediment. The process included a primary step of detaching the cells from the sediment grains. One gram of sediment was fixed with 1 mL of 1% glutaraldehyde for at least six hours and no more than 20 hours at 4ºC. During the fixation step, the sample was frequently mixed using a vortex. After six hours of decantating, the supernatant from the sample was collected and transferred to a 15 mL tube. The sample was rinsed five times with 1 mL of phosphate buffered saline (PBS) 0.01 mM, pH 7.4, mixed by vortex for 10 seconds, sonicated using an ultrasonic bath (35 khz) for one minute at room temperature, and allowed to settle at 4ºC for five min. The

supernatant was then collected and added to the 15 mL tube. As a final rinsing step, the sample was spun for five seconds, and the supernatant was collected. An aliquot of 250 µL of the sample was stained with SYBR Gold for 20 minutes, filtered onto 0.22 µm pore-sized black polycarbonate filters and observed under the epifluorescence microscope.

Scanning electron microscopy (SEM) of melted ice from samples Ice_2612 and Ice_2647 was conducted to observe integrity of the cells in Lake Vida deeper ice. Melted ice was fixed with 0.5% anoxic glutaraldehyde (v/v final concentration) and filtered onto 0.22 µm pore-sized polycarbonate. Cells were dehydrated by an ethanol series (30, 50, 70, 90, and 100%), dried in air, and coated with 1–3 nm of iridium or platinum to prevent charging during image acquisition.

### 3.2.4. DNA preparation

Ice, sediment, and transition zones (ice to sediment [Ice/Sed] and sediment to ice [Sed/Ice]) were treated differently for DNA extractions. DNA extraction from the ice segments was conducted as follows. The ice samples were melted inside an aseptic bag at 4°C and filtered on 0.22 µm pore-sized filters. Cells deposited on filters were resuspended in sucrose lysis buffer (SLB); chemically lysated with lysozyme, proteinase-K, and SDS; then DNA was purified through phenol:chloroform extractions (Massana *et al.*, 1997). The starting volume of melted ice used from each sample ranged from 58 mL to 108 mL. For DNA extraction from the sediment segments, approximately 1 g of

sediment was used as starting material, and DNA was extracted using the Power Soil DNA Isolation Kit (MO BIO) following manufacturer instructions. Samples within the transition zones were split into water and sediment phases. The water was separated from the sediment by melting and centrifugation, or by mechanical abrasion. DNA was extracted from each phase (water and sediment) using the same protocols described above and then combined as one transition segment sample. Before sending the DNA for iTag sequencing, the samples were quantified using the Quant-iT PicoGreen dsDNA Assay kit (Invitrogen, Carlsbad, CA, USA) and tested for amplification with the same bacterial/archaeal primers used for iTag sequencing. The amplification consisted of 35 cycles with annealing temperature of 50°C using the primers 515F (GTGCCAGCMGCCGCGG) and 806R (GGACTACHVGGGTWTCTAAT; Caporaso *et al.*, 2010; 2012).

### 3.2.5. RNA and cDNA preparation

Four sediment segments and one sediment/ice transition segment were selected for RNA extraction. RNA was extracted from approximately 1 g of sediment each from samples Sed_2147, Sed_2288, Sed/Ice_2298, Sed_2559, and Sed_2632 (sample ID with * in Figure 3.1). The extraction was conducted using a combination of Power Soil DNA Isolation Kit (MO BIO), Totally RNA protocol (Life Technologies), and RNeasy Clean up kit (Qiagen). The sediment was initially resuspended in 1 mL of SLB and mixed using a vortex. The mix was then centrifuged at 6000 rpm for 30 sec to separate the SLB from

the sediment. At this step, the sample was split. SLB was recovered, and RNA was extracted from cells in suspension using a modified Totally RNA protocol as follow.

RNA from cells within the sediment was extracted using the Power Soil DNA Isolation Kit followed by the modified Totally RNA protocol. Both extractions from the same sample were put through the RNeasy Clean-up kit as a final purification step and then combined. The modified Totally RNA protocol adjustments were as follows: The solution containing the cells (SLB or buffer C3 from MO BIO kit) passed through two phenol:chloroform extractions tailed by two chloroform extractions to ensure recovery of clean genomic material. Linear polyacrylimide (LPA) was added to the first isopropanol precipitation step. After extraction, two DNase I treatments (Ambion) were performed to ensure no genomic DNA contamination in the sample. The sediment fraction initially separated from the SLB was extracted using the MO BIO kit. Buffer C1 was added to the sediment and microtubes were bead beat for 30 sec at 4200 rpm and centrifuged at 10,000g for 30 sec. Supernatant was transferred to a new microtube, buffer C2 was added, and the sample was mixed and incubated for five min. After incubation, the sample was centrifuged at 10,000g for five min, and the supernatant was transferred to a new microtube where Buffer C3 was added, and the sample was mixed and incubated for five min. After incubation, the sample was centrifuged at 10,000g for five min, and the supernatant was recovered to a new microtube and processed with the modified Totally RNA extraction protocol described above.

The amount of total RNA extracted was below the detection levels of our current nucleic acid assays, Quant-iT RiboGreen RNA Assay kit (Invitrogen, Carlsbad, CA,

USA). To bypass this issue, total RNA (7.5 μL starting material) was amplified using the MessageAmp II-Bacteria kit (Life Technologies; Frias-Lopez *et al.*, 2008; Shi *et al.*, 2009). Briefly, total RNA was polyadenylated using *Escherichia coli* poly(A) polymerase. Polyadenylated RNA was converted to double-stranded cDNA via reverse transcription primed with an oligo(dT) primer containing a promoter sequence for T7 RNA polymerase and a recognition site for the restriction enzyme BpmI (T7-BpmI-(dT)16 VN, GCCAGTGAATTGTAATACGACTCACTATAGGGGCGACTGGAG TTTTTTTTTTTTTTTTVN) (Stewart *et al.*, 2010). Then, cDNA was transcribed *in vitro* at 37ºC yielding large quantities (2–50 μg) of single-stranded antisense RNA. Amplified RNA (~500 ng aliquot) was then reverse transcribed to double-stranded cDNA using the SuperScript III First-Strand Synthesis System (Life Technologies) with priming via directed 16S rRNA gene primer, using the archaeal/bacterial rRNA gene reverse primer 806R (Takai *et al.*, 2000) for first-strand synthesis. Then, the SuperScript Double-Stranded cDNA synthesis kit (Life Technologies) was used for second-strand synthesis. cDNA was purified with the MinElute PCR purification kit (Qiagen) and prepared for sequencing.

### 3.2.6. Illumina tag sequencing and pre-processing of reads

Illumina tag sequencing was conducted at the DOE Joint Genome Institute (Walnut Creek, CA) according to internal protocols. The V4 hypervariable region (from position 515 to 806) of the 16S rRNA gene from 18 DNA samples and the rRNA (as

cDNA) from five samples (average size of 291 bp) was amplified using the primers 515F and 806R (Caporaso *et al.*, 2012). Sequencing was performed on an Illumina MiSeq benchtop sequencer with a 2 x 250 bp reads configuration (Caporaso *et al.*, 2012). The 23 raw files generated were initially processed by the JGI iTags pipeline through which contaminants, barcode, and primers were removed, and sequences assembled. In summary, after removing the contaminants (e.g., adapters, PhiX reads, and unpaired reads), reads were trimmed to 165 bp and assembled to reconstitute the original 16S rRNA gene amplicon by using FLASH (Magoc and Salzberg, 2011). After assembling the reads, primers were removed, and reads were trimmed from both 5' and 3' ends using a 20 bp sliding window quality threshold having a mean quality of 30. The trimmed assembled reads were filtered for quality where reads that contained more than 5 Ns and 10 nucleotides below quality 15 were discarded. Reads that were not assembled by FLASH and passed by quality control were added to the high-quality/filtered assembled reads. High-quality/filtered reads added up to a total of 1,594,687 reads.

The 1,594,687 reads were run through pre-processing steps before community structure and composition analyses. FastqMcf was used to remove poor quality reads (Aronesty, 2011). FastqMcf was run with a minimum sequence length of 100 bp, ambiguous base of 0.1%, sliding window size of four, and quality threshold of 32. The quality of the sequences was checked before and after cleaning the reads. Chimera sequences were detected and removed using DECIPHER (Wright *et al.*, 2012). After quality control and removal of chimeras, reads were aligned using Clustal Omega (Sievers *et al.*, 2011). The 23 aligned libraries generated were manually checked for a

final filtering and trimming of reads for further microbial assemblage analyses. The average size of the sequences used in the analyses was 214 bp.

### 3.2.7. Microbial assemblage structure and composition

A total of 872,627 high quality reads were analyzed using the program mothur v.1.28.0 applying the standard operating procedure (SOP) described by the software developers (Schloss *et al.*, 2009; Schloss and Westcott, 2011). The 23 iTag libraries were processed separately but identically. A two pathway analysis was used to interpret the reads: phylotype-based and operational taxonomic unit (OTU)-based (Figure 3.2). For phylotype-based analyses, unique reads from each library from all the high quality data generated (data not normalized) were aligned to the RDP 16S rRNA training set 9 (http://rdp.cme.msu.edu) and taxonomically classified. A bootstrap cutoff ≥ 80% identity was used for class classification, and a bootstrap cutoff ≥ 95% identity was used for genus classification. Reads that were rated below the threshold values were assigned as unclassified (Claesson *et al*., 2009; Wang *et al.*, 2012). After taxonomic characterization, the 10 most abundant genera were ranked considering their relative abundance in each library.

OTU analysis was conducted after normalization of the samples. To normalize the number of iTags reads between samples, reads were randomly re-sampled to the number of reads in the iTag library with fewest reads (13,043 reads) using Daisychopper v. 0.6 (Gilbert *et al*., 2009). OTU-based analyses were calculated considering 0.03 distance

units as cutoff, and taxonomical assignment of OTUs was performed using the RDP 16S rRNA training set 9 (Figure 3.2).

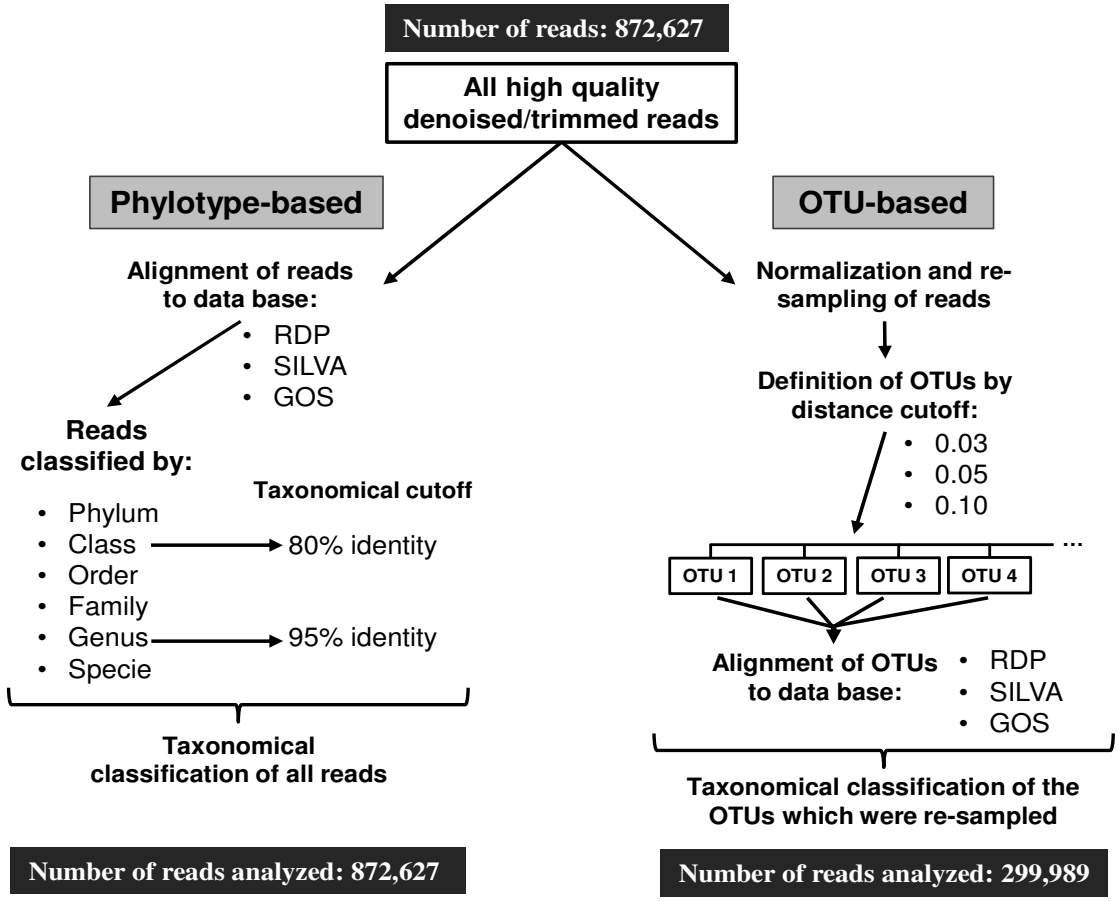



**Figure 3.2.** The two pathways applied to analyze the iTag reads: Phylotype-based and OTU-based.

Diversity metrics were calculated by mothur (Schloss *et al.*, 2009) using the OTUs generated from the 0.03 unit distance cutoff after data normalization. The alpha diversity (diversity within each sample) – such as Chao1 estimator of minimum number of OTUs expected (Chao, 1984), nonparametric Shannon diversity index (Shannon,

1948), and abundance-based coverage estimator (ACE; Chao and Lee, 1992) – were calculated for each iTag library separately. The rarefaction curves were the only estimator calculated based on all high quality reads generated. The beta diversity (between samples diversity) was investigated through Jaccard coefficient (membership comparison based on presence/absence of the OTU; Jaccard, 1901) and $Theta_{yc}$ coefficient (structure comparison considering the membership and abundance of the OTUs; Yue and Clayton, 2005). For the assemblage structure comparisons, reads were clustered using the average neighbor algorithm (Schloss and Westcott, 2011). Data plots were generated using mothur and the software Grapher 4.01 (Golden Software, CO, US).

### 3.2.8. Statistical analyses

Principal component analysis (PCA) was applied to test correlation of the ice and sediment environmental variables and the OTU assemblage composition through the core. Linear associations were calculated by Spearman's correlation coefficient. The PCA exploratory analysis calculated new synthetic variables (principal components) which accounted for much of the variance in the original data as possible (Hotelling, 1933). Correlation among the OTU and the environmental parameters was carried out in SPSS® Statistics (version 22, IBM®) software considering the extraction of the two principal components. The PCA results were displayed as biplots using Grapher 4.01 software (Golden Software, CO, US). Spearman correlation rank was used assuming that the variables relationship was not linear. For analyses of the significance of the correlation rank results, Student T-Test considering independent and dependent means was applied.

## 3.3. Results and discussion

### 3.3.1. Geochemical characterization of the ice and sediment layers

The 27 m core retrieved in 2010 revealed three sediment layers (SL) between the lake surface and 9 m depth, almost no sediment particles between 9 and 13 m, and 10 SL below 13 m. The SL investigated in the present study were from depths 2136–2150 cm (SL2136–2150 cm, 14 cm in length), 2288–2303 cm (SL2288–2303 cm, 15 cm in length), 2559–2570 cm (SL2559–2570 cm, 11 cm in length) and 2630–2644 cm (SL2630–2644 cm, 15 cm in length).

A clear distinction between the core geochemistry above and below SL2288–2303 cm was evident. A significant shift in concentrations of sulfate ($P < 0.00001$), calcium ($P < 0.00001$), and sodium ($P = 0.000031$) was identified at the depth around 2288 cm (Figures 3.3 and 3.4). The ice from 16.22 m to SL2288–2303 cm was dominated by chloride as the major anion, which is the anion that dominates the ice from SL2288–2303 cm up to the surface of the lake (data not shown). Below SL2288–2303 cm, the major anion shifted to sulfate, which was detected in concentrations ranging from 92 μM up to 242 mM with an average of 31.5 mM. The higher concentrations of sulfate below SL2288– 2303cm were accompanied with higher concentrations of sodium (Spearman $r = 0.88$, $P < 0.00001$) and calcium (Spearman $r = 0.53$, $P < 0.000395$). These ions are highly abundant in the McMurdo Dry Valleys because of the high quantity of the minerals mirabilite ($Na_2SO_4 \bullet 10H_2O$), thenardite ($Na_2SO_4$), and gypsum ($CaSO_4 \bullet 2H_2O$; Keys and Williams, 1981). We hypothesize that two events could have happened in the

past that could have significantly increased the concentration of $SO_4^{2-}$, $Na^+$, and $Ca^{2+}$ within the ice below SL2288–2303 cm: past weathering and/or ancient biological activity.

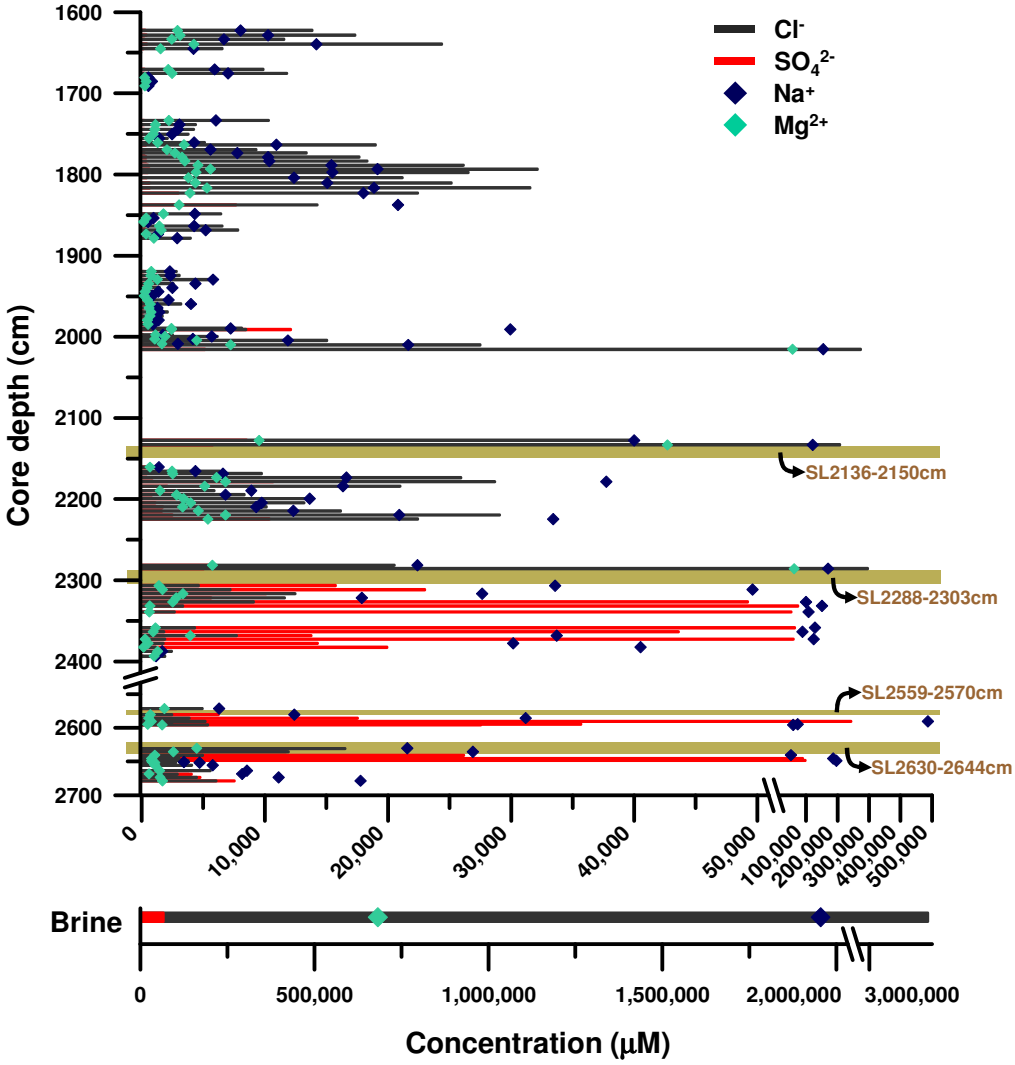



**Figure 3.3.** Concentrations of the major anions (chloride and sulfate) and major cations (sodium and magnesium) identified in the Lake Vida ice core from 1622 cm to 2681 cm depth. Location of the sediment layers (SL) analyzed by iTag libraries in the present study are shown in brown with their respective depths (SL2136–2150 cm, SL2288–2303 cm, SL2559–2570 cm, and SL2630–2644 cm). The concentration of chloride, sulfate, sodium and magnesium in the Lake Vida brine is shown in a separate graph because of its higher concentrations. The blank regions in the graph, with no plotted data, resulted from missing sections of the core.

Magnesium, ammonium, and potassium concentrations varied through the core but did not present any significant changes correlated to depth ($P = 0.09$, $P = 0.68$, $P = 0.09$, respectively; Figures 3.6 and 3.7). Ammonium concentrations appeared to be higher below SL2136–2150 cm. Nitrate was detected only between 1622 cm and 1960 cm with a concentration average of 7 μM (Figure 3.7). Nitrite and phosphate were not detected within the ice core. Fluoride, bromide, and lithium were detected only in a few subsamples. Fluoride presented an increase in concentration from 11 μM to a maximum of 509 μM around depth 1980 cm to 2015 cm, 13 μM to 125 μM from 2306 to 2326 cm, and a concentration peak of 292 μM around 2645 cm, below SL2630–2644 cm. Bromide was not detected below 1840 cm, and lithium – which was sporadically detected in low concentrations within the core – was detected in higher concentrations below SL2630–2644 cm. There was no significant correlation between depth and salinity between 16 and 27 m depth (Spearman $r = 0.48$; $P = 0.499343$), but TDS concentrations in the ice below SL2288–2303 cm were significantly higher ($P = 0.000214$) when compared to the TDS concentrations in the ice from 16 m down to SL2288–2303cm.

The concentrations of nutrients in Lake Vida brine (Murray *et al.,* 2012) are one-to-three orders of magnitude higher than the highest concentrations detected in the ice (Figures 3.3 and 3.4, bottom graph). Within the brine, $NO_3^-$, $NH_4^+$, $K^+$, $Mg^{2+}$, and $Ca^{2+}$ were detected in concentrations of at least one order of magnitude higher (905 μM, 3.88 mM, 82 mM, 665 mM, and 30 mM, respectively). $Na^+$ and $Cl^-$ were detected at least three orders of magnitude higher than the highest concentrations detected in the ice (1.91 M and 3.31 M, respectively). Sulfate below 2288 cm, alternatively, revealed peaks of

concentration higher than the concentration detected in the brine collected at 16 m depth (Murray *et al.,* 2012). The average concentration of $SO_4^{2-}$ in the ice core from 1622 cm to 2288 cm was 1.48 mM, and below 2288 cm it was 31.49 mM. Sulfate concentration in the brine was 58 mM (Murray *et al.,* 2012).

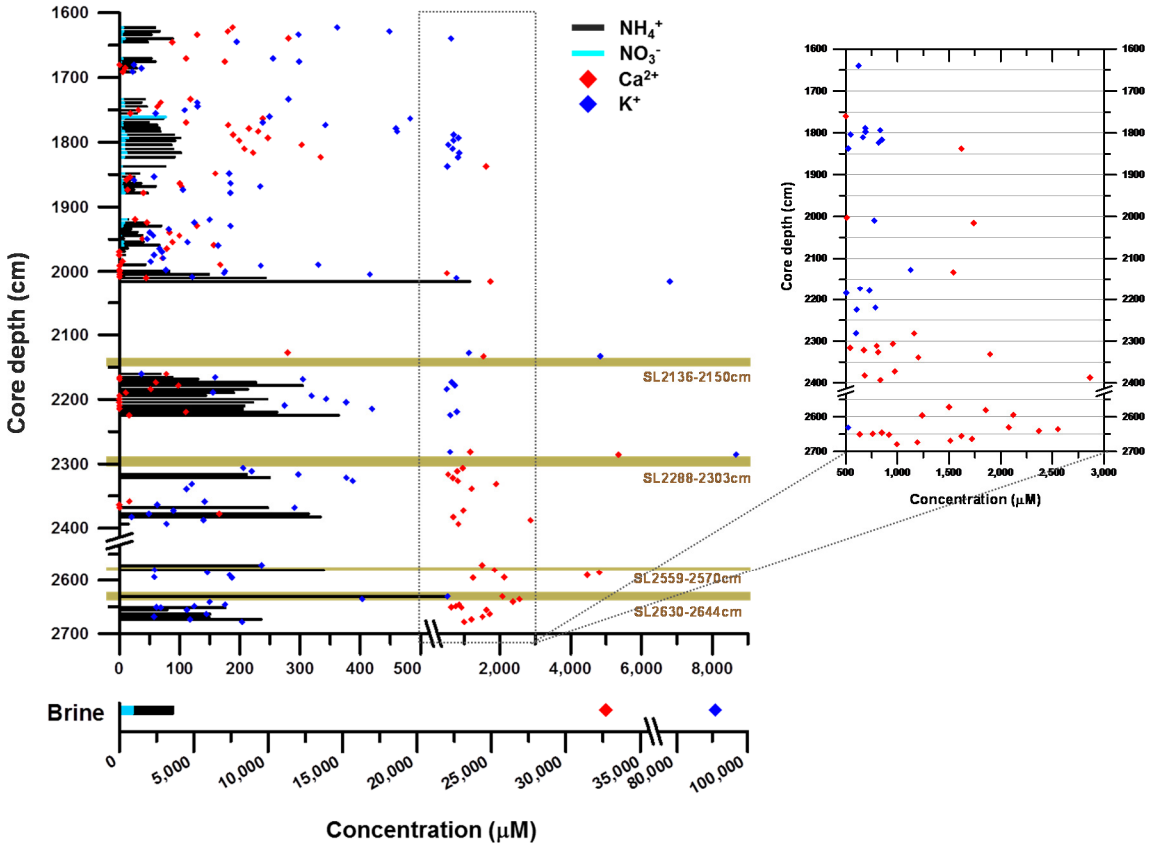



**Figure 3.4.** Concentration of the minor anion (ammonium and nitrate) and minor cation (calcium and potassium) identified in the Lake Vida ice core from 1622 cm to 2681 cm depth. Locations of the sediment layers (SL) analyzed by iTag libraries in the present study are shown in brown with their respective depths (SL2136–2150 cm, SL2288–2303cm, SL2559–2570cm, and SL2630–2644cm). The concentrations of ammonium, nitrate, calcium, and potassium in the Lake Vida brine are shown in a separate graph due to higher concentrations. The blank regions in the graph, with no plotted data, resulted from missing sections of the core. The plot in the right expanded cation concentrations between 500 and 3,000 µM for better visualization of the data.

Comparison among the ion ratios measured in the brine and ion ratios measured in the ice core suggested that the brine collected from the upper lake (15.9 and 18.5 m; Murray *et al.,* 2012) does not likely interact with the ice below SL2288–2303 cm (Table 3.1). Table 3.1 shows ion ratios above and below SL2288–2303 cm with ion ratios in the brine. Numbers approximate to 1.00 indicate smaller differences between the ion ratios in the brine and ion ratios in the ice. The brine ionic composition is more similar to the ionic composition above SL2288–2303 cm than the ionic composition below SL2288–2303 cm.

**Table 3.1.** Ratios between ionic composition proportions in the ice above and below SL2288–2303 cm and ionic composition ratios in the brine.

| | $Na^+/Cl^-$ | $Na^+/SO_4^{2-}$ | $Na^+/Mg^{2+}$ | $Na^+/Ca^{2+}$ | $Cl^-/SO_4^{2-}$ | $Ca^{2+}/Mg^{2+}$ | $(Na^+ + K^+)/(Ca^{2+}+Mg^{2+})$ | $Mg^{2+}/Na^+$ | $Mg^{2+}/Ca^{2+}$ |
|---|---|---|---|---|---|---|---|---|---|
| Above* SL2288-2303cm | 1.49 | 0.48 | 1.16 | 1.94 | 0.42 | 1.55 | 1.13 | 1.00 | 1.51 |
| Below SL2288-2303cm | 32.30 | 0.08 | 31.47 | 4.79 | 0.02 | 36.91 | 11.07 | 0.36 | 0.16 |

Calculation: $Ratio_{(ice/brine)}$ = ice ratio/brine ratio.
* from 16.22 m to sediment layer SL2288–2303cm.

The thick (≥10 cm) sediment bands observed within Lake Vida deep ice are not uncommon and are usually observed in glacier basal ice and basal ice sheets, where an intensive dynamic interaction between the ice and rock bed takes place (Gow *et al*., 1979; Knight, 1997; Doyle *et al*., 2013). Sediment layers thinner than 10 cm also have been described in other dry valley lake ice covers as a result of aeolian deposition followed by summer melting (Squyres *et al.,* 1991; Priscu *et al*., 1998), but this type of ice-sediment

banding is not present within the sediment of Taylor and Wright Valley lakes (Wagner *et al*., 2006; Wagner *et al*., 2011).

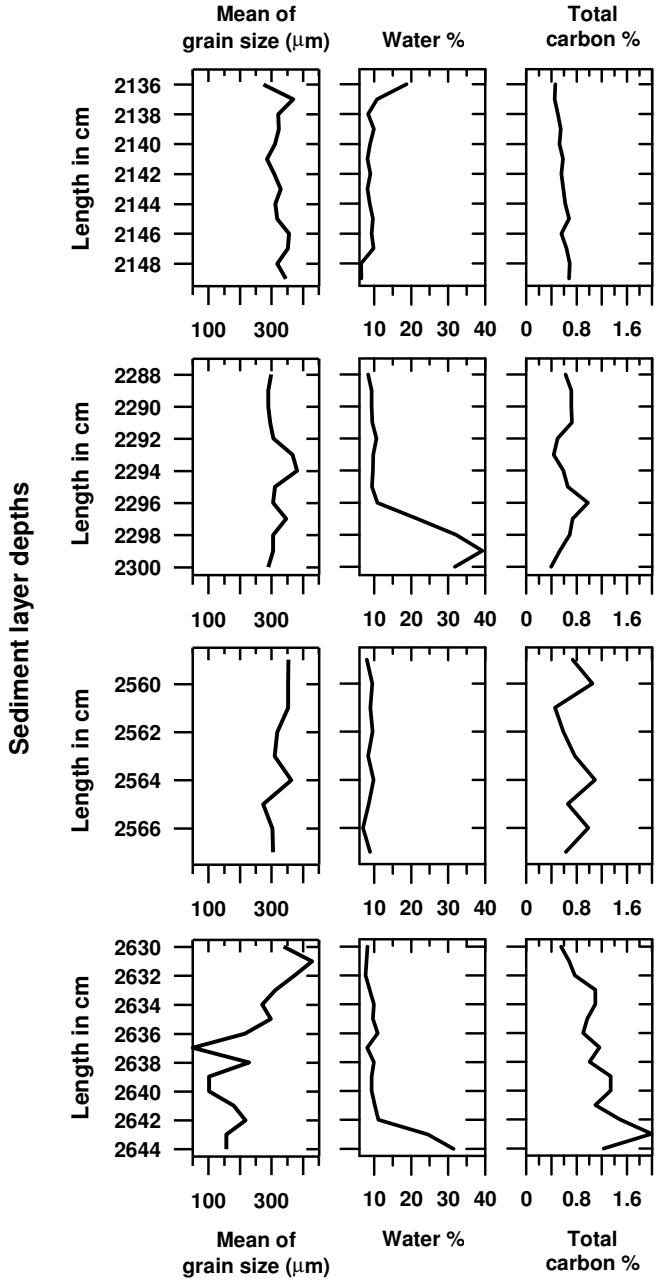



**Figure 3.5.** Profiles of the grain size (mean), water content, and total carbon content in the sediment layer from where the samples for iTag library analyses originated.

Sediments were analyzed for grain size (μm), water content (%), and TC (sum of organic and inorganic carbon; Figure 3.5). Sediment sections SL2136–2150 cm, SL2288–2303cm, and SL2559–2570 cm were not significantly different in grain size mean ($P > 0.1$), but the grain size mean decreased through SL2630–2644 cm and was smaller than the grain size mean in the other three sections ($P < 0.05$). Overall, there was a correlation between the increase in core depth and decrease in grain size when the four sediment layers were compared (Spearman $r = -0.28$, $P < 0.00001$). Total carbon content in sediment sections increased with the increase in depth (Spearman $r = 0.72$, $P < 0.00001$). Although SL2136–2150 cm and SL2288–2303 cm did not differ in TC ($\bar{x} = 0.57\%$, $\bar{x} = 0.63\%$, respectively, $P > 0.1$), there was a significant increase of TC in sections SL2559–2570 cm and SL2630–264 cm ($\bar{x} = 0.77\%$, $\bar{x} = 1.09\%$, respectively; $P < 0.05$). Water content in the sediment sections varied in each section due to proximity to ice layers (Figure 3.5).

Lake Vida cores recovered in 1996, 2005, and 2010 revealed cohesive sediment layers within the ice. Five sediment layers and dispersed layers of microbial mats were identified in the top 7 m of the lake ice (Doran *et al*., 2003). A core retrieved in 1961 identified layers of sand and silt at 3, 4, 10, 11, and 11.5 m below the lake surface (Calkin and Bull, 1967). At that time, it was suggested that the lake was frozen to the bottom and slowly built up by inflow of melted water from Victoria Upper and Lower Glaciers, as well as from smaller glaciers in the St. Johns Range (Calkin and Bull, 1967). Calkin and Bull established that the lake, as it exists now, has never completely melted, and that re-frozen seasonal cycles were the cause of formation of sediment layers within the ice.

Particles carried by wind year-round and transported by glacial meltwater streams during the summer formed layers on the lake ice surface and created thick horizons by differential absorption of radiation as well as melting and freezing processes (Calkin and Bull, 1967). Thirty-six years later, studies by Doran *et al.* (2003) supported findings by Calkin and Bull. Based on Calkin and Bull (1967) and Doran *et al.* (2003), Lake Vida sediment layers are the product of a four-step process: (1) during winter seasons, wind-carried sand and particles are deposited on the lake ice surface; (2) during warmer seasons, ephemeral streams flowing from surrounding glaciers flooded the ice surface with turbid particle-rich water; (3) particles in the flood water on top of the lake ice settled into layers; and (4) water on top of the lake froze, and the sediment layer was trapped within the ice. In addition, input of water in the solid phase (as ice or snow) has to be equally considered as a source of the particles present within the current ice (Sleewaegen *et al.*, 2002).

We hypothesize that the event(s) causing formation of the $SO_4^{2-}$ rich ice below SL2288–2303 cm was (were) coupled with a heavy input of particles enriched with mirabilite and gypsum. In Wright and Taylor Valleys, salts and lake water originated from marine sources and are $Cl^-$ rich with very little $SO_4^{2-}$ present (Torii and Yamagata, 1981). In Victoria Valley, where Lake Vida is located, the minerals that constitute the valley floor and soils predominantly originated from ice-shelf mineral beds dissolved by glacier meltwater (Ugolini *et al*., 1981; Brady and Batts, 1981). The valley floors in the Victoria Valley system have a maximum elevation of 1,000 m – with Upper and Lower Victoria, Barwick, eastern McKelvey, and Balham Valleys sloping irregularly down to

the shallow interior basin of Lake Vida, at 390 m above sea level (Calkin, 1971). Mirabilite crystals ($Na_2SO_4 \bullet 10H_2O$) exist as horizontal sheet-like aggregates beneath the sand and deeper in the soil in Balham Valley, 17 km west of Lake Vida (Gibson, 1962). We theorize that in an unknown warmer period in the past, glacial meltwater streams that flowed through mirabilite and gypsum beds would have deposited high amounts of sulfate into the lake. Later on, particles would have been deposited forming SL2559–2570 cm; the water rich in sulfate would have been frozen; and another cycle would have deposited another sediment layer (SL2288–2303 cm) on top of the sulfate-rich ice insulating it from the lake above. Thus, the high concentration of sulfate below SL2288–2303 cm could have been produced by the precipitation and redissolution of mirabilite that was brought into the lake by glacial meltwater streams (Brady and Batts, 1981).

The dissolution of iron-rich minerals also could be contributing to the origin of $SO_4^{2-}$ in Lake Vida ice. Ferrar dolerites rich in pyroxenes occur within the valley surrounding Lake Vida (Murray *et al.,* 2012; Salvatore *et al.,* 2013). The high concentration of iron previously detected in the lake brine was hypothesized to originate from the Ferrar dolerites in Victoria Valley (Murray *et al.,* 2012). Pyrite bedrocks were found to be the major source of $SO_4^{2-}$ in Arctic subglacial environments (Skidmore *et al.,* 2005; Marteinsson et al., 2013) as well as lakes and ponds located at Wright Valley, southbound of Victoria Valley (Nichols, 1990), playing an important role in the microbial community composition in these environments. Therefore, $SO_4^{2-}$ generated by the dissolution of pyrite ($FeS_2$) could be another reaction that added $SO_4^{2-}$ to the Lake Vida environment.

Ancient biological activity also could have been involved in the origin of the high concentrations of sulfate detected in the ice beneath SL2288–2303 cm, but this activity does not explain the primary input of sulfur into the system. Biological oxidation of hydrogen sulfide to sulfate is one of the major reactions in the global sulfur cycle (Friedrich *et al.,* 2001). Reduced inorganic sulfur compounds are exclusively oxidized by prokaryotes (in the presence or absence of oxygen), and sulfate is the major oxidation product. Table 3.2 shows the biological oxidation reactions that could be linked with high concentrations of sulfate within Lake Vida ice.

**Table 3.2.** Biological oxidation reactions that could be linked with high concentrations of sulfate within Lake Vida ice below SL2288–2303 cm.

| Sulfur oxidation reactions |
|---|
| Aerobic |
| (1) $H_2S_{(aq)} + 2O_{2(aq)} = 2H^+_{(aq)} + SO_4^{2-}{}_{(aq)}$ |
| (2) $S^0_{(s)} + 1.5O_{2(aq)} + H_2O = 2H^+_{(aq)} + SO_4^{2-}{}_{(aq)}$ |
| (3) $S_2O_3^{2-}{}_{(aq)} + 2O_{2(aq)} + H_2O = 2H^+_{(aq)} + 2SO_4^{2-}{}_{(aq)}$ |
| (4) $SO_3^{2-}{}_{(aq)} + 0.5O_{2(aq)} = SO_4^{2-}{}_{(aq)}$ |
| Anaerobic |
| (5) $S_2O_3^{2-}{}_{(aq)} + H_2O = SO_4^{2-}{}_{(aq)} + H_2S_{(aq)}$ |
| (6) $H_2S_{(aq)} + NO_3^-{}_{(aq)} + H_2O = SO_4^{2-}{}_{(aq)} + NH_4^+{}_{(aq)}$ |
| (7) $H_2S_{(aq)} + 1.6NO_3^-{}_{(aq)} = 0.4H^+_{(aq)} + SO_4^{2-}{}_{(aq)} + 0.8N_{2(aq)} + 0.8H_2O$ |
| (8) $S^0_{(s)} + 0.75NO_3^-{}_{(aq)} + 1.75H_2O = 0.5H^+_{(aq)} + SO_4^{2-}{}_{(aq)} + 0.75NH_4^+{}_{(aq)}$ |
| (9) $S^0_{(s)} + 1.2NO_3^-{}_{(aq)} + 0.4H_2O = 0.8H^+_{(aq)} + SO_4^{2-}{}_{(aq)} + 0.6N_{2(aq)}$ |
| (10) $4S^0_{(s)} + 4H_2O = 3H_2S_{(aq)} + SO_4^{2-}{}_{(aq)} + 2H^+_{(aq)}$ |

### 3.3.2. Microscopy observations

Observations of ice and sediment samples under the epifluorescence microscope revealed morphological diversity among the bacterial cells and the presence of diatom frustules at almost all analyzed depths (Figure 3.6). Coccoid, rod, and filamentous microbial cells were observed in different abundances through the core while diatom frustules were observed in all segments except in Ice_1866, Sed_2136, Sed_2288, and Sed_2559.

Cell enumeration revealed a distinction between ice and sediment segments, and a positive correlation with depth ($P < 0.001$). Cell counts from sediment samples were significantly higher than cell counts from the ice ($P = 0.00064$). The concentration of cells in the sediments and transition zones did not differ ($P = 0.06$), however. Cell counts in the sediment and transition zone segments ranged from $1.71 \times 10^6$ to $8.58 \times 10^6$ cells $g^{-1}$ while concentration of the cells in the ice segments ranged from $3.73 \times 10^4$ to $2.47 \times 10^6$ cells $mL^{-1}$ (Figure 3.7). Previous cell counts at depths of 4.8 m and 15.9 m in Lake Vida ice described a higher concentration of cells in the ice with sediment particles (depth 4.8 m; Mosier *et al.,* 2007). The most transparent ice, sample Ice_1886, contained the lowest concentration of cells, $3.73 \times 10^4$ cells $mL^{-1}$. Within ice, microbial cells were found attached to particles, in aqueous veins between ice crystals, and in brine pockets (Deming, 2002). In sediment and soils, cells are found attached to particles and mineral grains. Because of the high surface contact between grains and cells, the number of cells is expected to be higher. Indeed, a distinction between concentration of cells in ice and sediment was observed in the basal ice of Taylor Glacier (Doyle *et al*., 2013). The banded

basal ice under Taylor Glacier below 185 cm depth had a positive correlation between sediment content, solute concentration, and cell enumeration ($P < 0.05$). Cell counts in clear ice ranged from 2.6 to 4.9 x $10^2$ cells $mL^{-1}$; and in sediment-rich banded ice, counts ranged from 1.8 x$10^3$ to 1.6 x $10^4$ cells $g^{-1}$ (Doyle *et al*., 2013). Even though both cores have the same cell concentration pattern (the higher the sediment content, the higher the cell count), cell enumerations within the Lake Vida core were greater than within Taylor Glacier's basal banded ice. The accretion ice in subglacial Lake Vostok also contains higher cell concentration in ice with sediment inclusions, 3.6 x $10^4$ cells $mL^{-1}$, and lower cell concentration in clear ice, 3 x $10^2$ cells $mL^{-1}$ (Karl, 1999; J C Priscu *et al*., 1999). The cell count in the clearest ice segment from Lake Vida was two orders of magnitude higher than in the accretion ice in Lake Vostok. Moreover, the remainder of the ice segments in Lake Vida contains higher cell concentrations than the ice cover of Lake Bonney in Taylor Valley (Priscu, 1998). Lake Bonney particle-poor ice contains 5 x $10^3$ cells $mL^{-1}$, and particle-rich ice contains 0.1–4 x $10^5$ cells $mL^{-1}$. This indicates that the Lake Vida ice cover has a high concentration of cells throughout its entire core.

Figure 3.7 compiles results of cell enumeration in lower Lake Vida ice with data previously reported in the upper ice cover (Moiser *et al.,* 2007) and in the brine (Murray *et al.*, 2012; chapter II). The higher concentration of cells in the sediment layers is evident. Therefore, the slight positive correlation between depth and cell count (Spearman $r = 0.36$, $P = 0.000054$) is most likely related to the presence of sediments in the lower portion of the lake. If only the microbial cells in the ice are considered, there is no correlation between cell counts and depth (Spearman $r = - 0.06$).

Epifluorescence microscopy observations using a magnification of 1000x and fluorescent DNA stain did not indicate disruption of the cell membranes by scattered DNA surrounding the cells. To support this observation, scanning electron microscopy (SEM) was applied in the two deepest ice segments analyzed: Ice_2612 and Ice_2647. According to SEM observations, intact bacterial cells were in both samples (Figure 3.8). However, the bacterial cells observed exhibited abnormal morphology with wrinkled membranes and collapsed cell structure. These features are usually indicative of damaged cells that lost the rigidity and integrity of their cell walls (Mangoni *et al.,* 2004; Jo and Kim, 2013). We also observed that the bacterial cells identified in the deeper Lake Vida ice were bigger in size than the cells from Lake Vostok accretion ice and Antarctic permafrost (Priscu *et al.*, 1999; Gilichinsky *et al.*, 2007).

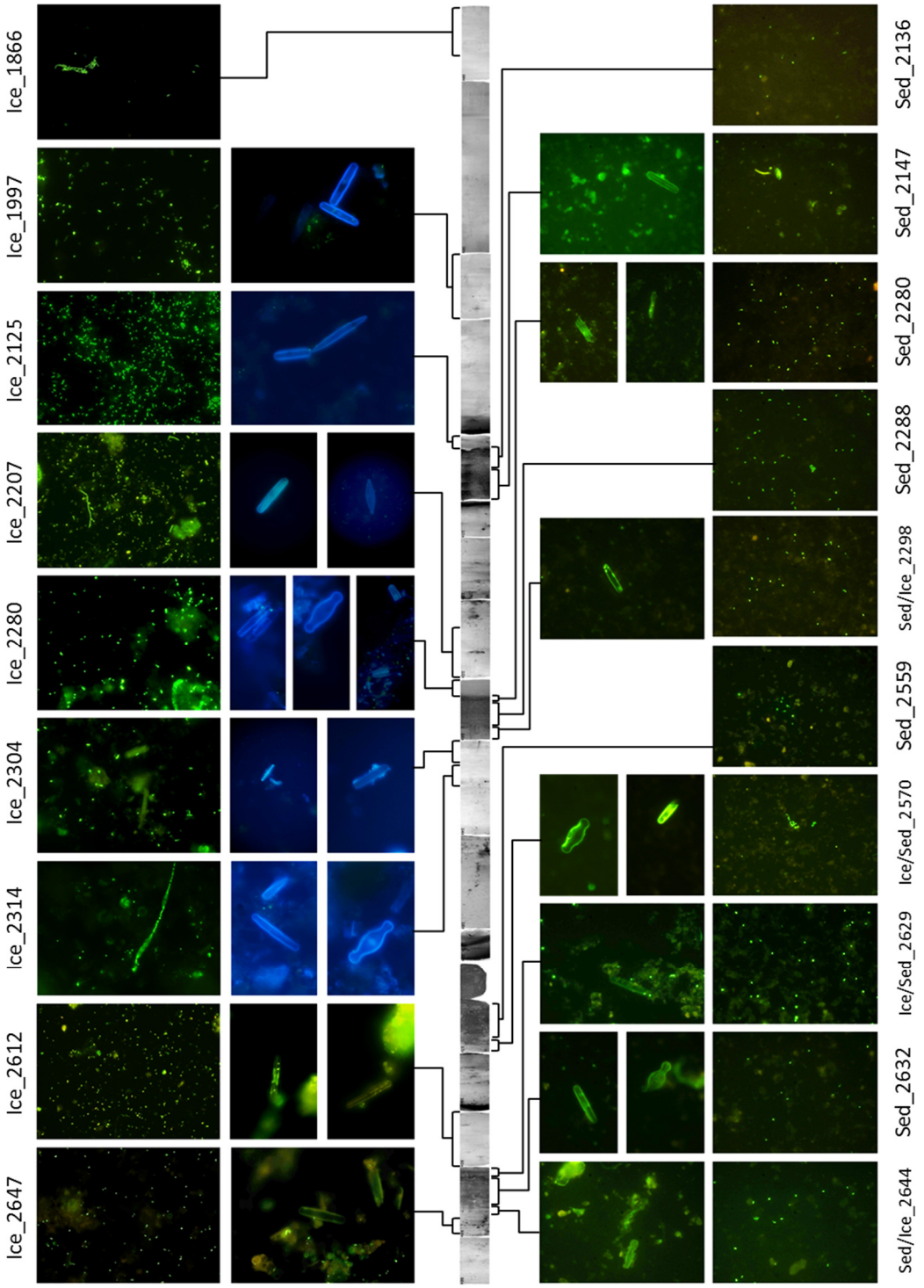



**Figure 3.6.** Epifluorescence micrographs of Lake Vida ice and sediment samples. 1000x magnification of microbial cells and diatom frustules detected in the ice (left) and sediment (right) segments. Ice core indicates where the samples were collected.

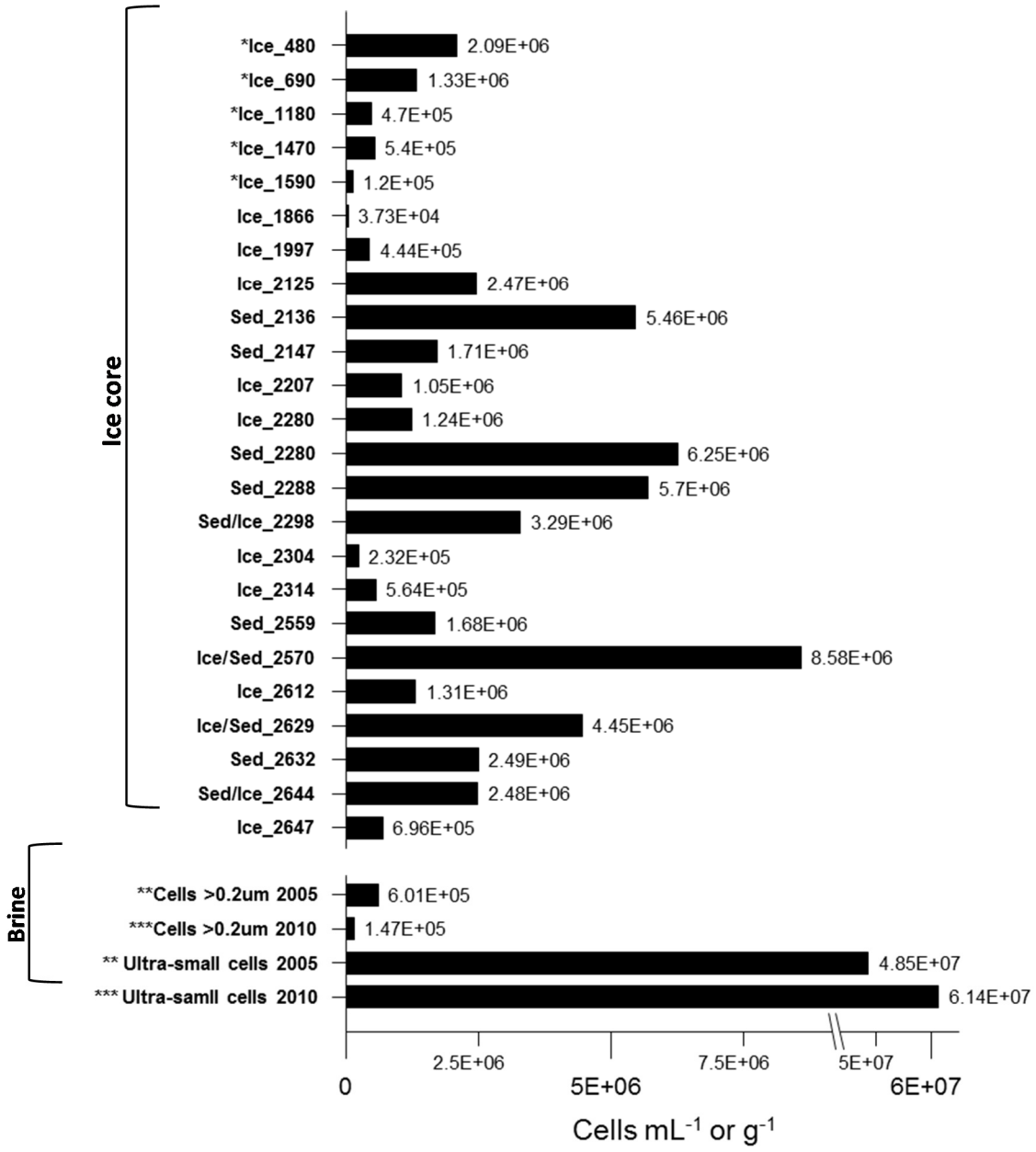



**Figure 3.7.** Cellular content in Lake Vida ice, sediment layers, and brine. Type of sample and depth are indicated in the name of the sample on the Y axis: sediment (Sed), ice (Ice), mixed sample (Ice/Sed); number indicates the depth in cm from the surface of the lake. The lower section of the graph shows the cell count of the two-cell size populations of brine collected in 2005 and 2010. * Results from Mosier *et al*., 2007. ** Results from Murray *et al*., 2012. *** Results from chapter II.

Diatom frustules observed in 15 of the 19 analyzed segments were identified as freshwater diatoms and are likely to have originated from glacier meltwater streams that flowed into the lake in the past. No planktonic diatoms were identified, suggesting that benthic taxa dominated the observed diatom assemblages. In the perennially ice-covered lakes of the McMurdo Dry Valleys, diatoms are confined to benthic mats within the photic zone and sediments (Spaulding *et al.,* 1997). In streams, diatoms are attached to benthic surfaces and within the microbial mat matrix (Esposito *et al.*, 2008). The frustules detected in the Lake Vida core were identified as belonging to the genera *Diadesmis* sp., *Hantzschia* sp., *Luticola* sp. *Muelleria* sp., *Navicula* sp., *Pinnularia* sp., and *Stauroneis* sp. (based on the Antarctic Freshwater Diatoms Web site; http://huey.colorado.edu/ diatoms). These genera were previously reported in Lake Hoare sediment and the major streams that flow into Lake Fryxell (Spaulding *et al.*, 1997; Stanish *et al.*, 2012). Indeed, Dry Valley lakes, streams, and soils are dominated by assemblages adapted to harsh conditions such as naviculoid taxa including the genera *Diadesmis*, *Luticola*, *Navicula* sp., and *Stauroneis* sp. (aerophilic diatoms). Nearly one-quarter of all dry valley stream diatom species are of the genus *Luticola* making it the most diverse genus in Dry Valley streams (West and West, 1911; Carlson, 1913; Esposito *et al*., 2008). *Hantzschia* is a genus that integrates numerous endemic species in the Dry Valleys; it occurs in abundance in soils and is moderately intermittent in low-flow streams (Stanish *et al*., 2012). SEM observations showed the presence of diatom frustules and frustule fragments in the deeper ice Ice_2612 and Ice_2647 (Figure 3.8). Frustules fragments were more often observed in Ice_2612 rather than Ice_2647. At least four distinct diatom frustules were observed in Ice_2167 by SEM (Figure 3.8 F, I, J and L).

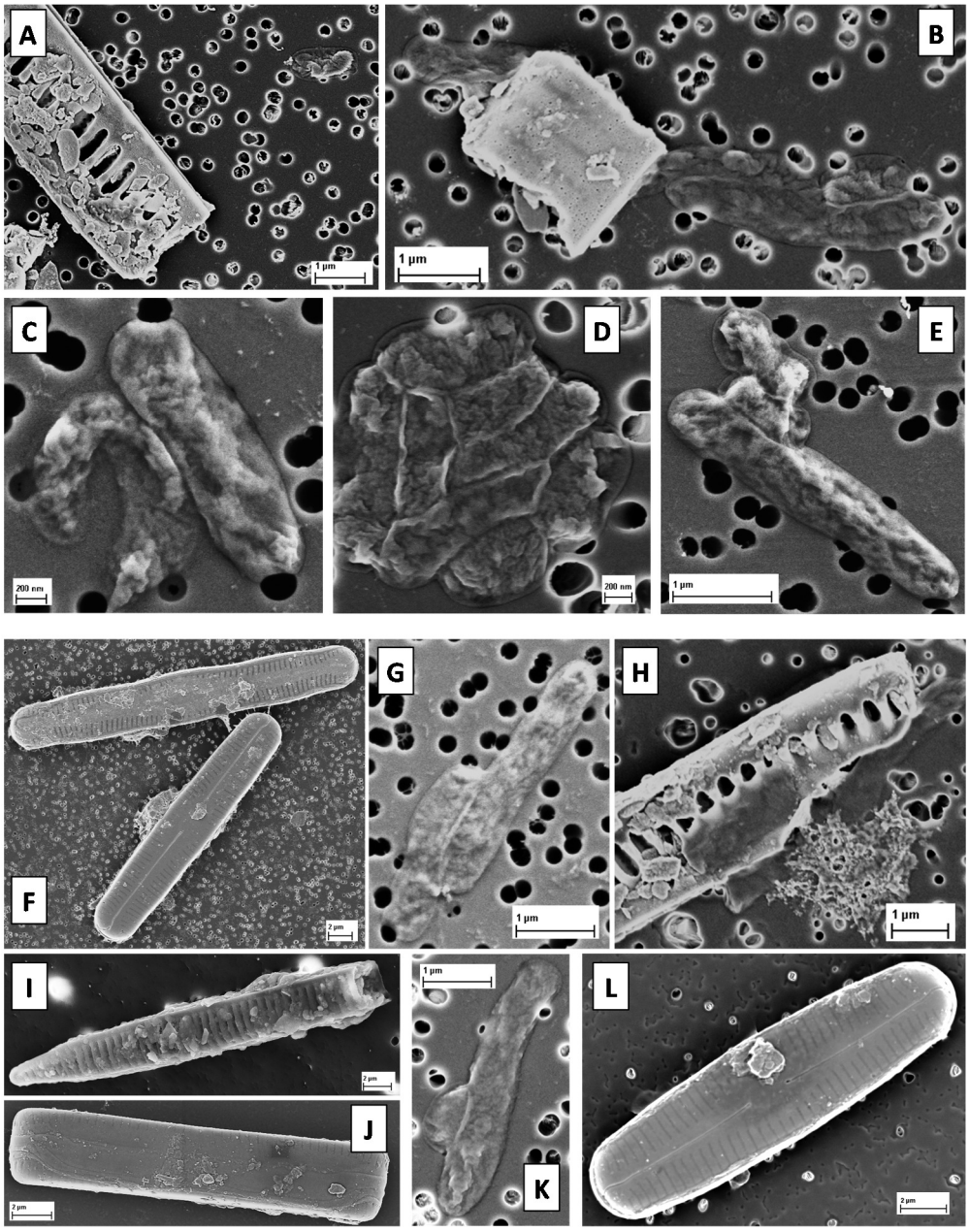



**Figure 3.8.** Scanning electron microscopy observations from bacterial cells and diatom frustules from Lake Vida deeper ice: (A-E) Ice_2612 and (F-L) Ice_2647.

### 3.3.3. Phylotype-based and OTU-based microbial assemblage composition

Phylotype-based analyses of the 872,627 high-quality reads (data not normalized) led to classification of 93% of known taxa and 7% (61,203 reads) of unclassified Bacteria. Archaeal 16S rRNA gene sequences were detected in the libraries with less than 0.02% of the total reads. *Proteobacteria, Bacteroidetes*, *Actinobacteria*, and *Firmicutes* were the most abundant phyla detected among the DNA iTag libraries. *Acidobacteria*, *Cyanobacteria*, *Planctomycetes*, and *Verrucomicrobia* also were detected but in less abundance. The few archaeal reads detected were affiliated with the Euryarchaeota orders *Methanosarcinales*, *Methanomicrobiales*, *Methanobacteriales*, *Halobacteriales*, *Thermoplasmatales* and with the Crenarchaeota order *Sulfolobales*. Recovery of only 0.02% of reads affiliated with archaeal sequences suggests that the *Archaea* Domain occurs among the rare phylotypes of the microbial assemblage in Lake Vida. This result corroborates with the lack of detection of archaeal 16S rRNA gene sequences in Lake Vida brine (Murray *et al*., 2012).

Reads affiliated with the genera *Demequina* (*Actinobacteria*), *Gillisia* (*Bacteroidetes*), *Sulfurovum* (*Epsilonproteobacteria*), and *Marinobacter* (*Gammaproteobacteria*) were observed among the top 10 most abundant taxa in almost all DNA iTag libraries (Figure 3.9). *Demequina*-affiliated reads were among the most abundant genera in 14 out of the 18 DNA iTag libraries with 0.77% to 3.30% occurrence. *Gillisia*-affiliated reads accounted for 8% to 15.5% of the microbial assemblage in 10 of the 18 iTag libraries and were identified among the top 10 genera in 15 of the 18 iTag libraries. *Sulfurovum*-affiliated reads were detected among the top 10 most abundant taxa

in 14 iTag libraries, corresponding to 13.22% to 25.53% of the microbial assemblage in seven of the iTag libraries (Figure 3.9). *Marinobacter*-affiliated reads were identified among the most abundant reads in 17 iTag libraries, although the genus was identified as the dominant group in only Ice_1997, 28.59% of the reads.

Using the phylotype-based analyses, we identified reads affiliated with two chloroplast sequences. Diatom chloroplast reads from the taxa *Bacillariophyta* were observed among the 10 most abundant reads in segment Ice_2314, and reads affiliated with chloroplast from the taxa *Streptophyta* were among the 10 most abundant reads in segment Sed/Ice_2644. Indeed, the diatom frustules detected in the microscopy observations belong to the phylum *Bacillariophyta*. Sequences affiliated with *Bacillariophyta* diatoms and *Streptophyta* chloroplasts had been previously detected in Lake Vostok accretion ice and in Arctic permafrost (Mackelprang *et al.*, 2011; Rogers *et al.*, 2013).

| Phylum | Class | Genus | Ice_1866 | Ice_1997 | Ice_2125 | Sed_2136 | Sed_2147 | Ice_2207 | Ice/Sed_2280 | Sed_2288 | Sed/Ice_2298 | Ice_2304 | Ice_2314 | Sed_2559 | Ice/Sed_2570 | Ice_2612 | Ice/Sed_2629 | Sed_2632 | Sed/Ice_2644 | Ice_2647 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Actinobacteria* | *Actinobacteria* | *Demequina* | | 1.02 | 1.79 | | | 2.31 | 2.24 | 1.64 | 2.46 | 3.30 | 1.16 | 0.77 | 2.48 | 1.66 | 1.03 | 1.55 | | 0.71 |
| | | *Ilumatobacter* | | | 0.48 | 0.90 | | 0.32 | 1.18 | | 3.60 | 1.35 | | | | | 0.70 | 1.30 | | |
| *Bacteroidetes* | *Flavobacteriia* | *Cloacibacterium* | 4.30 | | | | 0.41 | | | | | | | | | | | | | 0.49 |
| | | *Gelidibacter* | | | | | | | 0.86 | | | | | | | | | | | |
| | | *Gillisia* | | 6.60 | 13.44 | 10.06 | | 14.06 | 15.54 | 9.39 | 9.69 | 8.23 | 8.65 | 11.65 | 10.62 | 3.03 | | 0.89 | 3.29 | 6.06 |
| | | *Lutibacter* | | 1.23 | 7.48 | 2.15 | 0.75 | | | 2.41 | | | | | 5.16 | 9.98 | | | 0.82 | 5.37 |
| | | *Mesonia* | | 1.23 | | | | | | | | | | | | | | | | |
| | | *Psychroflexus* | 3.78 | 11.07 | 1.87 | | | 4.67 | | 0.87 | | 7.15 | 1.96 | | 3.29 | | | | | |
| | | *Salegentibacter* | | 1.58 | 0.88 | | | 3.86 | 2.03 | | | | | | 2.55 | | | | | |
| | *Cytophagia* | *Algoriphagus* | 1.20 | | | | | | | | | | | | | | | | | |
| | | *Rhodonellum* | | | | | | | | 1.07 | | | | | | | | | | |
| *Cyanobacteria/ Chloroplast* | Chloroplast | *Bacillariophyta* | | | | | | | | | | | 2.38 | | | | | | | |
| | | *Streptophyta* | | | | | | | | | | | | | | | | | 0.89 | |
| *Firmicutes* | *Bacilli* | *Alkalibacterium* | | | | | | | | | | | | | | | 1.39 | | | |
| | | *Carnobacterium* | | | | | | | 1.14 | | | | | | | | | | | |
| | | *Virgibacillus* | | | | | | 1.70 | | | | 1.85 | | | 2.36 | 0.57 | | | | |
| | *Clostridia* | *Clostridium sensu stricto* | | 0.43 | | 2.43 | 3.09 | | | | 0.78 | | 5.38 | 6.65 | 1.28 | | 6.83 | 12.92 | 4.09 | |
| | | *Alkaliphilus* | | | | | 0.64 | | | | | | | 0.62 | | | | | | |
| | | *Anaerovorax* | | | 1.75 | | | | | | | | | | | 0.70 | | | | |
| | | *Proteiniclasticum* | | | | | 0.40 | | | | | | | | | | | | | |
| | | *Saccharofermentans* | | | | | 3.94 | | | | | | | 1.51 | | 1.76 | | | | |
| *Planctomycetes* | *Planctomycetacia* | *Rhodopirellula* | | | | | | | 1.73 | | | 1.51 | | | | | 1.52 | | | |
| *Proteobacteria* | *Alphaproteobacteria* | *Hoeflea* | | | | | | | | 1.30 | 4.24 | | | | | | 5.05 | | 1.77 | 2.27 |
| | | *Novosphingobium* | 0.69 | | | | | | | | | | | | | | | | | |
| | *Betaproteobacteria* | *Acidovorax* | 2.46 | | | | | | | | | | | | | | | | | |
| | | *Comamonas* | 1.46 | | | | | | | 1.65 | | | | | | | | | | |
| | | *Polaromonas* | | | | | | | | | 1.32 | | | | | | | | | |
| | | *Sphaerotilus* | | | | | | | | | | | | | | | | | | 1.38 |
| | | *Thiobacillus* | | | | 0.88 | | | | | | | | 4.75 | | | 0.87 | 1.59 | | |
| | *Deltaproteobacteria* | *Desulfocapsa* | | | | | | | | | | | | | | 1.18 | | | | |
| | *Epsilonproteobacteria* | *Sulfurovum* | | 0.62 | 13.22 | 0.90 | 0.54 | 25.53 | 15.78 | 15.93 | | 1.04 | | | 17.68 | 16.32 | 2.33 | 4.58 | 2.48 | 22.52 |
| | *Gammaproteobacteria* | *Acinetobacter* | 36.74 | | | | 4.17 | 0.52 | | | 4.30 | 7.12 | 1.43 | 2.77 | | | 1.80 | 1.13 | 4.29 | 5.76 |
| | | *Haliea* | | | | 1.51 | | | | | | | | | | | | | | |
| | | *Lysobacter* | | | | 1.13 | | | | | | | 3.20 | | | | | | 2.21 | 1.66 |
| | | *Marinobacter* | 10.28 | 28.59 | 7.16 | 4.51 | 0.26 | 11.45 | 13.23 | 3.54 | | 3.85 | 2.75 | 2.98 | 5.31 | 1.60 | 0.60 | 2.96 | 0.70 | 1.11 |
| | | *Pseudoalteromonas* | 2.48 | | | | 0.69 | | | | 0.86 | 1.94 | | 2.65 | | | | 0.61 | 2.64 | |
| | | *Pseudomonas* | 1.22 | | | | | | | | 1.33 | | | | | | | | | |
| | | *Pseudoxanthomonas* | | | | | | | | | | | 1.82 | | | | | | | |
| | | *Psychrobacter* | | 19.34 | 5.88 | 1.21 | | 14.36 | 5.95 | 0.96 | | | | | 8.26 | | | | | |
| | | *Vibrio* | | | | | | | | | | | | 0.65 | | | | | | |
| *Verrucomicrobia* | *Opitutae* | *Opitutus* | | | | | | | | | 0.73 | | | | | | | | | |
| | Subdivision 5 | Genera incertae sedis | | | | | | | | | | | | | | 1.91 | | | | |
| | *Verrucomicrobiae* | *Luteolibacter* | | | | | | | | | | | 1.82 | | | | | 0.88 | | |
| Other genera (from total number of sequences analyzed) | | | 35.41 | 29.31 | 48.31 | 75.24 | 85.10 | 23.85 | 43.74 | 62.87 | 76.75 | 67.31 | 70.60 | 65.78 | 43.48 | 62.97 | 79.61 | 74.45 | 76.80 | 53.39 |

Lowest % Highest %

**Figure 3.9.** Relative abundance (in percentage) of the 10 most abundant genera detected in each iTag library according to phylotype-based analyses from non-normalized data (total of 872,627 reads). The numbers and color variation represent the relative frequency of the genus in the sample: lower percentage in blue; higher percentage in yellow.

The commonly used genus-level cutoff for the phylotype-based method generates such a broad distance between the reads that it masks the differences within lineages, and those differences can be perceived by the OTU-based method (Schloss and Westcott, 2011). Therefore, OTU-based analyses considering the normalized data and cutoff distance of 0.03 were conducted. Results led to identification of the same major taxa as identified by the phylotype-based analyses (except for *Demequina*), but with greater sensitivity to phylogenetic variability, accuracy in the relative abundance, and

distribution of taxa among the microbial assemblages. For example, by OTU-based analyses, five OTUs were assigned to the genus *Gillisia* and identified with distinct distribution patterns through the core (Figure 3.10). The same event (more than one OTU assigned to the same taxa and distinct abundance among iTag libraries) was observed with OTUs phylogenetically affiliated with the genera *Acinetobacter, Clostridium sensu stricto, Demenquina, Lutibacter, Marinobacter, Psychrobacter, Psychroflexus,* and *Sulfurovum*. The number of OTUs assigned to the same genus depends on at which distance threshold the OTUs were clustered (0.03, or 0.05, or 0.1). Distance units of 0.03 were found to be the threshold that provides the more refined definition of subgenus populations (Schloss and Westcott, 2011).

While both types of analyses showed similar results for three out of the four most abundant taxa (*Gillisia, Sulfurovum,* and *Marinobacter*), incongruities arose where taxa identified in phylotype-based analyses were missing in the OTU-based analyses (e.g., *Ilumatobacter*, *Bacillariophyta*, *Streptophyta*, *Saccharofermentans*, *Planctomycetes*, *Hoeflea*, *Pseudoalteromonas*, and *Verrucomicrobia*). Both phylotype-based and OTU-based methods are known to have unique challenges that affect the capability to implement the analyses and interpretation of the results (Schloss and Westcott, 2011). One of the reasons of the mismatch observed between the results could be related to the number of sequences analyzed in each method. For phylotype-based analyses, all the 872,627 reads were taxonomically characterized; and for OTU-based reads, almost three times fewer reads were analyzed (299,989 reads).


**Figure 3.10.** Heatmap of the 50 most abundant OTU distributed among the segments according to OTU-based analyses. Data originated from resampled reads and distance cutoff of 0.03. The scale indicates relative abundance of the OTUs in the iTag libraries.

Correlation analyses between the OTU abundance and depth revealed seven OTUs with negative correlation to depth, five OTUs with no correlation to depth ($0.0 \leq r \leq 0.05$), ten OTUs with a weak positive correlation to depth ($0.05 < r < 0.3$), and 14 OTUs with a stronger positive correlation to depth ($0.3 < r < 0.6$; Figure 3.11). *Gillisia, Actinobacteria*, *Marinobacter*, *Psychrobacter, Anaerovorax*, *Marisediminicola*, and *Psychroflexus* affiliated OTUs decreased in abundance with an increase in depth ($P < 0.00001$). *Desulfosporosinus, Desulfocapsa*, *Cellulomonas*, *Acinetobacter*, *Conexibacter*, *Clostridium sensu stricto*, and *Acetivibrio* affiliated OTUs exhibited the strongest positive correlation to depth ($P < 0.00001$). *Demequina* affiliated OTUs also increased in abundance with depth (Spearman $r = 0.35$, $P < 0.00001$). *Sulfurovum* affiliated OTUs, one of the most abundant OTUs throughout the core, presented a weak positive correlation to depth (Spearman $r = 0.28$, $P < 0.00001$). *Sulfurovum* is a facultative, denitrifying, chemolitoautotrophic sulfur-oxidizing bacterium capable of growth with elemental sulfur or thiosulfate as an electron donor, and oxygen or nitrate as electron acceptors (Park *et al.*, 2012). This group had been observed previously in Lake Vida brine and upper ice, and it is typically found in sulfidic environments (Macalady *et al.*, 2006; Murray *et al.*, 2012).

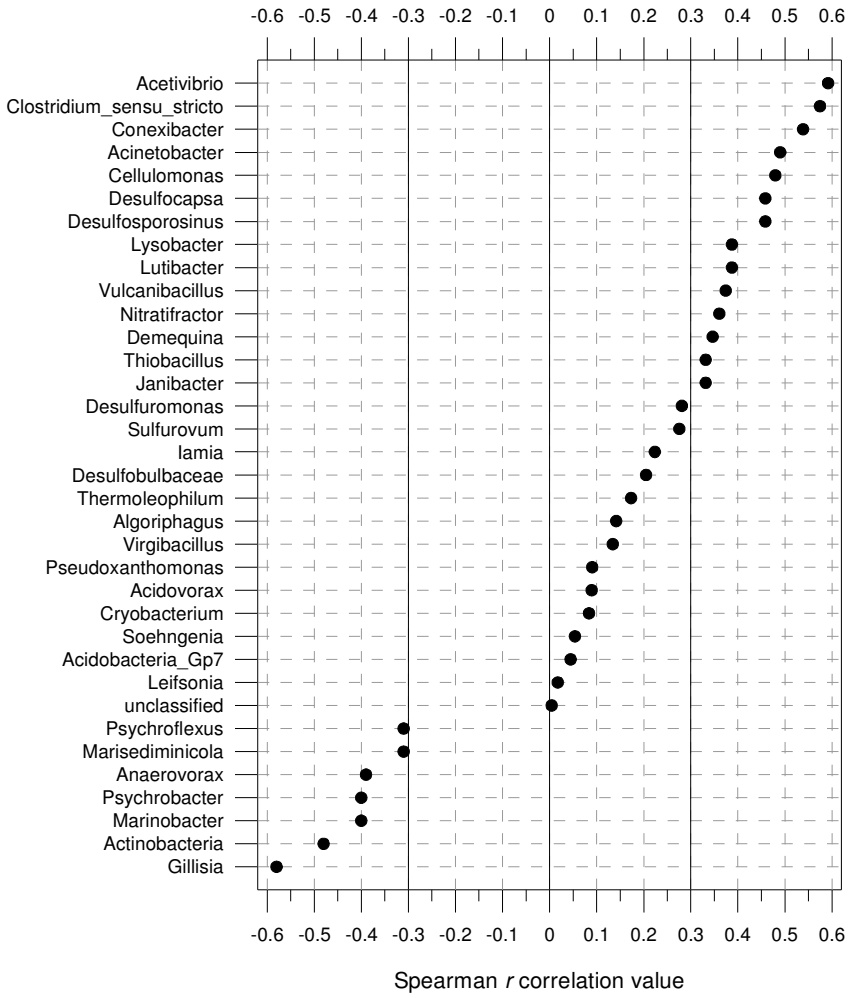



**Figure 3.11.** Correlation between OTU affiliated taxa and Lake Vida core's depth based on the Spearman *r* correlation coefficient.

A significant distinction of OTU abundance also was observed above and below SL2288–2303 cm. *Marinobacter*-affiliated OTUs were significantly more abundant throughout the core above sediment layer B ($P < 0.05$), where $Cl^-$ is the major anion. Members of the heterotrophic denitrifier bacterial genus *Marinobacter* are commonly found in marine environments and environments with high concentration of NaCl (optimum growth in 7–8% NaCl). The genus was identified in the eastern and western

lobes of Lake Bonney (Ward and Priscu, 1997; Glatz *et al.*, 2006), and in Blood Falls brine (Mikucki and Priscu, 2007). In Lake Vida, the genus was previously detected using molecular and cultivation approaches from both Lake Vida ice at 15.9 m depth and from the brine (Mosier *et al*., 2007; Murray *et al*., 1012; chapter II in this dissertation). In contrast to *Marinobacter's* higher abundance in the core above SL2288–2303 cm, OTUs affiliated with *Cellulomonas*, *Janibacter*, and *Lysobacter* were significantly more abundant throughout the core below SL2288–2303 cm ($P < 0.05$). *Cellulomonas* and *Janibacter* were previously detected in Lake Vida ice at depths of 4.8 m and 15.9 m, and in algal mats from 5 to 15 cm below the soil surface in Victoria Valley (Mosier *et al*., 2007; Antibus *et al*., 2012). *Cellulomonas* is an actinobacteria with cellulolytic capacities found in environments with relative high amounts of aromatic compounds such as cellulose (Stackebrandtt and Kandler, 1979). This genus was detected in high quantity within dry valleys permafrost and was identified among cryoconite hole microbial assemblages (Christner *et al.*, 2003). *Lysobacter* is a gammaproteobacteria often identified within dry valleys mineral soils and has the capacity to degrade aromatic compounds (Niederberger *et al.*, 2008; Bottos *et al.*, 2013).

Looking for a correlation between OTU abundance and sample type (ice, sediment, or transition zone), we found a group of OTUs significantly more abundant in sediment segments and transition zones compared to ice segments. Not surprisingly, the majority of the OTUs with higher abundance within sediment segments and transition zones were affiliated with *Actinobacteria*, a group that contributes to a significant portion of the microbial assemblages inhabiting dry valley soils (Babalola *et al.*, 2009). OTUs

affiliated with the *Actinobacteria* genera *Cellulomonas*, *Conexibacter*, *Marisediminicola*, *Thermoleophilum*, and *Demequina* accompanied by the betaproteobacteria *Thiobacillus* were more abundant in transition zone and sediment segments rather than in ice ($P < 0.05$). The actinobacteria *Janibacter* was found in higher abundance in transition zones ($P < 0.05$). The abundance of *Nitratifractor*, unclassified *Actinobacteria*, and *Desulfosporosinus* was significantly higher in sediment segments than in ice and transition zones ($P < 0.05$). Thus, we can infer that in Lake Vida, the microbial group associated with the sediment layers includes versatile heterotrophs that may be equipped to use various carbon sources, including complex aromatic polymers.

Principal component analyses (PCA) was used to investigate the association between (i) OTUs identified in the 18 core segments; (ii) ice and sediment geochemistry; and (iii) sample type (ice, sediment, or transition zone). PCA results revealed two distinct major OTU clusters among the core segments (Figure 3.12A). PC1 (which explained 20.4% of the variation observed) was composed of OTUs phylogenetically affiliated with *Acidobacteria*_Gp7, *Acidovorax*, *Algoriphagus*, *Cellulomonas*, *Conexibacter*, *Cryobacterium*, *Janibacter*, *Marisediminicola*, *Nitratifractor*, and *Thermoleophilum*. These bacteria are known to be abundant in soils and are associated with carbon cycling. Many of these bacteria were detected in higher abundance in the sediment segments and transition zones than in the ice segments. PC2 (which explained 14.5% of the variation) was composed of OTUs phylogenetically associated with *Desulfocapsa*, *Desulfuromonas*, *Soehngenia*, and *Sulfurovum* – which are known to participate in the biogeochemical

cycling of sulfur – and *Lutibacter*, *Acetivibrio*, and *Vulcanobacillus* which are polymer degraders (Table 3.3).

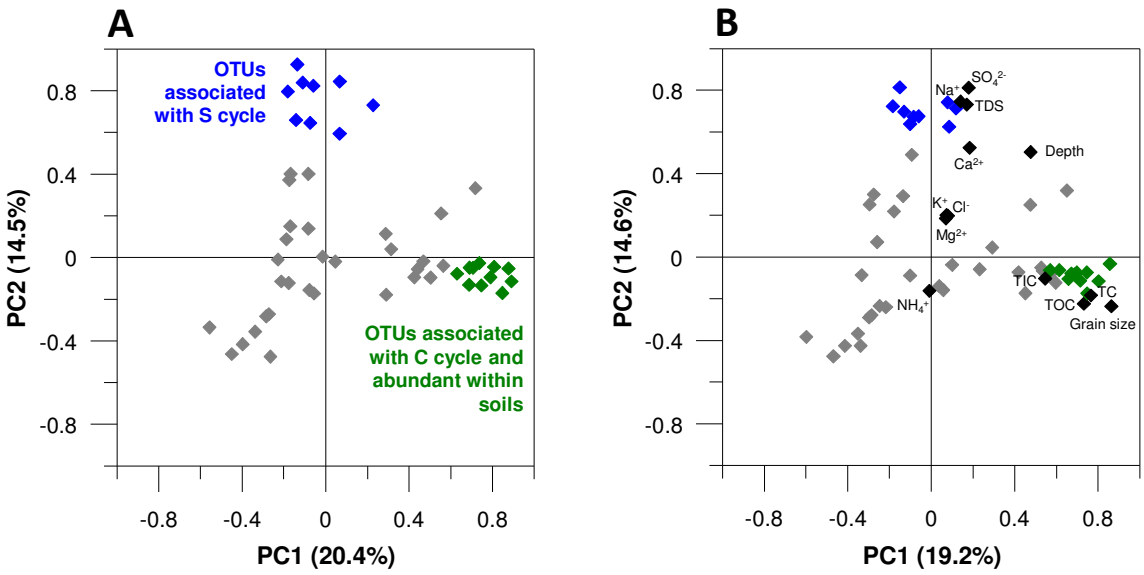



**Figure 3.12.** Principal component analysis (PCA) of microbial assemblages (as OTUs) and environmental parameters from Lake Vida core segments. PCA was used to explore the relationship among (A) OTUs and (B) OTUs and environmental parameters. PCA revealed two principal component loading of variables: PC1 (green) with OTUs phylogenetically associated with *Acidobacteria*_Gp7, *Acidovorax*, *Algoriphagus*, *Cellulomonas*, *Conexibacter*, *Cryobacterium*, *Janibacter*, *Marisediminicola*, *Nitratifractor*, and *Thermoleophilum*; and PC2 (blue) with OTUs phylogenetically associated with *Acetivibrio*, *Desulfocapsa*, *Desulfuromonas*, *Lutibacter*, *Soehngenia*, *Sulfurovum*, and *Vulcanibacillus*.

When the environmental parameters (ions, carbon, grain size, TDS, and depth) were added to the PCA, the PC1 OTU cluster in plot A (green loading) grouped with TOC, TIC, TC, and grain sizes (Figure 3.12 B). The PC2 OTU cluster in plot A (blue loading) grouped with sulfate, sodium, and TDS. A strong association between potassium, chloride, and magnesium was identified – with a weak association to the

principal components extracted ($K^+$, $Cl^-$, and $Mg^{2+}$ uniqueness ≅ 0.95). Calcium, water content, ammonium, and the OTUs shown in gray in Figure 3.12 also were weakly associated with the two principal components extracted.

**Table 3.3.** Biogeochemical characteristics of cultivated bacteria phylogenetically affiliated with OTUs from PC2 loading variables (blue loading) in Figure 3.12.

| Closest cultivated species | Tolerance to oxygen | Carbon sources | Electron donor | Electron acceptor | Comment | Ref. |
|---|---|---|---|---|---|---|
| *Lutibacter* | Aerobe | Organic | Organic | $O_2$ | • Polymer degrader | Choi and Cho, 2006 |
| *Sulfurovum* | Facultative | $CO_2$ | $S^0$, $S_2O_3^{2-}$ | $NO_3^-$, $O_2$ | • Sulfur reducer<br>• Needs salt to grow | Inagaki *et al.,* 2004 |
| *Soehngenia* | Anaerobe/ Aerotolerant | Organic | Organic, $H_2$ | $S_2O_3^{2-}$, $SO_3^{2-}$ | • $N_2$ fixer | Parshina *et al.*, 2003 |
| *Acetivibrio* | Anaerobe | Organic | Organic | Organic | • Polymer degrader<br>• Does not use $NO_3^-$ or sulfur compounds as electron acceptors | Tanaka *et al.,* 1991 |
| *Desulfocapsa* | Anaerobe | $CO_2$ | $S^0$, $S_2O_3^{2-}$, $SO_3^{2-}$ | $S^0$, $S_2O_3^{2-}$, $SO_3^{2-}$ | • Some are capable of $S^0$ disproportionation<br>• $N_2$ fixer<br>• Growth only occurs in the presence of $Fe^{3+}$ | Finster *et al.*, 1998; Finster *et al.*, 2013 |
| *Desulfuromonas* | Anaerobe | Organic | Organic | $S^0$, $Fe^{3+}$, $Mn^{4+}$ | • Sulfur reducer<br>• Oxidation of acetate to $CO_2$ is linked to the reduction of $S^0$ to $H_2S$. | Pfennig and Biebl, 1976; Roden and Lovley, 1993 |
| *Vulcanibacillus* | Anaerobe | Organic | Organic | $NO_3^-$ | • Spore forming<br>• $NO_3^-$ as sole electron acceptor and it is not further reduced to Ammonia or $N_2$ | L'Haridon *et al.,*2006 |

The iTag libraries from the Lake Vida core revealed a microbial assemblage known to be highly specialized in utilizing sulfur ions both as electron donors and acceptors, accompanied by an assemblage known to be highly specialized in oxidizing acetate and aromatic compounds (Table 3.3). *Sulfurovum* sp. is a sulfur oxidizer (Inagaki *et al.,* 2004) that metabolizes reactions 2, 3, 8, and 9 described in table 3.2. *Desulfocapsa*

sp. is capable of sulfur disproportionation (splitting one compound into an electron donor and an electron acceptor) using elemental sulfur ($S^0$), sulfite ($SO_3^{-2}$), and thiosulfate ($S_2O_3^{2-}$) to produce both hydrogen sulfide ($H_2S$) and sulfate ($SO_4^{2-}$) as shown in reactions 5 and 10 in Table 3.2 (Finster *et al.*, 2013). OTUs affiliated with these two genera were identified throughout the core, but there was no significant difference in the abundance of the OTUs when considering the ice above and beneath SL2288–2303 cm ($P > 0.05$). Therefore, if the microbial assemblage were the only contributor to the increase of sulfate in the ice, the levels of sulfate above SL2288–2303 cm also would be high. Accordingly, we inferred that the high concentrations of sulfate beneath SL2288–2303 cm might have originated from abiotic weathering of sulfur-containing rocks, such as mirabilite, and not from ancient biological activity.

Analyses of the OTU-based results indicated that there is a distinction between the microbial assemblage within the sediment layers and the microbial assemblage within the ice. We observed that Lake Vida sediment layers contain a higher abundance of microorganisms specialized in degradation of complex polymers in comparison to the ice. The OTUs that formed the PC2 (Table 3.3) all presented a positive correlation of abundance with an increase in depth. However, they did not present correlation with sample type and were not more or less abundant below SL2288–2303 cm. Furthermore, PC2 loaded OTUs were affiliated with microorganisms known to specialize in sulfur oxidation, reduction, and disproportionation indicating that the microbial community within the Lake Vida core from 18 to 27 m depth was (or may still be) active based on the sulfur cycle redox reactions and degradation of organic polymers.

### 3.3.4. Microbial assemblage diversity and structure within the core

The number of total reads, unique reads, OTUs, diversity index, and abundance estimators from the DNA iTag libraries before and after normalization of data are shown in Table 3.4. Chao1 richness estimator and Shannon diversity index (H') calculated before and after normalization of data did not present significant alteration (Chao1 $P$ = 0.40 and H' $P$ = 0.92, respectively). The number of OTUs generated from total high quality reads, however, differed slightly from normalized reads at the significance level of $P < 0.1$ ($P = 0.086$). To avoid unequal sequence coverage and under or overestimation of diversity among iTag libraries, alpha diversity (diversity within communities) and beta diversity (partitioning of diversity among communities) were analyzed using the normalized data and cutoff of 0.03 distance unit.

**Table 3.4.** Summary of the number of reads during pre-processing (raw reads - quality control – chimera check – unique reads) and measures of alpha diversity of the core segments analyzed before and after normalization of the data (distance of 0.03).

| iTag library ID | No. of raw reads | No. of reads after quality control | No. of reads after chimera check | No. of unique reads | Cutoff distance of 0.03 | | | No. of reads after resampling | No. of unique reads after resampling | Cutoff distance of 0.03 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | OTU | Chao1 | Shannon's index (H') | | | OTU | Chao1 | ACE | Shannon's index (H') |
| Ice_1866 | 51,369 | 31,018 | 30,545 | 5,114 | 610 | 837 | 3.04 | 13,043 | 3,508 | 426 | 698 | 948 | 3.05 |
| Ice_1997 | 62,685 | 39,357 | 38,842 | 5,188 | 708 | 1,289 | 2.84 | 13,043 | 3,318 | 387 | 797 | 1,096 | 2.82 |
| Ice_2125 | 126,821 | 83,535 | 80,622 | 12,091 | 1,247 | 2,569 | 3.3 | 13,043 | 3,339 | 474 | 977 | 1,345 | 3.24 |
| Sed_2136 | 52,233 | 35,660 | 31,268 | 8,771 | 1,970 | 5,011 | 4.97 | 13,043 | 4,130 | 1,142 | 2,767 | 4,008 | 4.89 |
| Sed_2147 | 132,096 | 95,490 | 90,479 | 16,819 | 1,945 | 3,955 | 3.77 | 13,043 | 5,135 | 658 | 1,428 | 2,004 | 3.71 |
| Ice_2207 | 62,444 | 37,889 | 37,460 | 4,933 | 624 | 1,175 | 2.86 | 13,043 | 3,035 | 346 | 586 | 785 | 2.84 |
| Ice/Sed_2280 | 105,524 | 66,001 | 63,363 | 11,539 | 1,272 | 2,374 | 3.66 | 13,043 | 3,929 | 592 | 1,321 | 2,126 | 3.67 |
| Sed_2288 | 39,031 | 24,578 | 23,940 | 5,517 | 753 | 1,260 | 4.05 | 13,043 | 3,895 | 554 | 1,017 | 1,327 | 4.06 |
| Sed/Ice_2298 | 61,959 | 37,868 | 36,877 | 9,597 | 893 | 1,355 | 4.69 | 13,043 | 4,785 | 646 | 901 | 892 | 4.68 |
| Ice_2304 | 63,510 | 41,849 | 40,652 | 9,652 | 1,038 | 1,543 | 4.14 | 13,043 | 4,839 | 262 | 1,064 | 1,395 | 4.11 |
| Ice_2314 | 63,568 | 36,297 | 34,455 | 9,813 | 1,404 | 2,362 | 4.71 | 13,043 | 5,118 | 884 | 1,718 | 2,358 | 4.69 |
| Sed_2559 | 73,888 | 38,973 | 36,981 | 8,566 | 1,064 | 1,485 | 4.56 | 13,043 | 4,777 | 757 | 1,123 | 1,128 | 4.57 |
| Ice/Sed_2570 | 125,870 | 76,103 | 73,066 | 13,686 | 1,711 | 3,306 | 3.91 | 13,043 | 4,010 | 659 | 1,445 | 2,239 | 3.85 |
| Ice_2612 | 36,646 | 20,448 | 17,603 | 4,992 | 1,110 | 1,977 | 4.63 | 13,043 | 4,601 | 519 | 887 | 1,334 | 3.89 |
| Ice/Sed_2629 | 40,401 | 22,676 | 21,414 | 5,190 | 949 | 2,505 | 4.25 | 13,043 | 4,891 | 934 | 2,022 | 3,476 | 4.62 |
| Sed_2632 | 135,205 | 69,012 | 63,378 | 13,177 | 1,137 | 1,875 | 3.92 | 13,043 | 4,204 | 729 | 1,335 | 1,709 | 4.23 |
| Sed/Ice_2644 | 35,994 | 20,862 | 20,115 | 5,804 | 892 | 1,111 | 4.82 | 13,043 | 4,429 | 782 | 1,014 | 1,006 | 4.81 |
| Ice_2647 | 70,812 | 36,651 | 36,019 | 8,458 | 950 | 1,440 | 3.82 | 13,043 | 4,269 | 584 | 986 | 1,306 | 3.81 |
| Total | 1,340,056 | 814,267 | 777,079 | 158,907 | | | | 234,774 | 76,212 | | | | |

Sediment and transition zone segments presented significantly higher values for Chao1 and ACE (abundance-based coverage estimator) estimators compared to ice estimators ($P < 0.05$). In addition, both estimators increased with an increase in depth in the core (Chao1 Spearman $r = 0.27$ and ACE Spearman $r = 0.12$, $P < 0.00001$). Among the 18 segments analyzed, segment Sed_2136 presented the highest number of OTUs (1,142 OTUs), the highest Chao1 estimator (2,767), the highest ACE estimator (4,008), and the rarefaction curve with the steeper slope (Table 3.4 and Figure 3.13). Chao1 is a robust estimator of minimum richness and is based on rare OTUs or singleton in the sample (Chao, 1984). The idea behind the estimator is that, if an assemblage is being sampled and rare species (singletons) are still being discovered, it is likely that rare species remain unidentified. So, the higher the Chao1 estimator, the more rare OTUs are likely to be found. ACE incorporates data from all species with fewer than 10 individuals, rather than just singletons and doubletons (Chao and Lee, 1992). For example, in Sed_2136 iTag library, 1,142 OTUs were detected whereas a minimum of 2,767 OTUs were estimated by Chao1, and 4,008 OTUs were estimated by ACE (considering the normalized data). This means that, for sample Sed_2136, the 13,043 reads from the normalized data covered 41% of the estimated richness (OTU:Chao1 ratio of 0.41). Sed/Ice_2644, Sed/Ice_2298, and Sed_2559 presented the highest OTU:Chao1 ratios of 0.77, 0.72, and 0.67, respectively. Sample Ice_2304 presented the lowest OTU:Chao1 ratio (0.25) and OTU:ACE ratio (0.19), considering the normalized data. However, considering the total number of reads generated by the iTag library, the OTU:Chao1 ratio for Ice_2304 was 0.67.

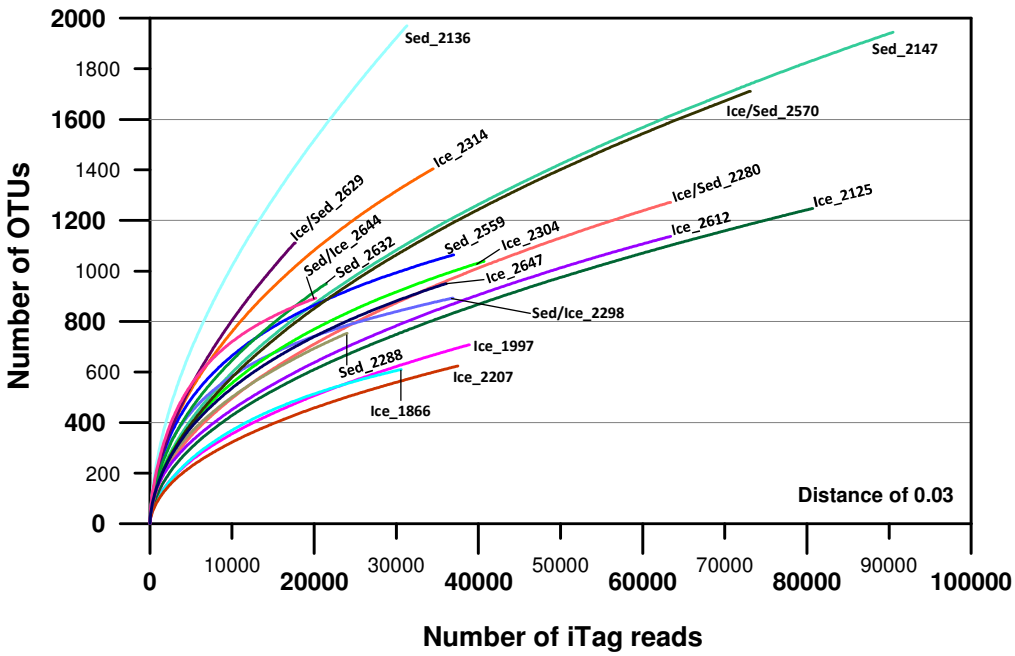



**Figure 3.13.** Rarefaction curve of not normalized DNA iTag libraries.

Shannon H' diversity index, which takes into account abundance and evenness of OTUs in the sample (Shannon, 1948), presented a positive correlation with depth and a weakly positive correlation with cell counts. An increase in microbial assemblage diversity was observed with increased depth (Spearman $r = 0.50$, $P < 0.00001$) and an increase in cell concentration (Spearman $r = 0.12$, $P < 0.00001$). An opposite correlation between the depth and diversity index was detected in a West Lake Bonney sediment core, where the diversity index decreased with depth in the core (10 cm depth presented H´ = 4.5 and 355 cm depth presented H' = 2.8; Tang *et al.*, 2013). No correlation was found between the Shannon diversity index and the sample type (ice, sediment or transition zone, $P < 0.05$). High Shannon index values (> 3) based on clone library approaches have been reported in dry valley soils but differed from apparently more

homogeneous microbial assemblages found in ice, lake sediment, and lake water within the Dry Valleys (Aislabie *et al*., 2006; Glatz *et al*., 2006; Mikucki and Priscu, 2007). Chao1 estimator versus Shannon index plot results revealed that the highest diversity values, except for Sed_2136, occurred below the 2288 cm depth in the core ($P = 0.005$; Figure 3.14). Sed_2136 presented the highest values for Chao1 and Shannon index among all the iTag libraries, followed by Ice/Sed_2629. Samples within 1866 cm to 2280 cm are colored black in the plot, and samples within 2288 cm to 2647 cm are colored green.

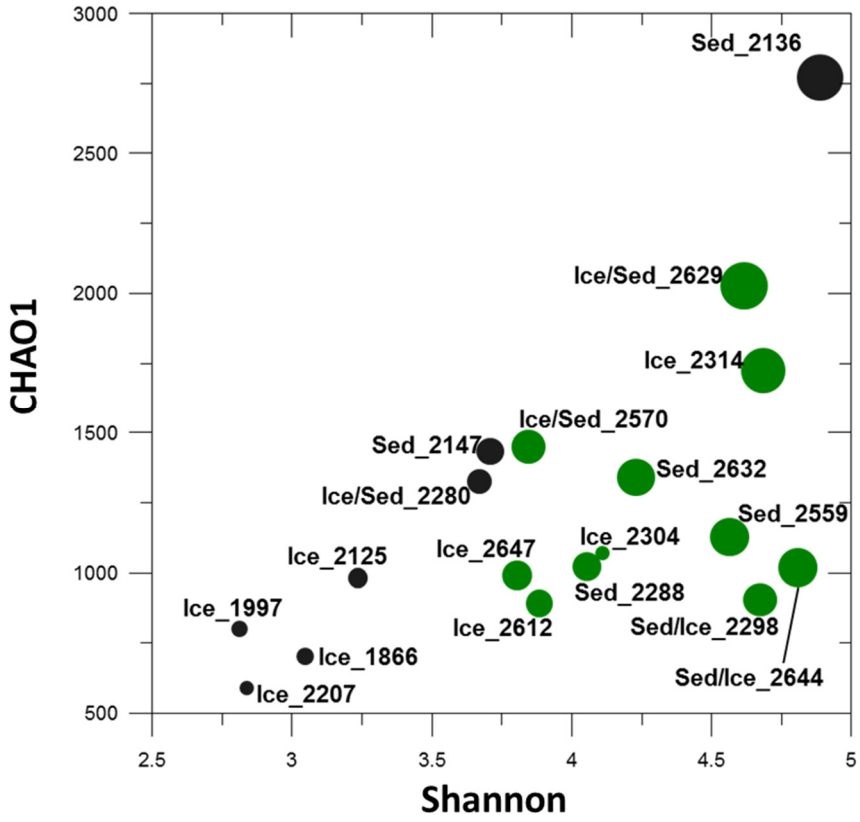



**Figure 3.14.** Plot of richness estimator Chao1 versus diversity index Shannon H' from DNA iTag library reads based on 0.03 distance unit of normalized data. The diameter of the circles correspond to the number of OTUs identified in each library (number of OTUs can be found in table 3.4). Samples collected between 1866 cm and 2280 cm are shown in black, and samples collected between 2288 cm and 2647 cm are shown in green.

Beta diversity was analyzed by the Jaccard coefficient, for similarities within assemblage membership, and $Theta_{yc}$ coefficient, for similarities within microbial assemblage structure (Schloss *et al.*, 2009). Jaccard and $Theta_{yc}$ dendrograms revealed distinct topologies, and the longer branches in the Jaccard dendrogram indicated that the microbial assemblages differed more from each other when membership was considered rather than structure (Figure 3.15). This means that, while the segments had considerably different membership (Jaccard), the relative abundances of the shared OTUs were similar ($Theta_{yc}$). Considering 55% membership similarity (Jaccard), seven clusters were identified (Figure 3.15). Ice_1866 clustered away from the other clusters because of its distinct assemblage dominated by OTUs affiliated with *Acinetobacter, Marinobacter,* and *Sulfurovum*. Three clusters contained only ice segments; one cluster contained only sediment segments; one cluster was formed by ice and Ice/Sed transition zones; and the biggest cluster was formed by sediment, Sed/Ice transition zone, and ice (Figure 3.15A). Considering the 60% structural similarity ($Theta_{yc}$), eight clusters were identified (Figure 3.15B). Indeed, the $Theta_{yc}$ coefficient identified a clearer distinction in clustering segments of the same nature (ice, sediment, and transition zones). These results, added to the high values of the Shannon H' diversity index, indicate that there are both high diversity and high spatial variability of microbial assemblages throughout Lake Vida and that the assemblage structure roughly clusters with sample type.

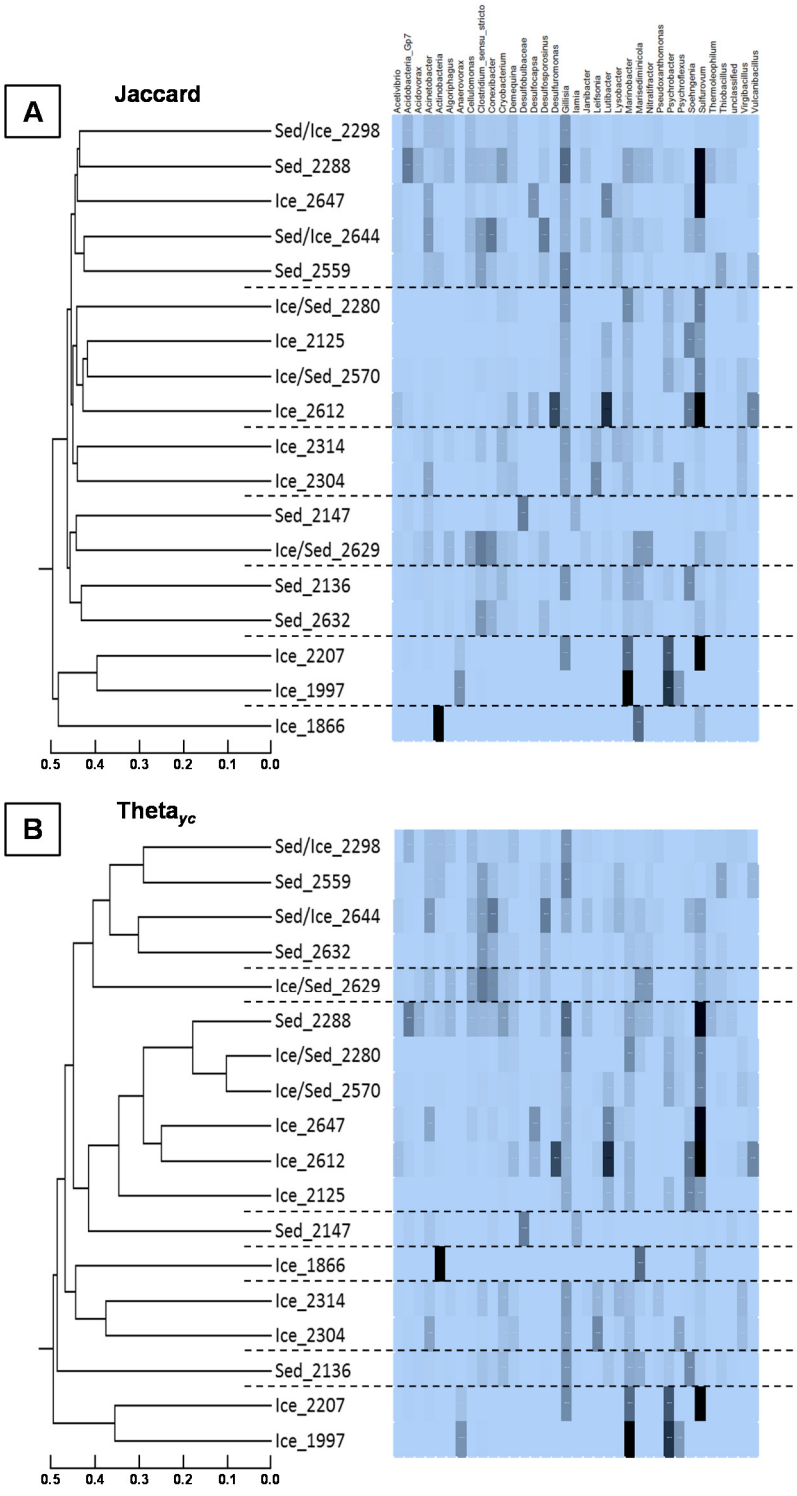



**Figure 3.15.** Clustering of DNA iTag libraries. Dendrogram (A) shows clustering established by the membership-based Jaccard coefficient. Dendrogram (B) shows clustering established by the structure-based Theta$_{yc}$ coefficient. Scale bars represent

Jaccard and Theta$_{yc}$ dissimilarity coefficients (1-coefficient). The heatmap indicates the abundance of OTUs within the microbial assemblages: lightest blue indicates 0% and darkest blue indicates 30% of OTU relative frequency. The dashed lines are used to indicate clustering of samples considering 45% dissimilarity in (A) and 40% dissimilarity in (B).

### 3.3.5. Relative abundance of rRNA (as cDNA) in the sediment layers

The occurrence of the banding pattern of thick (≥10 cm) sediment layers within Lake Vida deep ice is unusual for a lake. Thick sediment layers within ice have been reported in glacier basal ice, but not in polar or alpine lake ice. Therefore, we wanted to further explore the microbial assemblages in these layers by targeting their rRNA gene and also their rRNA (as cDNA). Detection of the rRNA gene is widely used in studies of microbial ecology as a measure of microbial community structure and diversity, regardless of the metabolic state of the cells (Kembel *et al.*, 2012; Blazewicz *et al.*, 2013). Detection of rRNA has been synonymous to microbial activity: the most abundant taxa identified by targeting rRNA would be the most active taxa in the community (Mannisto *et al*., 2013; Hunt *et al*., 2013). However, conflicting patterns between rRNA content and bacterial growth rate indicate that rRNA is not a reliable tool for measuring current activity and in some cases may be completely misleading (Blazewicz *et al*., 2013). It is still a general viewpoint that rRNA can be used as an indicator of potential activity, but not as an indicator of current activity (Blazewicz *et al*., 2013). Therefore, we interpreted our results based on the idea that microbial taxa detected in our cDNA iTag

libraries have the potential for identified metabolic activity, but the cells are not necessarily active at the present time.

OTU composition, distribution, and abundance between DNA and cDNA iTag libraries were distinct enough to separate DNA libraries from cDNA libraries in a clustering assay (Figure 3.16). The dendrograms based on Theta$_{yc}$ and Jaccard coefficients presented the same clustering topology and a strong distinction between the DNA iTag libraries and cDNA iTag libraries. This result indicates that the dominant members detected within the microbial assemblages (based on the rRNA gene) are not the same members that have potential to catalyze protein synthesis (based on rRNA).

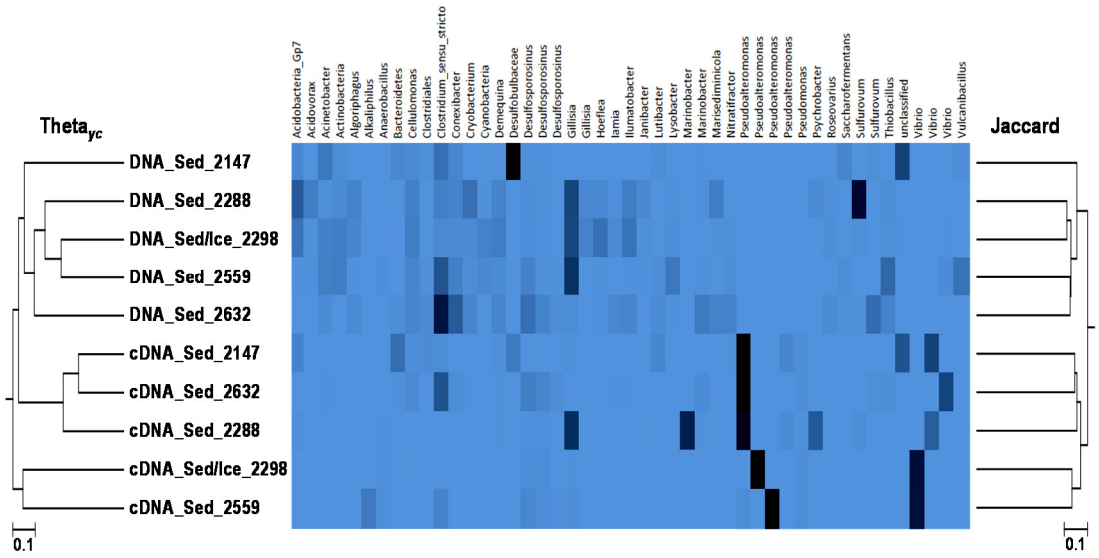



**Figure 3.16.** Clustering between the DNA and cDNA (rRNA) iTag libraries. The dendrogram on the left represents similarity of the samples based on the structure-based Theta$_{yc}$ coefficient. The dendrogram on the right represents similarity of the samples based on the membership-based Jaccard coefficient calculated using Chao1 estimated richness values. The heatmap indicates abundance of OTUs within the microbial assemblages: lightest blue indicates 0% and darkest blue indicates 32% of OTU relative frequency. OTU affiliated taxa are shown at the top of the heatmap.

The dominant OTUs within the sediment segments and transition zones in the DNA iTag libraries were affiliated with the phylum *Actinobacteria,* while the dominant OTUs detected in the cDNA iTag libraries were affiliated with the phylum *Proteobacteria*. As shown in figure 3.16, *Pseudoalteromonas* and *Vibrio*-affiliated OTUs dominated the five cDNA iTag libraries. These OTUs were detected in the DNA iTag libraries among the fewer phylotypes. *Pseudoalteromonas* relative abundance ranged between 19.01% (Sed_2288) to 33.89% (Sed/Ice_2298) of the reads (Table 3.5). *Vibrio*-affiliated reads were the second most abundant OTUs with a relative abundance ranging between 5.57% and 15.23%. Other taxa such as *Ferrimicrobium, Yersinia, Idiomarina, Halomonas,* and *Pseudomonas* also were detected among the fewer phylotypes in the DNA iTag libraries and were detected among the dominant OTUs in the cDNA iTag libraries. Their relative abundance in each cDNA iTag library can be found in Table 3.5. These taxa, in some way, are adapted to salinity and cold and were previously detected in polar regions. For example, *Pseudoalteromonas* sp. and *Vibrio* sp. are common representatives of *Gammaproteobacteria* found in Antarctic marine environments such as sea ice, seawater, and marine sediment (Médigue *et al.*, 2005; Yu *et al.,* 2009). Psychrotrophic *Yersinia* sp. has been reported in Siberian permafrost (Vishnivetskaya *et al.*, 2006), and *Halomonas* sp. and *Idiomarina* sp. are euryhalophiles (grown over a wide range of salinity between 3% and 15% NaCl) and are commonly found in Antarctic saline lakes as well as Antarctic sea ice and seawater (Bowman *et al.*, 2000).

Data produced from the rRNA gene iTag libraries indicated that Lake Vida sediment layers contain a high abundance of microorganisms known to be specialized in

degradation of complex organic polymers. Contrarily, rRNA analyses indicated that the major players carrying metabolic capacity in the same microbial assemblages are microorganisms, mostly aerobe/facultative capable of fermentation or anaerobic respiration in the absence of oxygen (Table 3.5). *Acidobacteria*_Gp7, *Cellulomonas*, *Conexibacter*, *Cryobacterium Marisediminicola*, *Thermoleophilum*, and *Thiobacillus* detected within the microbial assemblages are the most abundant microorganisms in number of cells in the sediment layers as shown by DNA data. However, the cDNA data indicate that the microorganisms containing the highest amount of rRNA in their cells are the genera *Pseudoalteromonas, Vibrio,* and *Marinobacter;* and the higher quantity of rRNA in these cells could be the result of the cell's life history in the lake. In Blazewicz *et al.* (2013) evaluation of the use of rRNA as an indicator of potential activity concluded that rRNA does not represent only potential future activity, but also historical cellular activity. Indeed, some microorganisms increase ribosome concentration as they enter a dormant state (Blazewicz *et al.*, 2013). This provides them with highest protein synthesis potential and also with a hypothetically higher capability to succeed when they return to a vegetative state (when environmental conditions become less favorable; Sukenik *et al.*, 2012). *Pseudoalteromonas, Vibrio,* and *Marinobacter* are organisms known for entering a dormancy phase when facing unfavorable conditions, but no literature was found indicating any evaluation of rRNA abundance in their cells during life history (Piette *et al.*, 2008; Singer *et al.*, 2011; Bari *et al.*, 2013). The microorganisms detected in the cDNA iTag libraries could be explained by their life history (the sequence of events that affected cellular activities up to a given time point added to the subsequent physiological

response to the events) associated with the paleolimnological history of Lake Vida (Stevenson and Schmidt, 2004).

Microbes in the ice of Lake Vida could, in theory, be metabolically active. It has been documented that the Lake Vida brine microbial assemblage presents very low rates of $^{3}$H-leucine incorporation indicating protein synthesis at -12$^{o}$C under anoxic conditions (Murray *et al*., 2012). The major limits to maintaining cellular activity in ice, permafrost, and frozen sediments are temperature, salinity, availability of nutrients, and water activity. Microbial activity (i.e., nutrient uptake, respiration, and minimum metabolism for cell maintenance) was previously reported in these environments, but this activity is at low rates or is undetectable (Karl, 1999; Rivkina *et al.*, 2000; Montross *et al.*, 2014). Synthesis of DNA and protein accompanied by detectable respiration rates were described at -15$^{o}$C basal ice from Taylor Glacier (Doyle *et al*., 2013); photoautotrophic and heterotrophic activities were detected in sediment inclusions in the Lake Bonney ice cover (Paerl and Priscu, 1998; Priscu *et al.,* 1998); low rates of microbial activity were detected in Arctic and Antarctic freezing brine (Mikucki and Priscu, 2007; Lay *et al.*, 2012); and microbial assemblages from Siberian permafrost were reported incorporating $^{14}$C-labeled acetate into lipids at -20$^{o}$C (Rivkina *et al.*, 2000). Considering the findings of these investigations, it is reasonable to assume that the microorganisms trapped in Lake Vida ice and sediment layers may be, to some degree, metabolically active.

**Table 3.5.** Features of the taxa affiliated with the 10 most abundant OTUs detected in the sediment cDNA iTag libraries.

| iTag libray ID | Taxa affiliated to OTUs | Relative Frequency (%) | Oxygen requirement | Phylum or Class |
|---|---|---|---|---|
| **cDNA_Sed_2147** | | | | |
| | *Pseudoalteromonas* | 21.84 | Aerobe/Facultative | Gammaproteobacteria |
| | *Vibrio* | 9.25 | Aerobe/Facultative | Gammaproteobacteria |
| | unclassified Bacteria | 6.77 | Anaerobe/Aerobe/Facultative | not applied |
| | unclassified Bacteroidetes | 3.81 | Anaerobe/Aerobe/Facultative | Bacteroidetes |
| | unclassified Desulfobulbaceae | 2.55 | Obligate anaerobe | Deltaproteobacteria |
| | *Nitratifractor* | 1.94 | Aerobe/Facultative | Epsilonproteobacteria |
| | unclassified Acidobacteria_Gp7 | 1.63 | Aerobe/Facultative | Acidobacteria |
| | *Lutibacter* | 1.36 | Aerobe/Facultative | Bacteroidetes |
| | *Ferrimicrobium* | 1.20 | Aerobe/Facultative | Actinobacteria |
| | unclassified Clostridiales | 0.83 | Obligate anaerobe | Clostridia |
| **cDNA_Sed_2288** | | | | |
| | *Pseudoalteromonas* | 19.01 | Aerobe/Facultative | Gammaproteobacteria |
| | *Marinobacter* | 12.60 | Aerobe/Facultative | Gammaproteobacteria |
| | *Gillisia* | 11.20 | Obligate aerobe | Bacteroidetes |
| | *Psychrobacter* | 6.21 | Aerobe/Facultative | Gammaproteobacteria |
| | *Vibrio* | 5.57 | Aerobe/Facultative | Gammaproteobacteria |
| | *Acidovorax* | 0.66 | Aerobe/Facultative | Betaproteobacteria |
| | *Yersinia* | 0.62 | Aerobe/Facultative | Gammaproteobacteria |
| | *Idiomarina* | 0.61 | Aerobe/Facultative | Gammaproteobacteria |
| | *Sulfurovum* | 0.58 | Aerobe/Facultative | Epsilonproteobacteria |
| | *Halomonas* | 0.55 | Aerobe/Facultative | Gammaproteobacteria |
| **cDNA_Sed/Ice_2298** | | | | |
| | *Pseudoalteromonas* | 33.89 | Aerobe/Facultative | Gammaproteobacteria |
| | *Vibrio* | 14.81 | Aerobe/Facultative | Gammaproteobacteria |
| | *Marinobacter* | 1.53 | Aerobe/Facultative | Gammaproteobacteria |
| | *Idiomarina* | 1.26 | Aerobe/Facultative | Gammaproteobacteria |
| | *Halomonas* | 1.23 | Aerobe/Facultative | Gammaproteobacteria |
| | *Desulfosporosinus* | 1.23 | Obligate anaerobe | Clostridia |
| | *Yersinia* | 1.20 | Aerobe/Facultative | Gammaproteobacteria |
| | *Acidovorax* | 0.67 | Aerobe/Facultative | Betaproteobacteria |
| | *Pseudomonas* | 0.61 | Aerobe/Facultative | Gammaproteobacteria |
| | unclassified Firmicutes | 0.58 | Aerobe/Facultative | Firmicutes |
| **cDNA_Sed_2559** | | | | |
| | *Pseudoalteromonas* | 32.87 | Aerobe/Facultative | Gammaproteobacteria |
| | *Vibrio* | 15.23 | Aerobe/Facultative | Gammaproteobacteria |
| | *Alkaliphilus* | 2.65 | Obligate anaerobe | Clostridia |
| | *Desulfosporosinus* | 1.96 | Obligate anaerobe | Clostridia |
| | *Clostridium_sensu_stricto* | 1.67 | Obligate anaerobe | Clostridia |
| | *Yersinia* | 1.37 | Aerobe/Facultative | Gammaproteobacteria |
| | *Halomonas* | 1.25 | Aerobe/Facultative | Gammaproteobacteria |
| | *Idiomarina* | 0.86 | Aerobe/Facultative | Gammaproteobacteria |
| | *Gillisia* | 0.61 | Obligate aerobe | Bacteroidetes |
| | *Pseudomonas* | 0.58 | Aerobe/Facultative | Gammaproteobacteria |
| **cDNA_Sed_2632** | | | | |
| | *Pseudoalteromonas* | 22.08 | Aerobe/Facultative | Gammaproteobacteria |
| | *Vibrio* | 9.23 | Aerobe/Facultative | Gammaproteobacteria |
| | *Clostridium_sensu_stricto* | 6.72 | Obligate anaerobe | Clostridia |
| | *Desulfosporosinus* | 3.69 | Obligate anaerobe | Clostridia |
| | *Yersinia* | 0.90 | Aerobe/Facultative | Gammaproteobacteria |
| | *Cellulomonas* | 0.87 | Aerobe/Facultative | Actinobacteria |
| | *Albidiferax* | 0.74 | Aerobe/Facultative | Betaproteobacteria |
| | *Marinobacter* | 0.71 | Aerobe/Facultative | Gammaproteobacteria |
| | *Conexibacter* | 0.70 | Obligate aerobe | Actinobacteria |
| | *Halomonas* | 0.68 | Aerobe/Facultative | Gammaproteobacteria |

Alternatively, microbial activity in Lake Vida ice and sediment layers could be static. Cells at dormant state are characterized by undetectable respiratory activity (Mulyukin *et al*., 1996; Demkina *et al.*, 2000); and we could be observing a microbial assemblage well preserved in a frozen, deep, anoxic, stable, isolated environment. Permafrost, frozen sediments, and ice are some of the coldest geological environments on Earth, and organic molecules of any size range (from DNA and RNA to intact cells) can be preserved in these environments for hundreds of thousands of years (Willerslev *et al.,* 2004). Preservation of biological material (cells, DNA, and ribosomes) within Lake Vida ice and sediment layers may lead to the detection and identification of a range of taxa (such as the obligate aerobes *Gillisia, Lutibacter,* and *Conexibacter*), which are not necessarily currently active *in loco*. Without any activity experiment, it is difficult to infer whether or not the microbial assemblages from Lake Vida ice and sediment present any kind of metabolic activity. We must bear in mind that the rRNA detected among the assemblages does not measure which taxa is more or less active; instead, it measures the potential of activity (Blazewicz *et al.,* 2013). Therefore, the taxa that we detected in our rRNA iTag libraries do not necessarily represent the active organisms in our samples, but these taxa may have a higher potential for cellular activity in the lake environment.

Diversity index and abundance estimators indicated that less evenness and more dominance were present among the OTUs from the cDNA iTag libraries when compared to those from the DNA iTag libraries (Table 3.6). This outcome may be a consequence of amplification of total RNA after extraction. Table 3.6 shows a summary of the number of reads, number of OTU counts, and diversity estimators and diversity index before and

after data normalization. The cDNA iTag library that presented the highest diversity index, Sed_2632 (H'= 3.66), also contained the second highest concentration of carbon among the sediment segments (TC = 1.0; Figure 3.5). The comparison between H' index and depth showed the existence of a positive correlation (Spearman $r = 0.20$, $P = 0.000013$). Rarefaction curves indicate that the Sed_2632 cDNA iTag library was the only sample that presented a slope almost asymptotic (Figure 3.17). Overall, alpha diversity results indicated that diversity within the cDNA iTag libraries was low and, as observed within the DNA results, also presented a positive correlation with depth.

**Table 3.6.** DNA and cDNA iTag library data from sediment segments. Summary of the number of reads during pre-processing (raw reads – quality control – chimera check – unique reads) and measures of alpha diversity of the segments analyzed before and after normalization of the data (distance of 0.03).

| iTag library ID | No. of raw reads | No. of reads after quality control | No. of reads after chimera check | No. of unique reads | Cutoff distance of 0.03 | | | No. of reads after resampling | No. of unique reads after resampling | Cutoff distance of 0.03 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | OTU | Chao1 | Shannon's index (H') | | | OTU | Chao1 | ACE | Shannon's index (H') |
| Sed_2147 | 132,096 | 95,490 | 90,479 | 16,819 | 1,945 | 3,955 | 3.77 | 13,043 | 5,135 | 658 | 1,428 | 2,004 | 3.71 |
| cDNA_Sed_2147 | 64,195 | 25,098 | 24,530 | 5,108 | 736 | 1,102 | 3.25 | 13,043 | 3,644 | 557 | 883 | 1,208 | 3.25 |
| Sed_2288 | 39,031 | 24,578 | 23,940 | 5,517 | 753 | 1,260 | 4.05 | 13,043 | 3,895 | 554 | 1,017 | 1,327 | 4.06 |
| cDNA_Sed_2288 | 39,031 | 17,147 | 16,833 | 3,457 | 532 | 784 | 2.8 | 13,043 | 3,554 | 448 | 673 | 703 | 2.77 |
| Sed/Ice_2298 | 61,959 | 37,868 | 36,877 | 9,597 | 893 | 1,355 | 4.69 | 13,043 | 4,785 | 646 | 901 | 892 | 4.68 |
| cDNA_Sed/Ice_2298 | 62,327 | 22,397 | 22,061 | 3,979 | 654 | 1,193 | 2.51 | 13,043 | 3,162 | 476 | 780 | 1,036 | 2.47 |
| Sed_2559 | 73,888 | 38,973 | 36,981 | 8,566 | 1,064 | 1,485 | 4.56 | 13,043 | 4,777 | 757 | 1,123 | 1,128 | 4.57 |
| cDNA_Sed_2559 | 53,077 | 19,382 | 19,081 | 3,581 | 530 | 739 | 2.52 | 13,043 | 3,140 | 438 | 655 | 685 | 2.51 |
| Sed_2632 | 135,205 | 69,012 | 63,378 | 13,177 | 1,137 | 1,875 | 3.92 | 13,043 | 4,204 | 729 | 1,335 | 1,709 | 4.23 |
| cDNA_Sed_2632 | 36,001 | 13,381 | 13,043 | 4,119 | 565 | 706 | 3.66 | 13,043 | 4,942 | 551 | 674 | 655 | 3.66 |
| Total | 696,810 | 363,326 | 347,203 | 73,920 | | | | 130,430 | 41,238 | | | | |

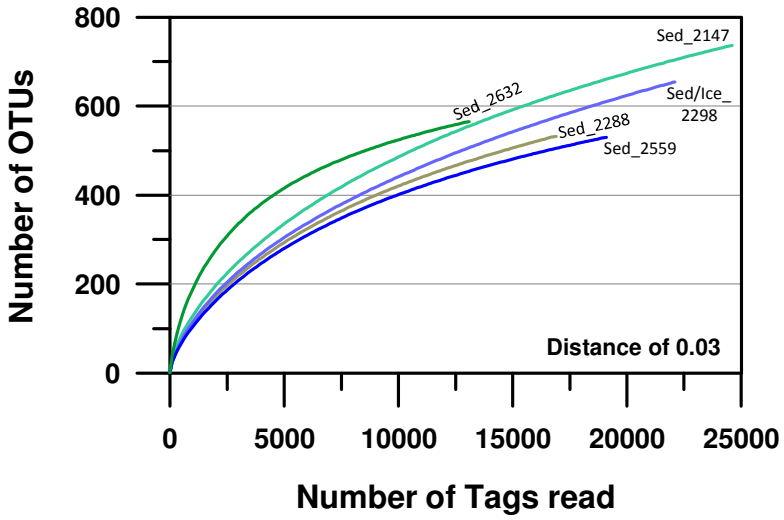



**Figure 3.17.** Rarefaction curve of non-normalized cDNA iTag libraries.

### 3.4. Summary and conclusions

Lake Vida is an unusual Antarctic lake, potentially frozen to its base, with sediment layers banding the ice at different depths, between approximately 3 m to 26.5 m. The upper portion of the lake (from its surface to 16 m depth) was previously explored; and a range of microorganisms was discovered in the ice, mats within the ice, and encapsulated cryogenic brine (Priscu *et al.*, 2005; Mosier *et al.,* 2007; Mondino *et al.*, 2009; Murray *et al.*, 2012). Hitherto, applying 18S and 16S rRNA gene amplification methods, archaeal amplification signal in the lake ice and brine was undetectable or minimal; eukaryotic organisms (diatoms, algae, protists, and fungi) were detected up to depth 15.9 m but were not detected in the brine (Mosier *et al.*, 2007; Murray *et al.*, 2012). The primers used for construction of the iTag libraries were developed to target the *Archaea* and *Bacteria* domains. For that reason, the existence of eukaryotes from depths

18 m up to 27 m was not explored, but chloroplast sequences from diatoms in deep sediment layers were detected. Additionally, microscopy observations confirmed the presence of diatoms known to be benthic taxa previously reported in dry valleys streams, lakes, and soils. Therefore, Lake Vida harbors bacterial and eukaryotic microorganism assemblages from top to bottom accompanied with a rare archaeal assemblage.

The sediment layers within deep ice (a characteristic found in glacial basal ice and permafrost) are not the only uncommon features of Lake Vida. Around the sediment layer at 22.88 m depth (SL2288–2303 cm), we found a shift in concentration of the major anion from $Cl^-$ to $SO_4^{2-}$. We hypothesize that the $SO_4^{2-}$ enriched-ice below SL2288–2303 cm originate from weathering event(s) combined with heavy input of particles enriched by mirabilite and gypsum. The switch in geochemistry above and below SL2288–2303 cm was accompanied by a distinction in the abundance of OTUs affiliated with the genera *Marinobacter, Cellulomonas*, *Janibacter*, and *Lysobacter*. *Marinobacter*-affiliated OTUs were significantly more abundant above SL2288–2303 cm, where $Cl^-$ is the dominant anion, and the other three genera were significantly more abundant below SL2288–2303 cm. Furthermore, microbial diversity below SL2288–2303 cm is statistically higher than microbial diversity above the layer. The event that caused the change in geochemistry and microbial assemblage above and below SL2288–2303 cm is unknown and may be discovered via a paleolimnological history of the lake. As shown in the PCA, $SO_4^{2-}$ concentrations, TDS and $Na^+$ concentrations explained a significant proportion of the microbial assemblage variation in the lake. Therefore, the origin of $Cl^-$

and $SO_4^{2-}$ appear to be an important factor controlling distribution of microbial assemblages within Lake Vida's ice.

Lake Vida contains a high number of cells within its ice core, sediment layers, and brine. The numbers are higher than the cell enumerations from comparable environments such as glacial ice, dry valley lake ice and Antarctic lake water and are more similar to cell enumerations from dry valley soils. We found a positive correlation between depth and cell count, which is most likely associated with the presence of sediments in the lower portion of the lake ice. Distinctions in cell enumeration between ice and sediment were evident and accompanied by a distinction in the major OTUs detected in each sample type. In general, the ice segments were dominated by heterotrophic (such as *Lutibacter* and *Desulforomonas*) and chemolithoautotrophic microorganisms (such as *Sulfurovum* and *Desulfocapsa*), which are known to oxidize reduced sulfur compounds as an energy source (ATP). The sediment segments (including the transition zone) were dominated by heterotrophic microorganisms (such as *Acidobacteria*_Gp7, *Cellulomonas*, and *Conexibacter*) known to be abundant in soils and specialized in oxidize acetate and complex organic compounds. However, rRNA screening indicated that the major players carrying metabolic capacity in the sediment layers are mostly heterotrophic bacteria affiliated with the genera *Pseudoalteromonas* and *Vibrio.* Considering the linkage between phylogeny and function (Langille *et al.,* 2013), the association of rRNA gene sequences to taxonomically classified organisms provided important information about the composition and, to some extent, potential physiological and ecological characteristics of Lake Vida microbial assemblages. rRNA detected within

the sediment layer can be directly associated with the potential to catalyze a specific function of protein synthesis by few populations and can therefore explain the relative expression of this function. For these reasons, we concluded that the lower portion of the Lake Vida ice cover (from 18 m to 27 m depth) harbors a diverse and cell-rich microbial assemblage composed of bacteria affiliated with microorganisms specialized in the reduction and oxidation of sulfur compounds, with the ability to degrade complex polymers. These microbial assemblages, which are comprised mostly of obligate aerobes or facultative aerobes, must contend with surrounding conditions of low temperature, low water activity, elevated salinity, no light, no oxygen, and sufficient carbon to support heterotrophic activity. The assumption that Lake Vida's microbial assemblage currently presents any rate of activity remains to be resolved. If no cellular activity is established (meaning that life in Lake Vida became static in some point during the lake's history), we can take one step further and infer that the environmental conditions of the lake did a good job preserving cellular and biological components. Therefore, interest in exploring and understanding Lake Vida's microbial assemblage from top to bottom lies in both biogeochemistry reactions and interactions of the microbial assemblages with Lake Vida's environment and also in understanding the origin, conservation, survival, and stability of deposited microorganisms in the lake.

### 3.5. Acknowledgments

EK was supported by Fulbright/CAPES-BRAZIL grant 2163-08-8. Funding for the Lake Vida project was provided by NSF grants ANT-0739681 to AEM and CHF, and ANT-0739698 to PTD. iTag sequencing was supported by an award to AEM, JGI-634, by the DOE Joint Genome Institute. Thanks to Vivian Peng, DRI, for helping in the RNA extractions; Robert Read, DRI, for guidance in computational analyses; Dr. Andrew S. Ichimura, San Francisco State University for providing support for the electron microscopy observations; and to K.C. King, DRI for reviewing this chapter and providing valuable feedback. Special thanks to members of the 2010 Lake Vida Expedition team: Dr. Peter Doran, Dr. Alison Murray, Dr. Chris Fritsen, Hilary Dugan, Dr. Bernd Wagner, Dr. Fabien Kenig, Dr. Brian Glazer, Dr. Seth Young, Peter Glenday, and Jay Kyne. Thanks also to Raytheon Polar Services for logistic support.

## 3.6. References

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 & Biochemistry, 38,* 3041–3056.

Antibus, D. E., Leff, L. G., Hall, B. L., Baeseman, J. L., & Blackwood, C. B. (2012). Cultivable bacteria from ancient algal mats from the McMurdo Dry Valleys, Antarctica. *Extremophiles: Life under Extreme Conditions*, *16*(1), 105–114.

Aronesty, E. (2011). Command-line tools for processing biological sequencing data. ea-utils. Expression Analysis, Durham, NC.

Babalola, O. O., Kirby, B. M., Le Roes-Hill, M., Cook, A. E., Cary, S. C., Burton, S. G., & Cowan, D. A. (2009). Phylogenetic analysis of actinobacterial populations associated with Antarctic Dry Valley mineral soils. *Environmental Microbiology*, *11*(3), 566–576.

Bari, S. M. N., Roky, M. K., Mohiuddin, M., Kamruzzaman, M., & Mekalanos, J. J. (2013). Quorum-sensing autoinducers resuscitate dormant *Vibrio cholerae* in environmental water samples. *Proceedings of the National Academy of Sciences, 110*(24), 9926–9931.

Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G., & Neufeld, J. D. (2011). Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Applied and Environmental Microbiology*, *77*(11), 3846–3852.

Blazewicz, S. J., Barnard, R. L., Daly, R. A, & Firestone, M. K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *The ISME Journal*, *7*(11), 2061–2268.

Bottos, E. M., Woo, A. C., Zawar-Reza, P., Pointing, S. B., & Cary, S. C. (2013). Airborne bacterial populations above desert soils of the McMurdo Dry Valleys, Antarctica. *Microbial Ecology*, *67*, 120–128.

Bowman, J. P., McCammon, S. A, Rea, S. M., & McMeekin, T. A. (2000). The microbial composition of three limnologically disparate hypersaline Antarctic lakes. *FEMS Microbiology Letters*, *183*(1), 81–88.

Brady, H. T., & Batts, B. (1981). Large salt beds on the surface of the Ross Ice Shelf near Black Island, Antarctica. *Journal of Glaciology, 27*, ll-18.

Calkin, P. E. (1971). Glacial geology of the Victoria Valley system, southern Victoria Land, Antarctica. In *Antarctic Research Series vol 16.* Washington, DC: American Geophysical Union, 363–412.

Calkin, P. E., & Bull, C. (1967). Lake Vida, Victoria Valley, Antarctica. *Journal of Glaciology*, *6*(48), 833–836.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-lyons, D., Lozupone, C. A., Turnbaugh, P. J., et at. (2010). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences, 108*(1), 4516–4522.

Caporaso, J. G., Lauber, C. L., Walters, W. A, Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, *6*(8), 1621–1624.

Carlson, G. W. (1913). Süsswasseralgen aus der Antarktis, Süd-Georgien und den Falkland Inseln. *In Wisserschaftliche Ergebnisse schwedische Südpolar-Expedition 1901–1903, Botanik*, *4,* 1–194.

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 in Microbiology, 8*(2)*,* 129-138.

Chao, A. (1984). Nonparametric-estimation of the number of classes in a population. *Scandinavian Journal of Statistics, 11*, 265–270.

Chao, A., & Lee S-M. (1992). Estimating the number of classes via sample coverage. *Journal of the American Statistical Association, 87*, 210–217.

Christner, B. C., Doyle, S. M., Montross, S. N., Skidmore, M. L., Samyn, D., Lorrain, R., Tison, J., & Fitzsimons, S. (2010). A subzero microbial habitat in the basal ice of an Antarctic glacier. AGU Fall Meeting 2010, Abstract B21F-04.

Christner, B. C., Kvitko, B. H., & Reeve, J. N. (2003). Molecular identification of bacteria and eukarya inhabiting an Antarctic cryoconite hole. *Extremophiles: Life under Extreme Conditions*, *7*(3), 177–183.

Claesson, M. J., O'Sullivan, O., Wang, Q., Nikkilä, J., Marchesi, J. R., Smidt, H., et al. (2009). Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PloS One*, *4*(8), e6669.

Degnan, P. H., & Ochman, H. (2012). Illumina-based analysis of microbial community diversity. *The ISME Journal*, *6*(1), 183–194.

Deming, J. W. (2002). Psychrophiles and polar regions. *Current Opinion in Microbiology*, *5*(3), 301–309.

Demkina, E. V, Soina, V. S., El Registan, G. I., & Zviagintsev, D. G. (2000). Reproductive resting forms of *Arthrobacter globiformis*. *Mikrobiologiia*, *69*(3), 377–382.

Doran, P. T., Fritsen, C. H., McKay, C. P., Priscu, J. C., & Adams, E. E. (2003). Formation and character of an ancient 19-m ice cover and underlying trapped brine in an "ice-sealed" east Antarctic lake. *Proceedings of the National Academy of Sciences, 100*(1), 26–31.

Doran, P. T., Fritsen, C. H., Murray, A. E., Kenig, F., Mckay, C. P., & Kyne, J. D. (2008). Entry approach into pristine ice-sealed lakes - Lake Vida, East Antarctica, a model ecosystem. *Limnology and Oceanography Methods*, *6*, 542–547.

Doran, P. T., Wharton, R. A, & Lyons, W. B. (1994). Paleolimnology of the McMurdo Dry Valleys, Antarctica. *Journal of Paleolimnology*, *10*, 85–114.

Doyle, S., Montross, S., Skidmore, M., & Christner, B. (2013). Characterizing microbial diversity and the potential for metabolic function at −15°C in the basal ice of Taylor Glacier, Antarctica. *Biology*, *2*(3), 1034–1053.

Esposito, R. M. M., Spaulding, S. A., McKnight, D. M., Van de Vijver, B., Kopalová, K., et al. (2008). Inland diatoms from the McMurdo Dry Valleys and James Ross Island, Antarctica. *Botany*, *86*(12), 1378–1392.

Finster, K. W., Kjeldsen, K. U., Kube, M., Reinhardt, R., Mussmann, M., Amann, R., & Schreiber, L. (2013). Complete genome sequence of *Desulfocapsa sulfexigens*, a marine deltaproteobacterium specialized in disproportionating inorganic sulfur compounds. *Standards in Genomic Sciences*, *8*(1), 58–68.

Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W., & Delong, E. F. (2008). Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences, 105*(10), 3805–3810.

Friedrich, C. G., Rother, D., Bardischewsky, F., & Quentmeier, A. (2001). Oxidation of reduced inorganic sulfur compounds by bacteria: emergence of a common mechanism? *Applied and Environmental Microbiology*, *67*(7), 2873–2882.

Gibson, G. W. (1962). Geological investigations in Southern Victoria Land, Antarctica. *New Zealand Journal of Geology and Geophysics*, *5*(3), 361–374.

Gilbert, J. A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., Somerfield, P. J., et al. (2009). The seasonal structure of microbial communities in the Western English Channel. *Environmental Microbiology*, *11*(12), 3132–3139.

Gilichinsky, D. A, Wilson, G. S., Friedmann, E. I., McKay, C. P., Sletten, R. S., Rivkina, E. M., et al. (2007). Microbial populations in Antarctic permafrost: biodiversity, state, age, and implication for astrobiology. *Astrobiology*, *7*(2), 275–311.

Glatz, R. E., Lepp, P. W.,Ward, B. B., & Francis, C. A. (2006). Planktonic microbial community composition across steep physical/chemical gradients in permanently ice-covered Lake Bonney, Antarctica. *Geobiology*, *4*, 53–67.

Gow, A. J., Epstein. S., & Sheehy, W. (1979). On the origin of stratified debris in ice cores from the bottom of the Antarctic ice sheet. *Journal of Glaciology*, *23*, 185-192.

Hall, B. L., Denton, G. H., Fountain, A. G., Hendy, C. H., & Henderson, G. M. (2010). Antarctic lakes suggest millennial reorganizations of Southern Hemisphere atmospheric and oceanic circulation. *Proceedings of the National Academy of Sciences, 107*(50), 21355–21359.

Hall, B. L., Denton, G. H., Overturf, B., & Hendy, C. H. (2002). Glacial Lake Victoria, a high-level Antarctic Lake inferred from lacustrine deposits in Victoria Valley. *Journal of Quaternary Science*, *17*(7), 697–706.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*, 417–441.

Hunt, D. E., Lin, Y., Church, M. J., Karl, D. M., Tringe, S. G., Izzo, L. K. et al. (2013). Relationship between abundance and specific activity of bacterioplankton in open ocean surface waters. *Applied and Environmental Microbiology, 79*(1), 177–184.

Inagaki, F., Takai, K., Nealson, K. H., & Horikoshi, K. (2004). *Sulfurovum lithotrophicum* gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the epsilon-Proteobacteria isolated from Okinawa Trough hydrothermal sediments. *International Journal of Systematic and Evolutionary Microbiology*, *54*(Pt 5), 1477–1482.

Jaccard, Paul. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura, *Bulletin de la Société Vaudoise des Sciences Naturelles 37*, 547–579.

Jo, W., & Kim, M. J. (2013). Influence of the photothermal effect of a gold nanorod cluster on biofilm disinfection. *Nanotechnology*, *24*(19), 195104-195112.

Karl, D. M. (1999). Microorganisms in the accreted ice of Lake Vostok, Antarctica. *Science*, *286*(5447), 2144–2147.

Kembel, S. W., Wu, M., Eisen, J. A, & Green, J. L. (2012). Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Computational Biology*, *8*(10), e1002743.

Keys, J. R., & Williams, K. (1981). Origin of crystalline, cold desert salts in the McMurdo region, Antarctica. *Geochimica et Cosmochimica Acta, 45*, 2299-2309.

Kirchman, D. L. (2002). The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS Microbiology Ecology*, *39*(2), 91–100.

Knight, P. G. (1997). The basal ice layer of glaciers and ice sheets. *Quaternary Science Reviews*, *16*(97), 975–993.

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A, et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, *31*(9), 814–821.

Lay, C.-Y., Mykytczuk, N. C. S., Niederberger, T. D., Martineau, C., Greer, C. W., & Whyte, L. G. (2012). Microbial diversity and activity in hypersaline high Arctic spring channels. *Extremophiles: Life under Extreme Conditions*, *16*(2), 177–1791.

Laybourn-Parry, J., & Pearce, D. A. (2007). The biodiversity and ecology of Antarctic lakes: models for evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1488), 2273–2289.

Macalady, J. L., Lyon, E. H., Koffman, B., Albertson, L. K., Meyer, K., Galdenzi, S., & Mariani, S. (2006). Dominant microbial populations in limestone-corroding stream biofilms, Frasassi cave system, Italy. *Applied and Environmental Microbiology*, *72*(8), 5596–5609.

Mackelprang, R., Waldrop, M. P., DeAngelis, K. M., David, M. M., Chavarria, K. L., Blazewicz, S. J., et al. (2011). Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature*, *480*(7377), 368–371.

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963.

Mangoni, M. L., Papo, N., Barra, D., Simmaco, M., Bozzi, A., Di Giulio, A., & Rinaldi, A. C. (2004). Effects of the antimicrobial peptide temporin L on cell morphology, membrane permeability and viability of *Escherichia coli*. *The Biochemical Journal*, *380*(3), 859–65.

Mannisto, M. K., Kurhela, E., Tiirola, M., & Haggblom, M. M. (2012). *Acidobacteria* dominate the active bacterial communities of Arctic tundra with widely divergent winter-time snow accumulation and soil temperatures. *FEMS Microbiology Ecology*, *84*, 47–59.

Marteinsson, V. T., Rúnarsson, Á., Stefánsson, A., Thorsteinsson, T., Jóhannesson, T., Magnússon, S. H., et al. (2013). Microbial communities in the subglacial waters of the Vatnajökull ice cap, Iceland. *The ISME Journal*, *7*(2), 427–37.

Massana, R., Murray, A. E., Preston, C. M., & DeLong, E. F. (1997). Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Applied and Environmental Microbiology*, *63*(1), 50–56.

Médigue, C., Krin, E., Pascal, G., Barbe, V., Bernsel, A., Bertin, P. N., et al. (2005). Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Research*, *15*(10), 1325–1335.

Mikucki, J. A., & John C. Priscu. (2007). Bacterial diversity associated with Blood Falls, a subglacial outflow from the Taylor Glacier. *Applied and Environmental Microbiology*, *73*(12), 4029–4039.

Mondino, L. J., Asao, M., & Madigan, M. T. (2009). Cold-active halophilic bacteria from the ice-sealed Lake Vida, Antarctica. *Archives of Microbiology*, *191*(10), 785–90.

Montross, S., Skidmore, M., Christner, B., Samyn, D., Tison, J.-L., Lorrain, R., et al. (2014). Debris-rich basal ice as a microbial habitat, Taylor Glacier, Antarctica. *Geomicrobiology Journal*, *31*(1), 76–81.

Mora, S. J. De, Whitehead, R. F., & Gregory, M. (2004). The chemical composition of glacial meltwater ponds and streams on the McMurdo Ice Shelf, Antarctica. *Antarctic Science*, *6*(1), 17–27.

Mosier, A. C., Murray, A. E., & Fritsen, C. H. (2007). Microbiota within the perennial ice cover of Lake Vida, Antarctica. *Fems Micriobiology Ecology*, *59*(2), 274–288.

Mulyukin, A. L., Kozlova, A. N., Kaprel'yants, A. S., & El'-Registan, G. I. (1996). Detection and investigation of the dynamics of autoregulatory factor $d_1$ in the culture liquid and cells of *Micrococcus luteus*. *Mikrobiologiya*, *65*(1), 20-25.

Murray, A. E., Kenig, F., Fritsen, C. H., McKay, C. P., Cawley, K. M., Edwards, R., et al. (2012). Microbial life at -13°C in the brine of an ice-sealed Antarctic lake. *Proceedings of the National Academy of Sciences, 109*(50), 20626–20631.

Nichols, R. L. (1990). Geology of Lake Vanda, Wright Valley, South Victoria Land, Antarctica. *Antarctic Research Series*, *7*, 47–52.

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), 1713–1724.

Paerl, H., & Priscu, J. (1998). Microbial phototrophic, heterotrophic, and diazotrophic activities associated with aggregates in the permanent ice cover of Lake Bonney, Antarctica. *Microbial Ecology*, *36*(3), 221–230.

Park, S.-J., Ghai, R., Martín-Cuadrado, A.-B., Rodríguez-Valera, F., Jung, M.-Y., Kim, J.-G., & Rhee, S.-K. (2012). Draft genome sequence of the sulfur-oxidizing bacterium "*Candidatus* Sulfurovum sediminum" AR, which belongs to the *Epsilonproteobacteria*. *Journal of Bacteriology*, *194*(15), 4128–4129.

Piette, F., Struvay, C., Godin, A., Cipolla, A., & Feller, G. (2008). Life in the cold: proteomics of the Antarctic bacterium *Pseudoalteromonas haloplanktis*. *In* Proteomic Applications in Biology, Dr. Joshua Heazlewood (Ed.), pp 93-114.

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, 106*(47), 19964–19969.

Priscu, J. C. (1998). Perennial Antarctic lake ice: an oasis for life in a polar desert. *Science*, *280*(5372), 2095–2098.

Priscu, J. C., Adams, E. E., Lyons, W. B., Voytek, M. A., Mogk, D. W., Brown, R. L., et al. (1999). Geomicrobiology of subglacial ice above Lake Vostok, Antarctica. *Science*, *286*(5447), 2141–2144.

Priscu, J. C., Adams, E. E., Paerl, H. W., Fritsen, C. H., Dore, J. E., John, T., et al. (2005). Perennial Antarctic lake ice: a refuge for cyanobacteria in an extreme environment. *In* Life in Ancient Ice, J.D.C. and S.O.R. (Ed.), Princeton University Press, 22–49.

Rivkina, E. M., Friedmann, E. I., McKay, C. P., & Gilichinsky, D. A. (2000). Metabolic activity of permafrost bacteria below the freezing point. *Applied and Environmental Microbiology*, *66*(8), 3230–3233.

Rogers, S., Shtarkman, Y., Koçer, Z., Edgar, R., Veerapaneni, R., & D'Elia, T. (2013). Ecology of subglacial Lake Vostok (Antarctica), based on metagenomic/metatranscriptomic analyses of accretion ice. *Biology*, *2*(2), 629–650.

Salvatore, M. R., Mustard, J. F., Head, J. W., Marchant, D. R., & Wyatt, M. B. (2013). Characterization of spectral and geochemical variability within the Ferrar Dolerite of the McMurdo Dry Valleys, Antarctica: weathering, alteration, and magmatic processes. *Antarctic Science*, *26*(01), 49–68.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423.

Schloss, P. D., & Westcott, S. L. (2011). Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Applied and Environmental Microbiology*, *77*(10), 3219–3226.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541.

Shi, Y., Tyson, G. W., & DeLong, E. F. (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature, 459*, 266–269.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(539), 1-6.

Singer, E., Webb, E. A, Nelson, W. C., Heidelberg, J. F., Ivanova, N., Pati, A., & Edwards, K. J. (2011). Genomic potential of *Marinobacter aquaeolei*, a biogeochemical "opportunitroph". *Applied and Environmental Microbiology*, *77*(8), 2763–2771.

Skidmore, M., Anderson, S. P., Sharp, M., & Lanoil, B. D. (2005). Comparison of microbial community compositions of two subglacial environments reveals a possible role for microbes in chemical weathering processes. *Applied and Environmental Microbiology*, *71*(11), 6986–6997.

Sleewaegen, S., Lorrain, R., Offer, Z., Azmon, E., Fitzsimons, S., & Souchez, R. (2002). Trapping of aeolian sediments and build-up of the ice cover of a dry-based Antarctic lake. *Earth Surface Processes and Landforms*, *27*(3), 307–315.

Spaulding, S. A., Mcknight, D. M., Stoermer, E. F., & Doran, P. T. (1997). Diatoms in sediments of perennially ice-covered Lake Hoare, and implications for interpreting lake history in the McMurdo Dry Valleys of Antarctica. *Journal of Paleolimnology*, *17*(9), 403–420.

Squyres, S. W., Andersen, D. W., Nedell, S. S., & Wharton, R. A. (1991). Lake Hoare, Antarctica: sedimentation through a thick perennial ice cover. *Sedimentology*, *38*, 363-379.

Stackebrandtt, E., Kandler, O., & Atcc, C. (1979). Taxonomy of the genus *Cellulomonas*, based on phenotypic characters and deoxyribonucleic acid-deoxyribonucleic acid homology, and proposal of seven neotype strains. *International Journal of Systematic Bacteriology, 29*(4), 273–282.

Stanish, L. F., Kohler, T. J., Esposito, R. M. M., Simmons, B. L., Nielsen, U. N., Wall, D. H., et al. (2012). Extreme streams: flow intermittency as a control on diatom communities in meltwater streams. *Journal of the North American Benthological Society*, *15*(2), 1–15.

Stevenson, B. S., & Schmidt, T. M. (2004). Life history implications of rRNA gene copy number in *Escherichia coli. Applied and Environmental Microbiology*, *70*(11), 6670–6677.

Sukenik, A., Kaplan-Levy, R. N., Welch, J. M., & Post, A. F. (2012). Massive multiplication of genome and ribosomes in dormant cells (akinetes) of *Aphanizomenon ovalisporum* (Cyanobacteria). *The ISME Journal*, *6*(3), 670–679.

Takai, K., & Horikoshi, K. (2000). Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. *Applied and Environmental Microbiology, 66*(11), 5066–5072.

Tang, C., Madigan, M. T., & Lanoil, B. (2013). Bacterial and archaeal diversity in sediments of west Lake Bonney, McMurdo Dry Valleys, Antarctica. *Applied and Environmental Microbiology*, *79*(3), 1034–1038.

Torii, T., & N. Yamagata (1981). Limnological studies of saline lakes in the Dry Valleys. *In* Dry Valley Drilling Project, L. D. McGinnis (Ed.). *Antarctic Research Series, 33*, 141-160.

Ugolini, F. C., Deutsch, W., & Harris, H. J. H. (1981). Chemistry and clay mineralogy of selected cores from the Antarctic Dry Valley Drilling Project. *In* Dry Valley Drilling Project, L. D. McGinnis (Ed.). *Antarctic Research Series*, *33*, 315-329.

Vishnivetskaya, T. A, Petrova, M. A, Urbance, J., Ponder, M., Moyer, C. L., Gilichinsky, D. A, & Tiedje, J. M. (2006). Bacterial community in ancient Siberian permafrost as characterized by culture and culture-independent methods. *Astrobiology*, *6*(3), 400–414.

Wang, X., Hu, M., Xia, Y., Wen, X., & Ding, K. (2012). Pyrosequencing analysis of bacterial diversity in 14 wastewater treatment systems in China. *Applied and Environmental Microbiology*, *78*(19), 7042–7047.

Ward, B. B., & Priscu, J. C. (1997). Detection and characterization of denitrifying bacteria from a permanently ice-covered Antarctic Lake, *Hydrobiologia*, *347*(3), 57–68.

West, W., & West, G. S. (1911). Freshwater algae. *In British Antarctic Expedition (1907–1909) Science Report, Biology, 1*, 263–298.

Willerslev, E., Hansen, A. J., & Poinar, H. N. (2004). Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology & Evolution*, *19*(3), 141–147.

Wright, E. S., Yilmaz, L. S., & Noguera, D. R. (2012). DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Applied and Environmental Microbiology*, *78*(3), 717–725.

Yu, Y., Li, H., Zeng, Y., & Chen, B. (2009). Phylogenetic diversity of culturable bacteria from Antarctic sandy intertidal sediments. *Polar Biology*, *33*(6), 869–875.

Yue, J. C., & Clayton, M. K. (2005). A similarity measure based on species proportions. *Communication in Statistics - Theory and Methods*, *34*(11), 2123-2131.

# Chapter IV

## Characterization of Thaumarchaeota marine Group I.1a environmental genome fragments from Antarctic Peninsula coastal surface seawaters


### Abstract

Distribution of the ubiquitous and abundant chemolithoautotrophic, ammonia-oxidizing Thaumarchaeota in marine environments is now well documented. In winter Antarctic Peninsula surface seawaters, Thaumarchaeota (earlier identified as marine group I.1a) has been reported to compose a significant fraction of bacterioplankton (10–30%) and to play a noteworthy role in carbon fixation coupled to ammonia oxidation during wintertime. Here, we analyzed 926,568 base pairs (bp) of Antarctic marine surface water environmental DNA in large-insert fosmid libraries associated with Thaumarchaeota. The goal was to characterize and categorize the phylogenomics of the Antarctic assemblage (Antarctic Group I.1a) by comparative genomic analyses. This process involved comparing Antarctic Group I.1a fosmids to five thaumarchaeal genome references. The

results revealed a rearranged genomic structure within Antarctic Group I.1a containing genes with high homology to the five genome references. We identified that 30% of the Antarctic Group I.1a open reading frames (ORFs) were rearranged in comparison to the genome of *Nitrosopumilus maritimus* SCM1. Genomic regions with highly preserved synteny, such as the gene neighborhood flanking the ammonia monooxygenase operon and chaperone operon were also observed. Results from the analyses of gene synteny between the Antarctic Group I.1a fosmids, the high nucleotide identity between fosmids during assemblage to larger scaffolds, and the high identity among ribosomal proteins within the fosmids (nucleotide identity higher than 92.8%) suggested that the Antarctic Group I.1a represents a very low diversity Thaumarchaeota Group I.1a assemblage. The average of nucleotide identity (ANI) score of approximately 82% between the Antarctic Group I.1a and reference genomes indicated that Antarctic Group I.1a represents a "*Ca.* Nitrosopumilus" not yet exemplified by any enriched or isolated Thaumarchaeota organism. Pairwise distance between homologous proteins indicated that Antarctic Group I.1a is more closely related to the Arctic isolate "*Ca.* Nitrosopumilus sedimentis" AR2 than the other thaumarchaeal genomes investigated. Topology of phylogenetic trees of concatenated proteins varies with the protein trees calculated – although SCM1, AR1, and AR2 were observed branching together and apart from Antarctic Group I.1a. Therefore, our findings suggest enrichment and isolation of a Thaumarchaeota Group I.1a strain from Antarctic oceans for future analyses to better evaluate physiological adaptation of Thaumarchaeota Group I.1a to the Antarctic marine environment.

## 4.1. Introduction

In two recent studies, Borchier-Armanet *et al*. (2008) and Pester *et al*. (2011) proposed that the Archaea phylum Thaumarchaeota marine Group I.1a (Group I.1a) embraces a diverse group of uncultivated mesophilic and psychrophilic Crenarchaeota widely distributed in different ecosystems (e.g., seawater, freshwater, soil, and sediment) (Brochier-Armanet *et al.,* 2008; Pester *et al.*, 2011). In marine environments, Group I.1a is known to be a highly abundant group inhabiting deep oceans (Church *et al.,* 2010; Santoro, *et al*., 2010; Amano-Sato *et al*. 2013), oxygen-deficient zones (Labrenz *et al*., 2010; Molina *et al.*, 2010), and surface waters of the Southern Ocean (Massana *et al*., 1998; Murray *et al.*, 1999; Grzymski *et al.*, 2012). After the discovery that some Thaumarchaeota are able to assimilate carbon autotrophically and oxidize ammonia aerobically, it was recognized that the new phylum plays an important role in global carbon and ammonia assimilation (Könneke *et al.*, 2005). Ammonia-oxidizing archaea (AOA) are among the most abundant organisms on Earth, although AOA have proven difficult to isolate in culture and many questions remain about their physiology, metabolism, and ecology within distinct environments.

*Nitrosopumilus maritimus* SCM1 was the first cultivated AOA that grew chemolithoautotrophically by aerobically oxidizing ammonia to nitrite and also was the first completed genome sequenced from a free-living AOA (Könneke *et al*., 2005; Walker *et al.,* 2010). Since then, other AOA such as "*Candidatus* Nitrosoarchaeon limnia" SFB1 (Blainey *et al*., 2011), "*Ca.* Nitrosopumilus koreensis" AR1 (Park *et al.*, 2012a), and "*Ca.* Nitrosopumilus sedimentis" AR2 (Park *et al*., 2012b) have been

enriched from geographically distinct environments, and genomic databanks have been loaded by Thaumarchaeota Group I.1a genomes. Comparative genomics indicated that the average nucleotide identity (ANI ; a robust measurement of genomic relatedness among strains) between *N. maritimus* SCM1 and the other AOA range from 70% to 85%. For example, "*Ca.* Nitrosopumilus koreensis" AR1 presents high genomic similarities to *N. maritimus* SCM1 with 71.9% of its genes homologous to *N. maritimus* SCM1 genes, and the genome's relatedness is supported by an ANI of 85% (Park *et al.,* 2012a). "*Ca.* Nitrosopumilus sedimentis" AR2 presents only 65.1% of its genes homologous to *N. maritimus* SCM1 genes and ANI relatedness of approximately 79% (Park *et al.,* 2012b). Despite the similarities between genomes, dissimilarities such as a unique set of genes and low amino acid identity, in comparison to the *N. maritimus* SCM1 genome, indicates that the Group I.1a genomes from different AOA might offer distinct genetic potential that might be environmentally oriented (Blainey *et al.*, 2011; Park *et al.,* 2012b).

The gene content and gene organization of *N. maritimus* SCM1 have been shown to be highly similar to environmental populations represented by genome fragments from marine environments suggesting that the *N. maritimus* SCM1 genome is representative of many of the globally abundant Group I.1a (Walker *et al.,* 2010). From Antarctic seawater, an environmental genome fragment, fosmid 74A4 (Béjà *et al.,* 2002), was shown to share similarities with the *N. maritimus* SCM1 genome – revealing, for the first time, the functional diversity of Group I.1a in Antarctic oceans. Fosmid 74A4 shared high synteny with portions of the *N. maritimus* SCM1 genome and 59% of its open reading frames (ORFs) presented ≥ 40% amino acid identity to ORFs codified by *N.*

*maritimus* SCM1genome (Walker *et al.,* 2010). Additionally, fosmid 74A4 contained one 16S rRNA gene that was 98% identical to the *N. maritimus* SCM1 rRNA gene, confirming the close phylogenetic affiliation between both (Walker *et al.,* 2010).

A metagenomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters showed, by targeting the 16S rRNA through clone libraries, that 84% of the clone library from the winter assemblage was closely affiliated (> 99% sequence identity) with the fosmid 74A4, which, in turn, is related to *N. maritimus* SCM1 (Grzymski *et al.*, 2012). In surface seawater of Antarctica during winter seasons, Group I.1a was recognized to comprise 10 to 30% of the bacterioplankton assemblage (DeLong *et al*., 1994; Massana *et* al., 1998; Murray *et al*., 1998) and to play a noteworthy role in carbon fixation coupled to ammonia oxidation (Grzymski *et al.*, 2012; Williams *et al.,* 2012). To date, little is known about the genomic capacities and characteristics of Group I.1a in Antarctic oceans (Béjà *et al.,* 2002; Grzymski *et al.* 2012; Williams *et al.,* 2012), and the present study was conducted to advance our knowledge of the subject.

Motivated by identification of abundant environmental genome fragments affiliated with Group I.1a in the Antarctic winter surface seawater large insert library (Grzymski *et al.*, 2012), our research focused on characterizing the Antarctic Group I.1a genome-encoded capacities, genome organization, and coding gene phylogenomics. To achieve this goal, we analyzed 1,882 ORFs encoded by Antarctic Group I.1a environmental genome fragments by comparing them to well characterized Thaumarchaeota Group I.1a genomes. Our results suggested that the Antarctic Group I.1a

environmental genome fragments represent a very low diversity Group I.1a not yet represented by any enriched or isolated Thaumarchaeota organism. Phylogeny analyses indicated that Antarctic Group I.1a clustered separately from the genome references used and pairwise distance between homologous proteins indicated a closer affiliation with the Arctic isolate "*Ca.* Nitrosopumilus sedimentis" AR2 than the other thaumarchaeal genomes investigated. Therefore, these findings support the effort to enrich and isolate a Thaumarchaeota Group I.1a strain from Antarctic oceans.

## 4.2. Materials and methods

### 4.2.1. Antarctic Group I.1a data set, assemblage, and recruitment

The data used in the present investigation was previously generated by the GOLD Project ID GM00006 and is available at the Integrated Microbial Genomes and Metagenomes, IMG Submission numbers 1212 and 1775 (IMG-M)/ DOE Joint Genome Institute (JGI; https://img.jgi.doe.gov/). The surface seawater samples were collected at Arthur Harbor, Anvers Island, west of the Antarctica Peninsula, during the austral winter of 2002. Methods for sample collection, nucleic acid preparation, and production of the metagenomic data are detailed in Grzymski *et al*. (2012). The Antarctic Group I.1a environmental genome fragments were selected from the metagenomic data set based on distribution of best BLAST hits of protein-coding genes. All fosmids with archaeal ORFs were mapped to the *N. maritimus* SCM1 genome and a total of 1,533,317 bp from 44

fosmids with some affiliation to Thaumarchaeota Group I.1a sequences were selected from the metagenomic data set for analysis.

The environmental genome fragments were initially assembled using Sequencher$^{TM}$4.7 software. In the first assembly step, contigs from the same fosmid were assembled with a minimum overlap of 90 bases and 95% nucleotide match. In the second step, the fosmids were assembled in scaffolds using the same set of parameters. All assembled scaffolds were manually verified and confirmed by checking the nucleotide alignments and gene content of sequences in the alignment.

Recruitment (with an overlap coverage of ≥85%) of nucleotide sequences from the Antarctic Group I.1a scaffolds was performed with the whole-genome sequences of *N. maritimus* SCM1 (NC_010085) using the program wgVISTA (Couronne *et al*., 2003; Frazer *et al*., 2004). The program runs a combination of global and local methods (glocal) alignment algorithm - Shuffle-LAGAN - that detects rearrangements such as insertions, deletions, transpositions, inversions, duplications, and point mutations in a global alignment framework (Brudno *et al.*, 2003).

### 4.2.2. Gene content within Antarctic Group I.1a genome fragments and comparative genomic analyses

Gene annotation and cluster of orthologous groups (COG) distribution from ORFs encoded within the Antarctic Group I.1a fosmids were assigned by and analyzed in the IMG-M/JGI system (Markowitz *et al.*, 2008; 2011). In parallel with the COG analyses,

the Artemis comparison tool (Carver *et al.*, 2005) was used to analyze gene organization of the Antarctic scaffolds and compare them with the gene organization of the genome references *N. maritimus* SCM1 (Könneke *et al.*, 2005), "*Ca.* Nitrosoarchaeon limnia" SFB1 (Blainey *et al*., 2011), "*Ca.* Nitrosopumilus koreensis" AR1 (Park *et al.*, 2012a), "*Ca.* Nitrosopumilus sedimentis" AR2 (Park *et al.*, 2012a), and *Cenarchaeum symbiosum* A (Preston *et al.,* 1996; Hallam *et al*., 2006). Genomic features of data sets used are listed in Table 4.1. The basic local alignment search tool (BLAST; Altschul *et al*., 1990) of ORFs (nucleotides and amino acid sequences) was used to account for gene homology among the Antarctic Group I.1a scaffolds and the thaumarchaeal genomes. Average nucleotide identity (ANI), which estimates the average nucleotide identity using both best hits (one-way ANI) and reciprocal best hits (two-way ANI) between two genomic data sets, was calculated as the mean identity of all BLASTN matches (nucleotide-based) that showed more than 30% overall sequence identity (recalculated to an identity along the entire sequence) over a matching region of at least 70% of their length (Goris *et al.*, 2007). The ANI was calculated using the script generated by Goris *et al.* (2007).

**Table 4.1.** Features of the Antarctic Group I.1a fosmids environmental genome fragments and thaumarchaeal genomes used in the comparative genomic analyses.

| | Antarctic Group I.1a fosmids | *N. maritimus* SCM1 | Candidatus *N. limnia* SFB1 | Candidatus *N. koreensis* AR1 | Candidatus *N. sedimentis* AR2 | *C. symbiosum* A |
|---|---|---|---|---|---|---|
| **Acronym** | AF | SCM1 | SFB1 | AR1 | AR2 | CS |
| **Draft/ Finished** | D | F | D | F | F | F |
| **Source** | Antarctic ocean | Salt-water aquarium | Low-salinity sediment SF Bay | Arctic marine sediment | Arctic marine sediment | Sponge symbiont |
| **Size (bp)** | 926,568 | 1,645,259 | 1,769,573 | 1,639,964 | 1,690,905 | 2,045,086 |
| **Avg % GC** | 31 | 34 | 32 | 34 | 34 | 58 |
| **No.of ORFs** | 1,882 | 1,797 | 2,038 | 1,890 | 1,974 | 2,017 |
| **Total % hypothetical** | 27 | 40 | 45 | 47 | 45 | 51 |
| **Avg ORF length (nt)** | 675 | 825 | 696 | 747 | 740 | 924 |
| **16S-23S rRNAs** | 0 | 1 | 1 | 1 | 1 | 1 |
| **5S rRNAs** | 2 | 1 | 1 | 1 | 1 | 1 |
| **tRNAs** | 27 | 44 | 45 | 39 | 37 | 45 |
| **Other RNAs** | 0 | 2 | 2 | 0 | 0 | 1 |
| **GenBak Accession** | - | NC_010085.1 | NZ_CM001158.1 | NC_018655.1 | CP003843.1 | NC_014820.1 |
| **Reference** | This study | Walker et al., 2010 | Blainey et al., 2011 | Park et al., 2012 | Park et al., 2012 | Hallam et al., 2006 |

### 4.2.3. Protein homology analysis

Protein homology was conducted by BLAST (Altschul *et al.*, 1990). Evolutionary distances were employed to determine the relationship between homologous proteins using MEGA6 software (Tamura *et al.*, 2011). Protein distance matrices and phylogenetic trees were calculated using the Jones-Taylor-Thornton model with bootstrap analysis of 1000 repetitions. The genomes of the hyperthermophile Crenarchaeota *Sulfolobus tokodaii* strain 7 (NC_003106.2) and the mesophile Euryarchaeota *Methanococcus maripaludis* C7 (CP000745.1) were selected as out groups for homology analysis because of their 33% GC content (the same GC content present by the thaumarchaeal genomes, except *C. symbiosum* A). In addition to the protein homology analyses based in the amino acid identity, we also analyzed the percentage of nucleotide identity between proteins and the amino acid composition of the proteins.

## 4.3. Results and discussion

This study classified a set (44 fosmids) of Thaumarchaeota Group I.1a large insert environmental genome fragments from surface seawater collected from the Antarctic Peninsula during winter. We focused on the description of genomic organization, gene content, and gene homology to phylogenetically categorize the Antarctic Group I.1a by comparing them to the thaumarchaeal reference genomes. The approach involved (i) recruitment and analyses of rearrangement events, (ii) analysis of COG distribution and unique ORFs identified within the Antarctic Group I.1a, and (iii) analyses of protein homology.

### 4.3.1. Recruitment of Antarctic Group I.1a genome fragments to SCM1 genome

Recruitment analysis from nucleotide data showed that approximately 56% of the SCM1 genome was covered by the Antarctic Group I.1a fosmid DNA at a low level of coverage (Figure 4.1). By assembling the 44 fosmids, we were able to build 33 scaffolds with lengths ranging from 13,131 to 113,139 bp (recruitment plot shown in Figure 4.1; fosmids used to build the scaffolds are listed in Table S4.1). Each scaffold matched the SCM1 genome in overlapping size with a minimum correspondence of 85% (where 100% is the total size of the fosmid). All 33 scaffolds presented similar percentages of GC content (28% to 35%) when compared to the thaumarchaeal genome references listed in Table 4.1 (31–34%), but each was significantly different from the 58% GC content reported for *C. symbiosum* (Hallam *et al.*, 2006). The 33 scaffolds aligned to 11 distinct

regions of the SCM1 genome (indicated by letters A to K in Figure 4.1). Overlap between the scaffolds was identified in nine of the 11 regions. Regions D and E were covered by only one scaffold each (Figure 4.1).

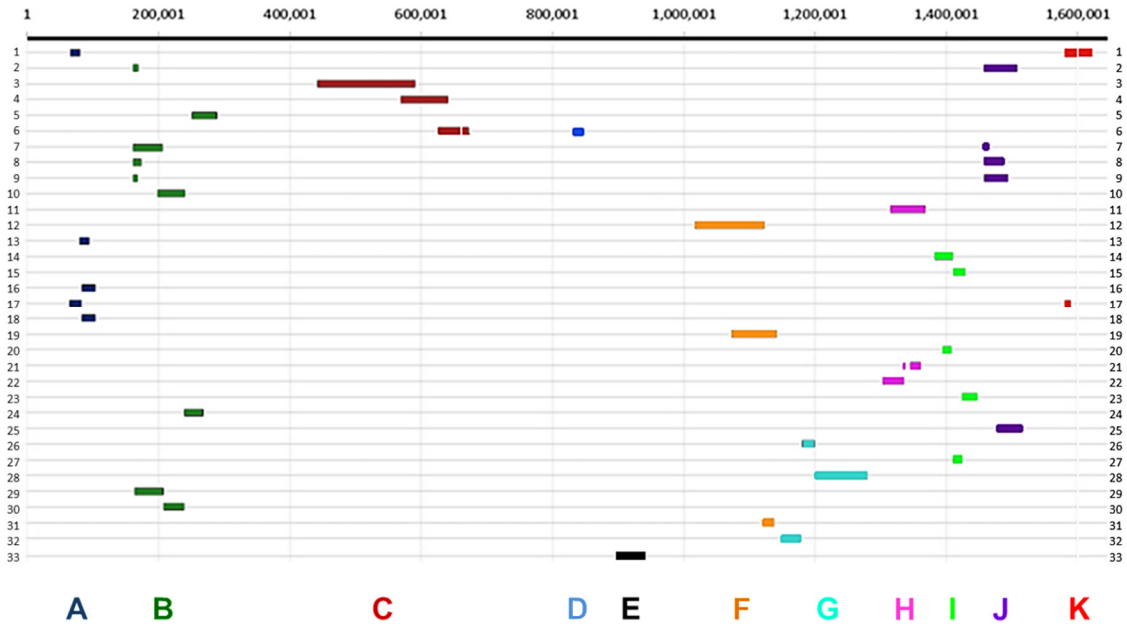



**Figure 4.1.** Antarctic Group I.1a recruitment plot using *N. maritimus* SCM1 as the genome reference. The black line at the top of the figure represents the reference genome of *N. maritimus* SCM1 from nucleotide 1 to 1,645,259. Each line below the reference genome represents the recruitment position of each of the scaffolds built from the fosmids. Different colors indicate distinct overlapping regions which were named regions A to K.

The recruitment exercise between the environmental genome fragments and the SCM1 genome identified significant genomic rearrangements within the Antarctic Group I.1a environmental genome fragments. Eight of the 33 scaffolds aligned with two different regions of the SCM1 genome (Figure 4.1; scaffold numbers 1, 2, 6, 7, 8, 9, 17, and 21). Scaffold 1 aligned partially to the SCM1 genome region from position 64,870 to

position 80,130 and partially from position 1,580,684 to position 1,622,447 (Figure 4.1). Synteny analysis, based on nucleotide sequences, indicated the occurrence of an inverted translocation event in scaffold 1 (Figure 4.2A); scaffolds 2, 8, and 9 (Figure 4.2B); and scaffold 7 (Figure 4.2C) when compared to the SCM1 genome.

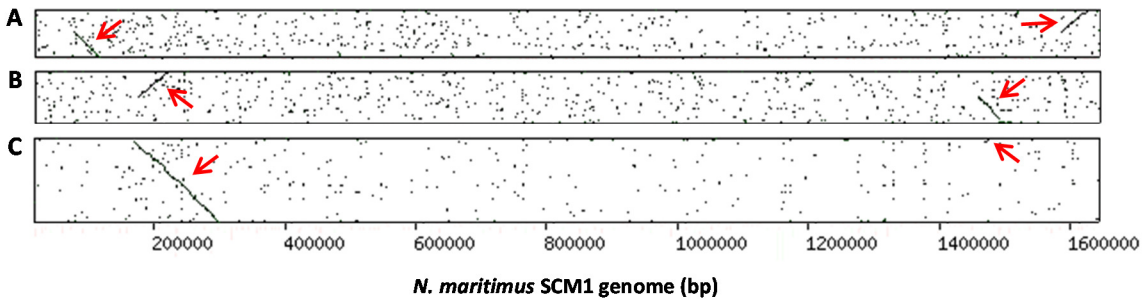



**Figure 4.2.** Inverted translocation events observed in the Antarctic Group I.1a scaffolds. A, B, and C show the synteny plots comparing the SCM1 genome (bp) with scaffold 1 (in A); scaffold 2, 8, and 9 (in B); and scaffold 7 (in C). Dot matrices were built using a minimum and maximum window of 20 and 40,000, respectively.

### 4.3.2. Gene content and comparative genomic analyses

A total of 1,882 ORFs were identified within the 926,568 bp of the Antarctic Group I.1a fosmids. By clustering the ORFs at 95% amino acid identity, we found 1,172 unique ORFs. Hypothetical proteins corresponded to 30% (357 ORFs) of the ORFs, 69% (813 ORFs) were assigned to COGs, and 3% (34 ORFs) presented a conserved domain but were not assigned to COGs. Two 5S rRNA genes, 27 tRNA genes, and no 16S-23S rRNA genes were detected. For the comparative genomic analyses and a description of the gene content within Antarctic Group I.1a genome fragments, we described (i) COG distribution, (ii) breaks in synteny observed among genomes, and (iii) unique ORFs identified within Antarctic Group I.1a.

**(i) COG distribution.** Comparison of COG categories between Antarctic Group I.1a and genome references indicated that the environmental COGs followed the standard distribution of COGs within thaumarchaeal genomes (Figure 4.3). The category S (function unknown) represented almost 10% of the COGs identified (9.10%). By adding the hypothetical proteins with the proteins from COG category S, we identified that 49% of the Antarctic Group I.1a ORFs could not be assigned to a function.

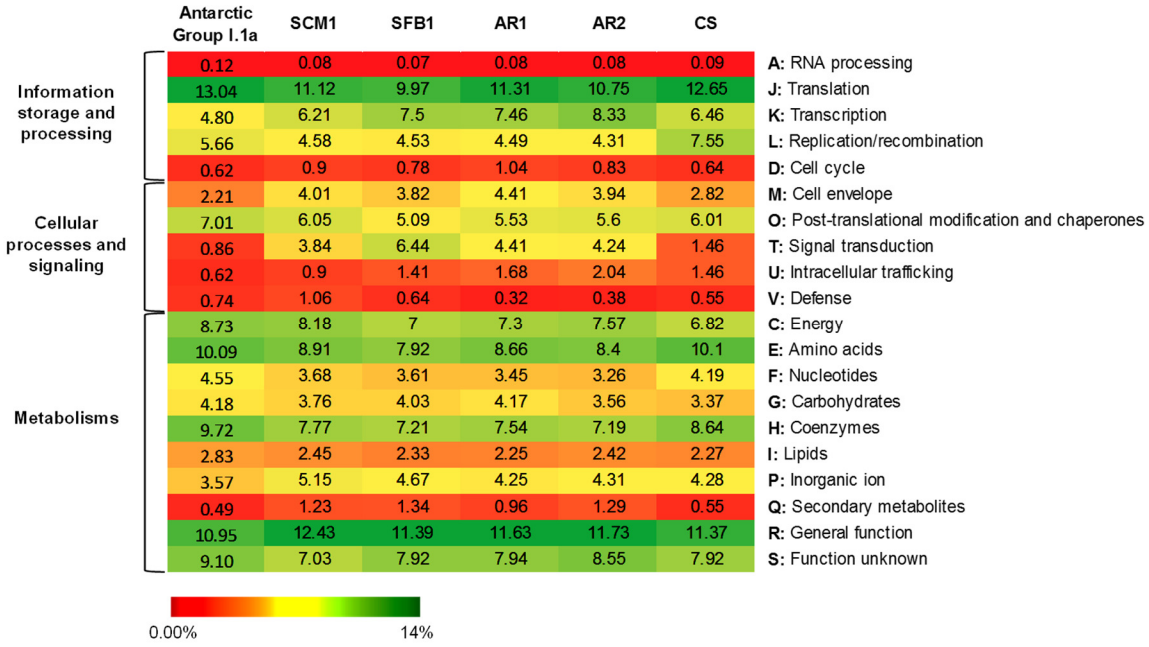



**Figure 4.3.** Relative abundance distributions of Clusters of Orthologous Groups of Proteins (COG) categories within Antarctic Group I.1a, and thaumarchaeal genome references. SCM1: *N. maritimus* SCM1; SFB1: "*Ca.* Nitrosoarchaeon limnia" SFB1; AR1: "*Ca.* Nitrosopumilus koreensis" AR1; AR2: "*Ca.* Nitrosopumilus sedimentis" AR2; CS: *C. symbiosum* A.

Among the 51% of the ORFs characterized, 28 ORFs associated with the pathway for carbon assimilation 3-hydroxypropionate/4-hydroxybutyrate (3-HP/4-HB; Berg *et al.*, 2007) were identified (Table 4.2). Genes associated with this pathway were previously observe by Grzymski *et al.* (2012). With the exception of one gene (succinate semialdehyde reductase), all of the genes implicated in the 3-HP/4-HB were detected and showed the highest similarity with the genes of "*Ca.* Nitrosopumilus sedimentis" AR2. Some of the genes in the pathway, including the key enzymes of the pathway (4-hydroxybutyryl-CoA dehydratase), presented more than one homolog within the Antarctic Group I.1a genome sequences. Homologous genes presented 93.7 to 99.3% nucleotide identity and 99% amino acid identity indicating that little genetic variability was identified among these predicted proteins within the Antarctic Group I.1a.

**Table 4.2.** Genes involved in the pathway for carbon assimilation 3-hydroxypropionate/4-hydroxybutyrate identified within the Antarctic Group I.1a (AntGroupI.1a) genome fragments, the thaumarchaeal genomes, one crenarchaeotal genome (st7), and one euryarchaeotal genome (C7). SCM1: *N. maritimus* SCM1; SFB1: "*Ca.* Nitrosoarchaeon limnia" SFB1; AR1: "*Ca.* N. koreensis" AR1; AR2: "*Ca.* N. sedimentis" AR2; CS: *C. symbiosum* A; St7: S*ulfolobus tokodaii* strain 7; C7: *Methanococcus maripaludis* C7.

| Gene product | COG | Locus Tag in genomes | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | SCM1 | AntGroupI.1a | SFB1 | AR1 | AR2 | CS | St7 | C7 |
| acetyl-CoA carboxylase | COG4799 | Nmar_0272 | AWFS_00004390 | Nlim_1195 | NKOR_01205 | NSED_01260 | CENSYa_1660 | ST0591 | 0 |
| acetyl-CoA carboxylase | COG4770_1 | Nmar_0273 | AWFS_00004400 | Nlim_1194 | NKOR_01210 | NSED_01265 | CENSYa_1661 | ST0593 | MmarC7_1342 |
| acetyl-CoA carboxylase | COG4770_2 | Nmar_0274 | ANTWFO_05620<br>AWFS_00004410 | Nlim_1193 | NKOR_01215 | NSED_01270 | CENSYa_1662 | ST0592 | 0 |
| malonate semialdehyde reductase | COG1028 | Nmar_1043 | AWFS_00005600 | Nlim_0722 | NKOR_05815 | NSED_06130 | CENSYa_1197 | 0 | 0 |
| 3-hydroxypropionyl-CoA synthetase | COG1042_1 | Nmar_1309 | AWFS_00043060 | Nlim_1038 | NKOR_07375 | NSED_07250 | CENSYa_0167 | 0 | 0 |
| 4-hydroxybutyryl-CoA synthetase | COG1042_2 | Nmar_0206 | ANTWFO_18720<br>AWFS_00003220 | Nlim_1257 | 0 | NSED_00830 | CENSYa_0021 | ST1537 | MmarC7_1256 |
| acryloyl-CoA reductase | COG0604 | Nmar_1622 | AWFS_00018750<br>ANTWFO_03300<br>ANTWFO_18160<br>ANTWFO_17400 | Nlim_1382 | NKOR_08825 | NSED_08780 | CENSYa_1751 | ST2605 | 0 |
| acryloyl-CoA reductase | COG1064 | Nmar_0523 | ANTWFO_21160<br>ANTWFO_16640 | Nlim_0075 | NKOR_02700 | NSED_02630 | CENSYa_1407 | ST0038n | 0 |
| methylmalonyl-CoA mutase, large subunit | COG1884 | Nmar_0954 | ANTWFO_19160 | Nlim_0652 | NKOR_05405 | NSED_05460 | CENSYa_0476 | ST0552 | 0 |
| methylmalonyl-CoA mutase | COG2185 | Nmar_0958 | ANTWFO_19230 | Nlim_0656 | NKOR_05425 | NSED_05515 | CENSYa_0482 | ST2096 | 0 |
| methylmalonyl-CoA epimerase | COG3185 | Nmar_0953 | ANTWFO_19150 | Nlim_0650 | NKOR_05400 | NSED_05435 | CENSYa_0475 | 0 | 0 |
| succinyl-CoA reductase | COG1012 | Nmar_1608 | ANTWFO_03950 | Nlim_1367 | NKOR_08755 | NSED_08705 | 0 | ST0064 | MmarC7_0731 |
| succinate semialdehyde reductase | COG1454 | Nmar_1110 | 0 | Nlim_0812 | NKOR_06610 | NSED_06440 | CENSYa_0549 | | |
| 4-hydroxybutyryl-CoA dehydratase | COG2368 | Nmar_0207 | AWFS_00003270<br>ANTWFO_18730 | Nlim_1256 | NKOR_00855 | NSED_00835 | CENSYa_0022 | ST1659 | 0 |
| crotonyl-CoA hydratase | COG1024 | Nmar_1308 | AWFS_00001800<br>AWFS_00043080<br>AWFS_00042300<br>AWFS_00042290 | Nlim_1037 | NKOR_07370 | NSED_07245 | CENSYa_0166 | ST0048 | 0 |
| 3-hydroxyacyl-CoA dehydrogenase | COG1250 | Nmar_1028 | AWFS_00005780 | Nlim_0708 | NKOR_05740 | NSED_06060 | CENSYa_0413 | ST0069 | 0 |
| acetoacetyl-CoA beta-ketothiolase | COG0183 | Nmar_1631 | AWFS_00018890<br>ANTWFO_17510<br>ANTWFO_02860 | Nlim_1392 | NKOR_08875 | NSED_08890 | CENSYa_1780 | ST0079 | 0 |

The Antarctic Group I.1a genome fragments also harbored ORF coding for the three ammonia monooxygenases (*amo*A, *amo*B, *amo*C) plus "amoX" hypothetical protein, but no ORF coding for ammonia transporters was found (Table 4.3).

**Table 4.3.** Genes involved in ammonia oxidation pathway identified within the Antarctic Group I.1a (AntGroupI.1a) genome fragments (fosmid FIUX1083) and the thaumarchaeal genomes. SCM1: *N. maritimus* SCM1; SFB1: "*Ca.* Nitrosoarchaeon limnia" SFB1; AR1: "*Ca.* N. koreensis" AR1; AR2: "*Ca.* N. sedimentis" AR2; CS: *C. symbiosum* A.

| Gene product | Locus Tag in genomes | | | | | |
|---|---|---|---|---|---|---|
| | SCM1 | AntGroupI.1a | SFB1 | AR1 | AR2 | CS |
| ammonium transporter | Nmar_0588 | 0 | Nlim_1564 | NKOR_03065 | NSED_09440 | 0 |
| ammonium transporter | Nmar_1698 | 0 | Nlim_1421 | NKOR_09015 | NSED_09040 | CENSYa_1453<br>CENSYa_0526 |
| ammonia monooxygenase subunit A | Nmar_1500 | AWFS_00011010 | Nlim_1890 | NKOR_08170 | NSED_08255 | CENSYa_0402 |
| hypothetical protein | Nmar_1501 | AWFS_00011000 | Nlim_1891 | NKOR_08260 | NSED_08175 | CENSYa_0401 |
| ammonia monooxygenase subunit C | Nmar_1502 | AWFS_00010990 | Nlim_1892 | NKOR_08180 | NSED_08265 | CENSYa_0399<br>CENSYa_0670 |
| ammonia monooxygenase subunit B | Nmar_1503 | AWFS_00010970 | Nlim_1893 | NKOR_08185 | NSED_08270 | CENSYa_0394 |

**(ii) Breaks in synteny observed within Antarctic Group I.1a.** The comparison of gene organization between Antarctic Group I.1a genome fragments and *N. maritimus* SCM1 genome revealed a high percentage of ORF deletion within the Antarctic genome fragments. Among the 1,013 ORFs aligned between the Antarctic Group I.1a genome fragments and the SCM1 genome, we observed 272 events of ORF deletion in the Antarctic genome fragments (26.85% of the Group I.1a ORFs), 23 events of insertion of an ORF in the Antarctic genome fragments (2.27% of the Group I.1a ORFs), and 16 mismatches of ORFs between genome alignments (1.57% of the Group I.1a ORFs).

The 23 ORFs identified in the Antarctic Group I.1a genome fragments without homologous genes in the same *loci* in the genome of *N. maritimus* SCM1 are shown in Table 4.4. Interestingly, nine of these 23 ORFs were not detected within SCM1 ORFs, and two of the 23 ORFs were phylogenetically affiliated with homologous genes in bacteria indicating the possibility of horizontal gene transfer (HGT). The remaining 12

ORFs were identified in distinct regions of the SCM1 genome. ORF ANTWFO_18210 showed significant similarity (58%) to an alphaproteobacterium putative sugar O-methyltransferase protein; ANTWFO_02380 showed significant similarity to a Holliday junction resolvase (HJR) protein highly conserved within the *Bacilli* class; and ORFs AWFS_00001580 and AWFS_00001590 hypothetical proteins were shown to be conserved within archaea, but absent in Thaumarchaeota.

The analyses of gene organization within the rearrangement regions of larger fragments suggested that the gene neighborhood contained in Antarctic Group I.1a is conserved but distinct from the gene neighborhood observed within the other thaumarchaeal genomes. For example, the gene organization within 50,000 bp around position 1,456,797 in SCM1 genome (Figure 4.4). We observed that SFB1 and AR1 genomes contain similar gene neighborhoods to those observed in the SCM1 genome through the 50,000 bp fragment (Figure 4.4). However, the genomes of CS, AR2, and four Antarctic Group I.1a scaffolds presented a rearrangement downstream of the region 1,456,797. Additionally, we observed that the overlapped region of the Antarctic Group I.1a scaffolds exhibit a highly conserved gene organization. Strain AR2 contained a conserved gene organization upstream position 1,456,797 but presented a gene neighborhood distinct from the other genomes downstream of position 1,456,797. Not surprisingly, the genome of *C. symbiosum* A (CS) contained a gene neighborhood distinct from the other thaumarchaeal genomes throughout the entire region analyzed, a result expected because of the lack of synteny of CS genome with SCM1 genome (Walker *et al.,* 2010).

**Table 4.4.** Insertions of ORFs observed in the Antarctic Group I.1a genome fragments. Percentage of identity was based in amino acid sequences.

| Antarctic Group I.1a Locus Tag | Length (aa) | Product name | COG ID or Pfam Domain | COG description | Best Identity hit (%) | e-value | Best Genome hit | Best Gene hit ID in IMG* | Percent Identity | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | SCM1 | SFB1 | AR1 | AR2 | CS |
| ANTWFO_18210 | 347 | Hypothetical protein | - | - | 58.00 | 2.00E-66 | *Alphaproteobacterium* | WP_018036975* | - | - | - | - | - |
| ANTWFO_02380 | 259 | Predicted endonuclease distantly related to archaeal Holliday junction resolvase and Mrr-like restriction enzymes | COG1787 | [V] Defense mechanisms | 64.00 | 2.00E-19 | *Bacillus cereus* | WP_002091772* | - | - | - | - | - |
| ANTWFO_03330 | 535 | Hypothetical protein | - | - | 38.34 | 5.00E-121 | *Ca* . N. sedimentis AR2 | 2518892124 | 36.86 | 39.78 | 36.48 | **38.34** | 31.03 |
| ANTWFO_13470 | 327 | Oxidoreductase | COG0667 | [C] Energy production and conversion | 75.68 | 0.00 | *Ca* . N. sedimentis AR2 | 2518890547 | - | 68.28 | - | **75.68** | 55.62 |
| ANTWFO_18230 | 371 | Hypothetical protein | COG2221 | [C] Energy production and conversion | 60.75 | 4.00E-170 | *Ca* . N. sedimentis AR2 | 2518892112 | 60.00 | 61.52 | 59.94 | **60.75** | 47.37 |
| AWFS_00037190 | 254 | Xaa-Pro aminopeptidase | COG0006 | [E] Amino acid transport and metabolism | 87.40 | 0.00 | *Ca* . N. sedimentis AR2 | 2518892175 | 86.61 | 77.95 | 87.01 | **87.40** | 62.6 |
| ANTWFO_13430 | 316 | Alcohol dehydrogenase | COG1063 | [E] Amino acid transport and metabolism | 81.01 | 0.00 | *Ca* . N. sedimentis AR2 | 2518890532 | 77.22 | 34 | 76.90 | **81.01** | 30.53 |
| AWFS_00037170 | 335 | tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase | COG0482 | [J] Translation, ribosomal structure and biogenesis | 91.29 | 0.00 | *Ca* . N. sedimentis AR2 | 2518892176 | 81.38 | 78.81 | 81.38 | **91.29** | 73.19 |
| ANTWFO_03340 | 99 | Ribonuclease P component 4 | COG2023 | [J] Translation, ribosomal structure and biogenesis | 83.84 | 9.00E-58 | *Ca* . N. sedimentis AR2 | 2518892109 | - | 79.07 | 78.89 | **83.84** | 59.18 |
| ANTWFO_13440 | 381 | Dolichol-phosphate mannosyltransferase | COG0463 | [M] Cell wall/membrane/envelope biogenesis | 74.54 | 0.00 | *Ca* . N. sedimentis AR2 | 2518891124 | 65.16 | 65.61 | 76.64 | **74.54** | 56.61 |
| ANTWFO_01500 | 269 | DSBA oxidoreductase | COG1651 | [O] Posttranslational modification, protein turnover, chaperones | 71.89 | 2.00E-132 | *Ca* . N. sedimentis AR2 | 2518891658 | 65.37 | 55.20 | 64.75 | **71.89** | 32.65 |
| AWFS_00005430 | 374 | Peptidase U62 modulator of DNA gyrase | COG0312 | [R] General function prediction only | 81.33 | 0.00 | *Ca* . N. sedimentis AR2 | 2518891577 | 81.45 | - | 80.59 | **81.33** | - |
| ANTWFO_03350 | 222 | Ribosome biogenesis protein | COG1756 | [S] Function unknown | 79.37 | 2.00E-133 | *Ca* . N. sedimentis AR2 | 2518892111 | 75.68 | 70.72 | 76.13 | **79.37** | 50.91 |
| ANTWFO_13460 | 263 | Hypothetical protein | COG1801 | [S] Function unknown | 79.09 | 5.00E-164 | *Ca* . N. sedimentis AR2 | 2518890534 | - | - | - | **79.09** | - |
| ANTWFO_02530 | 359 | Hypothetical protein | - | - | 59.88 | 6.0E-119 | *Ca* . Nitrosoarchaeon limnia SFB1 | 651418931 | 31.33 | **59.88** | 31.94 | 33.13 | - |
| ANTWFO_02450 | 448 | Deoxyribodipyrimidine photolyase | COG0415 | [L] Replication, recombination and repair | 60.04 | 0.00 | *Ca* . Nitrosoarchaeon limnia SFB1 | 651418014 | - | **60.04** | - | - | - |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AWFS_00040220 | 497 | Hypothetical protein | COG1331 | [O] Posttranslational modification, protein turnover, chaperones | 86.58 | 0.00 | *Ca* . Nitrosopumilus koreensis" AR1 | 2518760151 | 86.58 | 77.12 | **86.58** | 81.36 | 68.16 |
| ANTWFO_00650 | 455 | Zn-dependent hydrolase of the beta-lactamase fold-like protein | COG2220 | [R] General function prediction only | 57.48 | 0.00 | *Ca* . Nitrosopumilus koreensis" AR1 | 2518758870 | - | 30.41 | 57.48 | - | - |
| AWFS_00001580 | 517 | Hypothetical protein | COG0433 | [R] General function prediction only | - | - | *Euryarchaeota* | - | - | - | - | - | - |
| ANTWFO_02470 | 1483 | FG-GAP repeat | PG-GAP | FG-GAP repeat | 65.73 | 2.00E-57 | *N. maritimus* SCM1 | 641315852 | **65.73** | 38.21 | 64.34 | 62.76 | 58.68 |
| AWFS_00005450 | 451 | Carbamoyltransferase | COG2192 | [O] Posttranslational modification, protein turnover, chaperones | 76.33 | 0.00 | *N. maritimus* SCM1 | 641315747 | **76.33** | - | - | - | - |
| AWFS_00005490 | 719 | Lipolytic protein G-D-S-L family | COG1357 | [S] Function unknown | 32.98 | 2.00E-17 | *N. maritimus* SCM1 | 641315746 | **32.98** | 40.23 | - | 30.77 | - |
| AWFS_00001590 | 299 | Hypothetical protein | COG1468 | [L] Replication, recombination and repair | 41.00 | 4.00E-06 | uncultured archaeon | AAU83873* | - | - | - | - | - |

* indicates Genbank (NCBI) gene ID number.

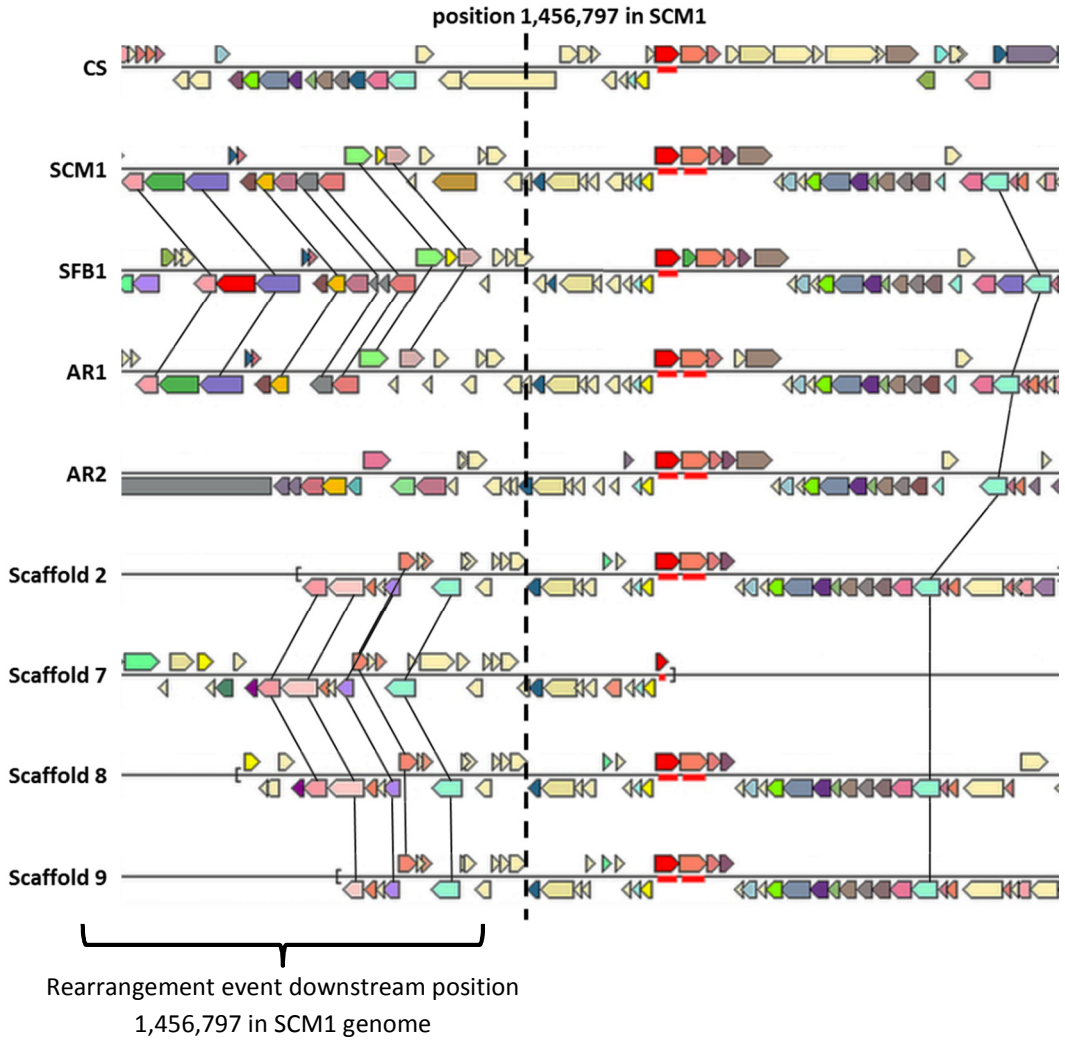



**Figure 4.4.** Genomic organization and evidence of a rearrangement event in the region near position 1,456,797 in the SCM1 genome (identified by the dashed line in the middle of the figure). Size of the region aligned: 50,000 bp. Lines connecting ORFs were added to help in synteny guidance. Genome references: CS: *C. symbiosum* A; SCM1: *N. maritimus* SCM1; SFB1: "*Ca.* Nitrosoarchaeon limnia" SFB1; AR1: "*Ca.* Nitrosopumilus koreensis" AR1; AR2: "*Ca.* Nitrosopumilus sedimentis" AR2; and scaffolds 2, 7, 8, and 9: Antarctic Group I.1a environmental genome sequences. Colors indicate homologous genes. Underlined genes (Nmar_1604 - phosphate uptake regulator and Nmar_1605 - GTP1/OBG protein) were used to anchor the region for analysis.

**(iii) Unique ORFs identified within Antarctic Group I.1a.** We detected 176 ORFs in Antarctic Group I.1a genome fragments that were not identified in the other thaumarchaeal genomes (using a 40% amino acid identity threshold). From these, 19 contained a conserved domain (specific or multi-domain) or were assigned to a protein superfamily (Table 4.5). The remaining unique ORFs were identified as hypothetical proteins with unknown functions.

Three of the unique ORFs identified in three distinct fosmids (ANTWFO_02380, ANTWFO_18360, and ANTWFO_19200 – COG1787) were predicted as Holliday junction resolvase and Mrr-like restriction enzymes. These ORFs were distantly related to the archaeal Holliday Junction resolvases and clustered with bacterial homologous genes indicating the occurrence of HGT in these ORFs (Figure S4.1). The other two ORFs that we suspect originated from HGT (considering their lack of homology to the thaumarchaeal genomes and distinct GC content) were ANTWFO_17560 (COG3398) and ANTWFO_17550. ORF ANTWFO_17560, an uncharacterized protein conserved in *Archaea,* had a GC content of only 22%, which is distinct from the 30–34% in the thaumarchaeal genomes, suggesting HGT. ANTWFO_17550 was identified as BNR/Asp-box repeat, which is a structural motive (Copley and Russell, 2001) and had a GC content of 38%.

**Table 4.5.** Unique ORFs found within Antarctic Group I.1a genome fragments that contained a conserved domain.

| Locus Tag | Product name | Sequence length (aa) | PSSM-ID* | e-value | Conserved domain name |
|---|---|---|---|---|---|
| ANTWFO_17550 | BNR/Asp-box repeat | 462 | 221914 | 1.26E-03 | BNR/Asp-box repeat. |
| ANTWFO_04470 | hypothetical protein | 102 | 246664 | 9.70E-03 | Animal haem peroxidase-like superfamily |
| AWFS_00042200 | hypothetical protein | 28 | 247774 | 8.50E-04 | Arsenical Resistance Operon Repressor |
| AWFS_00043050 | hypothetical protein | 68 | 236548 | 1.43E-20 | beta-D-galactosidase |
| AWFS_00039390 | hypothetical protein | 21 | 242830 | 3.93E-09 | Chloramphenicol acetyltransferase suprefamily |
| AWFS_00037580 | hypothetical protein | 30 | 235205 | 1.67E-03 | Glutamyl-tRNA(Gln) amidotransferase subunit E |
| AWFS_00042500 | hypothetical protein | 26 | 226168 | 3.48E-03 | Mn2+-dependent serine/threonine protein kinase [COG3642-Signal transduction mechanisms] |
| AWFS_00001530 | hypothetical protein | 96 | 185594 | 7.55E-03 | Not assigned to any domain superfamily. |
| ANTWFO_00110 | hypothetical protein | 212 | 222176 | 4.06E-03 | Phage integrase, N-terminal SAM-like domain |
| AWFS_00040150 | hypothetical protein | 141 | 248433 | 1.61E-04 | Putative 2OG-Fe(II) oxygenase superfamily |
| AWFS_00011070 | hypothetical protein | 59 | 216254 | 3.39E-03 | Receptor L domain superfamily |
| AWFS_00011870 | hypothetical protein | 130 | 236185 | 6.65E-03 | Succinate dehydrogenase flavoprotein subunit |
| ANTWFO_01720 | hypothetical protein | 112 | 243034 | 4.95E-03 | Tetratricopeptide repeat domain |
| AWFS_00038820 | Methylase involved in ubiquinone/menaquinone biosynthesis | 298 | 100107 | 3.70E-16 | S-adenosylmethionine-dependent methyltransferases |
| ANTWFO_02380 | Predicted endonuclease distantly related to archaeal Holliday junction resolvase and Mrr-like restriction enzymes | 259 | 218102 | 6.45E-23 | Predicted endonuclease distantly related to archaeal Holliday junction resolvase and Mrr-like restriction enzymes |
| ANTWFO_18360 | Predicted endonuclease distantly related to archaeal Holliday junction resolvase and Mrr-like restriction enzymes | 217 | 218102 | 5.36E-35 | Predicted endonuclease distantly related to archaeal Holliday junction resolvase and Mrr-like restriction enzymes |
| ANTWFO_19200 | Predicted endonuclease distantly related to archaeal Holliday junction resolvase and Mrr-like restriction enzymes | 205 | 218102 | 9.03E-31 | Predicted endonuclease distantly related to archaeal Holliday junction resolvase and Mrr-like restriction enzymes |
| AWFS_00005490 | Uncharacterized low-complexity proteins | 719 | 238141 | 1.12E-09 | SGNH_hydrolase |
| ANTWFO_17560 | Uncharacterized protein conserved in archaea | 131 | 247774 | 3.59E-04 | Arsenical Resistance Operon Repressor |

* PSSM-ID: unique identifier for the position-specific scoring matrix

### 4.3.3. Protein homology analyses

Best homologous hits analysis between the ORFs identified in the Antarctic Group I.1a genome fragments and the archaeal genome references revealed a high number of best hits with SCM1, SFB1, AR1, and AR2 genomes and a low number of best hits with the CS genome and even lower with the Crenarchaeota *S. tokodaii* strain 7 genome and the Euryarchaeota *M. maripaludis* C7 (Figure S4.2). Considering a minimum BLASTp identity of 40%, the Antarctic Group I.1a environmental genome shared 83.5% of its 1,882 ORFs with the AR2 genome, 81% with the AR1 genome, 82.6% with the

SCM1 genome, and 79.4% of its gene content with the SFB1 genome (Figure 4.5). Antarctic Group I.1a shared only 69.6% of its gene content with the *C. symbiosum* genome. The ANI shared among the thaumarchaeal genomes and Antarctic Group I.1a ranged from 80.60% to 82.86% (Figure 4.5). Among the highly conserved genes within the thaumarchaeal genomes, we focused on 42 for homology analysis. The goal was to identify the phylogenetic relationship between Antarctic Group I.1a and the thaumarchaeal genomes using conserved genes among the genomes. We concentrated our homology analysis on genes with phylogenetic and physiological relevance including the following: (i) a concatenated set of ribosomal proteins, (ii) the chaperone encoding operon, (iii) genes associated with the 3HP/4HP pathway, and (iv) the ammonia monooxygenase operon.

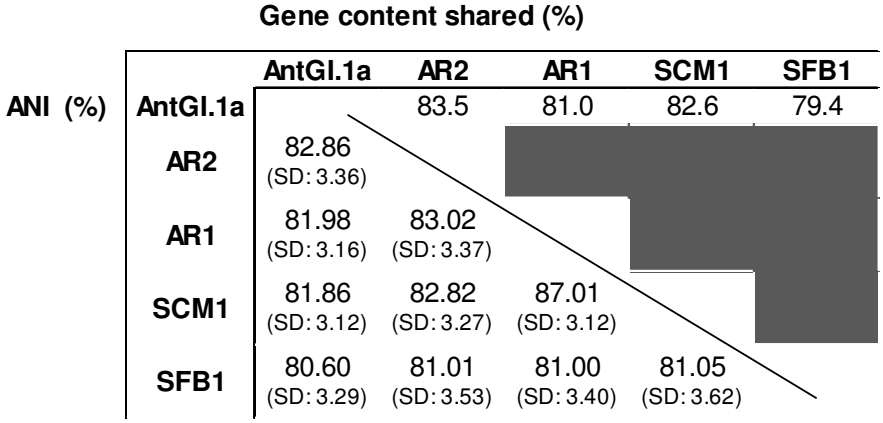



**Figure 4.5.** Average nucleotide identity (ANI) values and percentage of gene content shared among the Antarctic Group I.1a genome fragments (AntGI.1a) and the thaumarchaeal genomes are as follow: AR2: "*Ca.* N. sedimentis" AR2; AR1: "*Ca.* N. koreensis" AR1; SCM1: *N. maritimus* SCM1; SFB1: "*Ca.* Nitrosoarchaeon limnia" SFB1. The values above the diagonal line are the shared gene content and below the diagonal line are the ANI scores. SD is standard deviation. Gene content shared was based on 40% amino acid identity threshold.

**(i) Concatenated set of ribosomal proteins.** We identified 22 ribosomal (R) proteins within the Antarctic Group I.1a genome fragments that were represented by one to seven homologous ORF copies (nucleotide identity higher than 92.8% and amino acid identity higher than 99%; Table S4.2). Not all homologs of the R proteins were identified within the genome references (e.g., R protein S30E). Consequently, for homology analyses, a concatenated alignment using only R proteins shared among all five thaumarchaeal genomes and Antarctic Group I.1a was built. The maximum likelihood phylogenetic tree based on 2,641 amino acid positions from 21 R proteins showed that SCM1 and AR1 presented the strongest association with a pairwise distance of 0.06 and 96% amino acid identity (bootstrap value [BV] of 100%; Figure 4.6A). Antarctic Group I.1a, on the other hand, showed a pairwise distance of 0.24-0.25 with SCM1, AR1, and AR2 and grouped outside of the clade formed by SCM1, AR1, and AR2. Therefore, based on the R proteins' homology, Antarctic Group I.1a presented equal relatedness with the genomes of SMC1, AR1, and AR2 with a percentage of amino acid identity of 86% (BV of 86%).

**(ii) Chaperone operon.** The chaperone operon is composed of the genes *grpE*, h*sp70 (dnaK),* and h*sp40 (dnaJ).* The homology analysis based on 994 amino acid positions of the concatenated proteins indicated that the Antarctic Group I.1a operon is closely affiliated with the operon codified by the AR2 genome with a pairwise distance of 0.24 and BV of 98% (Figure 4.6B). The tree topology for the chaperone operon analysis is similar to the tree topology of the genes associated with the 3HP/4HB pathway and differs from the tree topology observed with the R protein and ammonia monooxygenase operon analyses.

**(iii) Genes associated with the 3HP/4HP pathway.** For the genes associated with the 3HP/4HP pathway, 14 of the proteins shown in Table 4.2 were aligned. The proteins 4-hydroxybutyryl-CoA synthetase, succinyl-CoA reductase, and succinate semialdehyde reductase were not included in the analyses because of the absence of homologs in AR1, Antarctic Group I.1a, and CS, respectively. The phylogenetic tree based on 5,186 amino acid positions revealed a topology distinct from the topology observed in the R protein phylogenetic tree (Figure 4.6C). Antarctic Group I.1a showed a robust association (BV of 100%) with AR2 proteins with a pairwise distance of 0.17 and 91% amino acid identity. Once more, as observed in the R protein analyses, SCM1 and AR1 presented the strongest association with a pairwise distance of 0.1 (BV of 100%), while AR1 and AR2 presented a pairwise distance of 0.18.

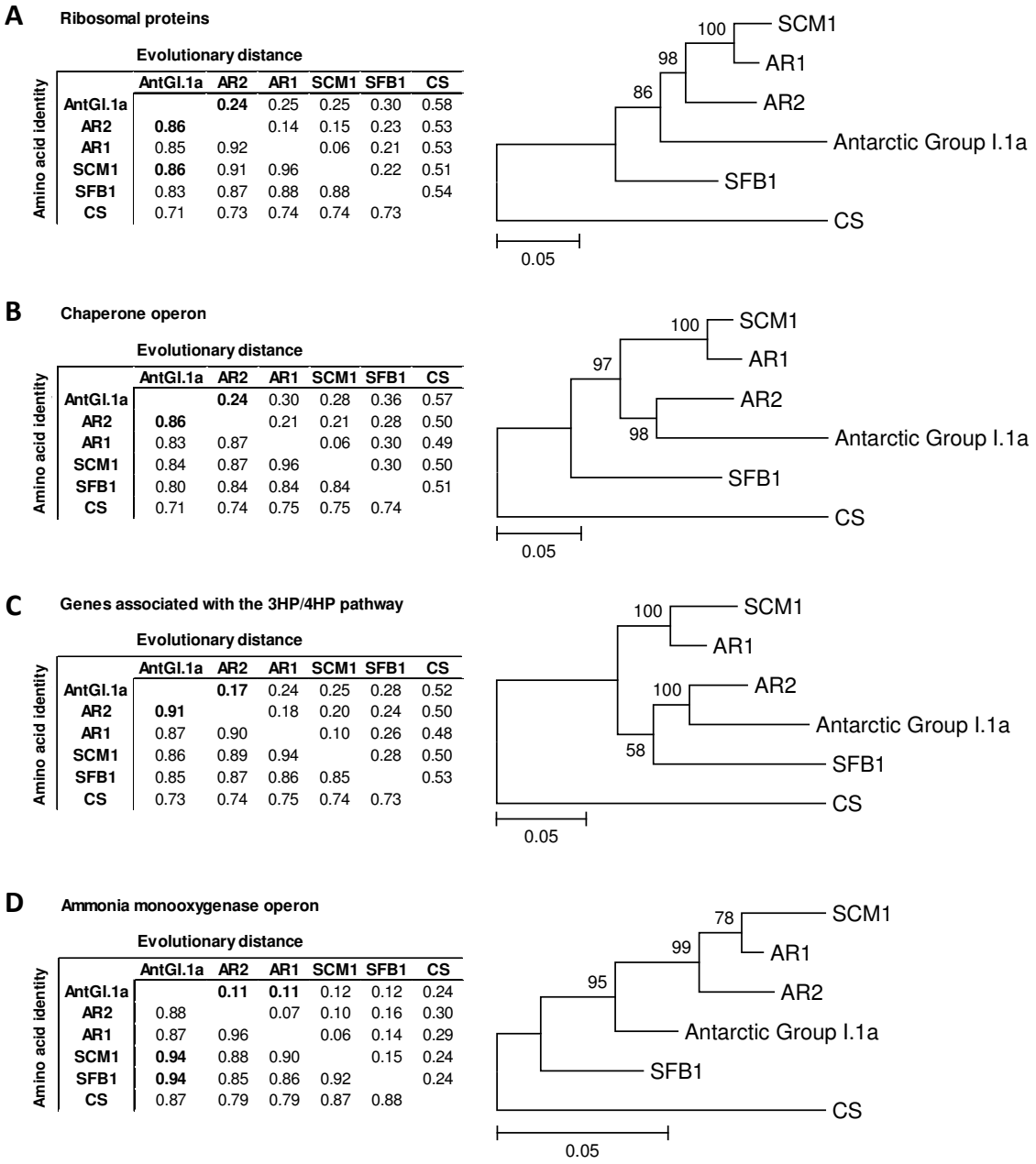



**Figure 4.6.** Amino acid identity values, evolutionary distance, and maximum likelihood phylogenetic tree based on alignment of concatenated genes contained in Antarctic Group I.1a (abbreviated as AntGI.1a) genome fragments and thaumarchaeal genome references. (A) Phylogenetic tree of 21 concatenated R proteins based on the alignment of 2,641 amino acid positions; (B) Phylogenetic tree of the three proteins from the chaperone operon based on the alignment of 994 amino acid positions; (C) Phylogenetic tree of 14 proteins involved in the 3HP/4HB pathway based on the alignment of 5,186 amino acid positions; (D) Phylogenetic tree of the four proteins from the ammonia monooxygenase operon based on the alignment of 714 amino acid positions. SCM1: *N. maritimus* SCM1;

SFB1: "*Ca.* Nitrosoarchaeon limnia" SFB1; AR1: "*Ca.* Nitrosopumilus koreensis" AR1; AR2: "*Ca.* Nitrosopumilus sedimentis" AR2; CS: *C. symbiosum* A.

**(iv) Ammonia monooxygenase operon.** The archaeal operon for ammonia monooxygenase is composed of four conserved genes: *amo*A, *amo*C, "amoX" hypothetical protein, and *amo*B. All four genes were detected within Antarctic Group I.1a genome fragments in a unique fosmid (Table 4.3). The phylogenetic tree based on alignment of 714 amino acid positions showed the same topology previously observed in the R protein analyses where SCM1 clustered with AR1 and a distance of 0.10 (BV of 78%), and AR2 also was closely related to AR1 with a distance of 0.07 (BV of 99%; Figure 4.6D). Antarctic Group I.1a ammonia monooxygenase operon presented the lowest evolutionary distance to AR1 and AR2 (0.11 with BV of 98%), but the highest amino acid identity to SCM1 and SFB1.

The analyses of the evolutionary distance of the four sets of proteins shown in Figure 4.6 indicated that the Antarctic Group I.1a is an outgroup of the cluster formed by SCM1, AR1, and AR2 – which always branched together. In three of the four analyses (R proteins, chaperone operon, and the genes involved with the 3H/4H pathway), the homology analyses showed that Antarctic Group I.1a proteins were most closely affiliated with AR2 proteins.

### 4.3.4. Categorizing the Antarctic Group I.1a assemblage

Comparison of gene organization between Antarctic Group I.1a and the reference genomes revealed a highly conserved set of genes in a rearranged genomic structure. Breaks in synteny within Antarctic Group I.1a and SCM1 genome were observed where one or multiple ORFs organized in tandem were rearranged. The rearrangement observed with highest frequency when comparing Antarctic Group I.1a genome fragments to the SCM1 genome was deletion of one ORF or a fragment with multiple ORFs from the Antarctic Group I.1a genome. Rearrangement events of long fragments between Antarctic Group I.1a and the other genome references also were identified. Our findings support, on a large scale, observations by Béjà *et al.* (2002) and López-García *et al.* (2004) identifying poor conservation of gene order within Antarctic Thaumarchaeota Group I.1a environmental fosmids (fosmids 74A4 and DeepAnt-EC39). These previous findings, however, were limited to regions surrounding the 16S rRNA gene. The absence of synteny within the 16S rRNA neighborhood is not restricted to Thaumarchaeota and also has been observed within Euryarchaeota environmental genome fragments from distinct geographic regions (Brochier-Armanet *et al.*, 2011). We expanded Béjà's finding to 33 environmental genome scaffolds that do not contain the rRNA operon, but we continued to detect a reorganized genomic structure.

The immediate genes surrounding the ammonia monooxygenase operon and chaperone operon showed a different story. We observed that the gene neighborhoods flanking these operons were highly conserved among Thaumarchaeota genomes (except *C. symbiosum*). The ammonia monooxygenase operon identified was located 6,000 to

6,700 bp downstream of the R protein S3Ae in all genome references. The Antarctic fosmid that harbored the ammonia monooxygenase operon was short in length (9,662 bp long) and did not contain the S3Ae gene; thus, we could not infer the same finding to Antarctic Group I.1a. The conserved neighborhood flanking the chaperone operon was identified among all Thaumarchaeota genomes and three fosmids (Figure 4.7). Upstream of the GrpE-Hsp70 (DnaK)-Hsp40 (DnaJ) operon, we identified ORFs for DEAD/DEAH box helicase, OB-fold tRNA, and ABC transporter. Downstream of the gene set, we identified ORFs from Glu-tRNA/Gln aminotransferase and from the $AAA^+$ family ATPse. The two Arctic isolates' gene neighborhood and one of the Antarctic fosmids (ANTWF_4084609) were not as highly conserved as the other genomes. AR1 was missing the ORFs for one of the ABC transporters and contained an insertion for predicted Zn-dependent hydrolases of the beta-lactamase fold (COG2220) between Glu-tRNA/Gln aminotransferase and the $AAA^+$ family ATPse. WAF ANTWF_4084609 contained an insertion of the same ORFs at the same *loci* as the AR1 isolate. AR2 was missing the ORFs for DEAD/DEAH box helicase and one of the Glu-tRNA/Gln aminotransferase subunits (Figure 4.7). The gene neighborhood flanking the chaperone operon included genes associated with stress response. Therefore, it appears likely that the selective advantage of keeping these genes clustered lies in maintaining approximately the same level and temporal pattern of expression of these genes. This could be related to adaptation of Thaumarchaeota to lower temperatures. Chaperones so far were detected only within the phyla Thaumarchaeota and Euryarchaeota (Spang *et al.*, 2010). We looked into the gene neighborhood of the *hsp70 (dnaK)* gene in 74 Euryarchaeota genomes and did not find the conserved neighborhood in the

Euryarchaeota sequences. Additionally, to support the finding in conservation of this genomic region within the Thaumarchaeota phylum, the region around the *hsp70 (dnaK)* gene from an environmental genome fragment from the North Pacific Sub-Tropical Gyre also was highly conserved (Swan *et al.,* 2011). Therefore, we identified within our analyses a unique gene organization surrounding the Thaumarchaeota's ammonia monooxygenase operon and chaperone operon. These findings can now be integrated to the list of conserved features found among the Thaumarchaeota phylum (Brochier-Armanet *et al.*, 2012; Spang *et al.*, 2010).

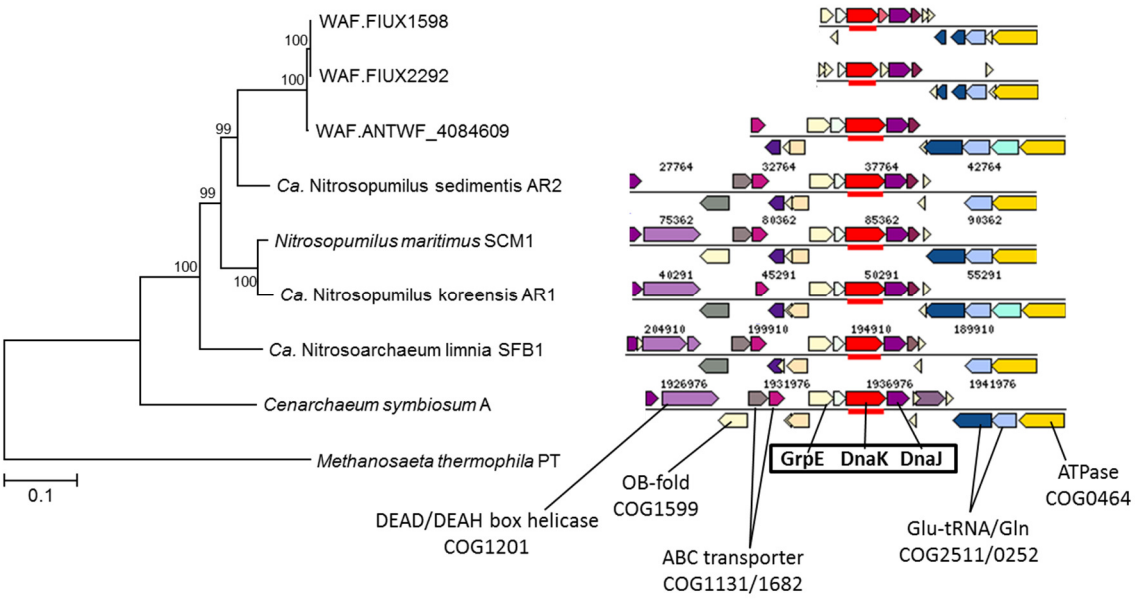



**Figure 4.7.** The GrpE/Hsp70 (DnaK)/Hsp40 (DnaJ) neighborhood and phylogenetic affiliation. (A) Maximum likelihood phylogenetic tree of 994 amino acid positions of concatenated GrpE (COG0576) - Hsp70 (DnaK; COG0443) - Hsp40 (DnaJ; COG0484). Numbers at nodes indicate bootstrap values (%) from 100 repetitions. Identical colors indicate homologous genes in the corresponding aligned genomic fragments. External group: euryarchaeota *Methanosaeta thermophila* PT (NC_008553).

Recombination is a fundamental process that occurs in DNA resulting in the repair of damage (mutation) and production of combinatorial variations that provide the substrate for adaptation to a condition or environment (Aravind *et al.*, 2000). We found 17 ORFs associated with recombination activity within Antarctic Group I.1a. We identified three recombinases (COG0468), that are involved in homologous strand exchange; 11 helicases (COG0513, COG0553, COG1204, and COG1205), that promote strand migration within the DNA; and three predicted Holliday junction resolvase and Mrr-like restriction enzymes (COG1787/COG1591). Holliday junction resolvase promotes recombination by forming a mobile junction structure between four DNA strands (Ruvabc *et al.*, 1999). Therefore, we hypothesize that the high evidence of these DNA structural modifiers (recombinases, helicases, and Holliday junction resolvase) is associated with the high level of recombination observed within Antarctic Group 1.1a genome fragments.

The predicted Holliday junction resolvase and Mrr-like restriction enzymes identified in the Antarctic fosmids contained the three domains which identify the protein family (Aravind *et al.*, 2000). These three proteins were recognized, however, as being phylogenetically associated with bacterial homologs suggesting acquisition by HGT (Philippe and Douady, 2003). Acquisition of exogenous genes by HGT can allow the recipient species to adapt more rapidly to novel environments (Brochier-Armanet *et al.*, 2011). High frequency of genes acquired by HGT in the Antarctic fosmids was previously observed by Béjà *et al.* (2002) and López-García *et al.* (2004). López-García and coauthors suggested that events of HGT might have been important for the

adaptation and evolutionary history of the organisms related to the fosmids 74A4 and DeepAnt-EC39 (López-García *et al.,* 2004). In our set of environmental DNA, we identified only a few of the ORFs affiliated with other archaeal phyla and to Bacteria: three hypothetical proteins and the predicted Holliday junction resolvases (Table 4.4). This difference in the number of possible genes acquired by HGT between our findings and the previous findings might result from the availability of more Thaumarchaeota genome sequences at present. For example, our set of fosmids did not overlap with the SCM1 genome region in which fosmid 74A7 (Béjà *et al.*, 2002) overlapped. However, we found 10 homologous genes (with average nucleotide identity of 89.26%) shared between our fosmids and fosmid 74A4 (Table S4.3). From these, five were previously suggested to be acquired by HGT: cysteinyl-tRNA synthetase; DNAJ molecular chaperone; hypothetical protein (AAK96104); and two double-stranded, beta-helix fold enzymes (Béjà *et al.*, 2002). With the genomes available now, we were able to determine that these genes are highly conserved within most of the Thaumarchaeota genomes.

Our findings based on the gene homology and ANI results indicated that the Antarctic Group I.1a represents a very low diversity group of Thaumarchaeota Group I.1a not yet exemplified by any enriched or isolated Thaumarchaeota organism that had been genomically characterized. The ANI index from the comparison of the Antarctic Group I.1a and the thaumarchaeal genomes ranged from 79.4 to 83.5% identity. According to the genomic demarcation of species by Konstantinidis *et al.* (2006), that proposed 95% ANI as the threshold for species definition, we can infer that Antarctic Group I.1a represents a novel species of the genus "*Ca.* Nitrosopumilus." The pairwise distance and

amino acid and nucleotide identity of R proteins support the concept that the Antarctic assemblage is more closely related to "*Ca.* Nitrosopumilus sedimentis" AR2, a strain isolated from Arctic marine sediment (Park *et al.*, 2012b). It was not possible to infer whether our fosmids represent the same organisms represented by fosmid 74A4 because the homologs shared cannot be used as molecular markers for phylogeny resolution. We observed, however, a high percentage of nucleotide identity between the 10 shared genes ranging from 84.60% to 93.20% (Table S4.3). The R proteins have been shown to be an excellent molecular marker to resolve complicated nodes (Gribaldo & Brochier-Armanet, 2006).

More conserved than R proteins, the ammonia monooxygenase operon revealed the lowest pairwise distance values between the genomes analyzed. The gene *amo*A has been used as a molecular marker for bacteria for a long time (Purkhold *et al.*, 2000) and is now being used as a molecular marker in the study of archaea assemblages (Labrenz *et al.*, 2010; Bartossek *et al.*, 2012). The phylogenetic trees of the chaperone operon and the genes associated with the 3HP/4HB pathway presented the same topology and supported the finding that Antarctic Group I.1a is more closely affiliated with "*Ca.* Nitrosopumilus sedimentis" AR2. The major observation from the phylogenetic affiliation of the concatenated proteins analyzed was that even Antarctic Group I.1a was more closely affiliated with the proteins of AR2 (in three of the four analyses); Antarctic Group I.1a always branched as an external group from the cluster formed by SCM1, AR1, and AR2. These results indicated that phylogenetic affiliation within the Thaumarchaeota strains

SCM1, AR1, and AR2 are higher that the affiliation of Antarctic Group I.1a with these three strains.

Antarctic Group I.1a contained almost all genes associated with the 3HP/4HB pathway, indicating the potential for $CO_2$ fixation (Berg *et al.*, 2007). It also contained a nearly complete set of genes for the TCA cycle (possible reductive cycle) indicating mixotrophic metabolisms – likely *N. maritimus* SCM1 (Walker *et al.*, 2010). Investigating the genes that constitute other metabolic pathways, we observed that most of the pathways were incomplete. In the oxidative phosphorylation pathway, for example, Antarctic Group I.1a contained all the genes involved in formation of complexes I, II, II, and IV, plus NADH; ubiquinone reductase, succinate dehydrogenase, ubiquinol-cytochrome-c, and cytochrome-c oxidase, but did not contain the complete set of genes to build Complex 5 (it contained just one gene out of eight). These observations confirmed that we recovered just a fraction of the genomic information expected to be harbored by an Antarctic Group I.1a cell.

The seawater temperature from which the Antarctic Group I.1a genome fragments originated was -1.73 $^{o}$C (Grzymski *et al.*, 2012). Organisms that thrive in cold environments develop strategies to overcome challenges dictated by low temperature. One of these strategies is a change in amino acid content in proteins allowing the organism to achieve better flexibility to counterbalance the decrease in affinity for the substrate (Chiuri *et al.*, 2009). Significant changes in amino acid composition observed in cold-adapted proteins are as follows: (1) substitution of arginine-to-lysine (which lowers the Arg/[Arg+Lys] ratio); (2) reduction in salt-bridge-forming residues such as arginine

(Arg), glutamic acid (Glu), and aspartic acid (Asp); and (3) reduction in proline (Pro) residues (Grzymski *et al.*, 2006; Srinivas and Ray, 2006). We analyzed amino acid distribution within the proteins in our data set looking for these aspects of adaptation to cold temperatures. We first analyzed all the proteins from the Thaumarchaeota genomes including proteins from the hyperthermophile Crenarchaeota S*ulfolobus tokodaii* strain 7 (33% GC) and the mesophile Euryarchaeota *Methanococcus maripaludis* C7 (33% GC). By comparing amino acid composition from Antarctic Group I.1a proteins to the other Thaumarchaeota proteins (not considering CS because of distinct results), we did not identify any statistically significant distinction between Antarctic Group I.1a and the other thaumarchaeal genomes (Figure S4.3). We then looked into amino acid composition of the proteins used in our homology analysis (R proteins, genes associated with the 3HP/4HB pathway, the ammonia monooxygenase operon, and the chaperone operon), and no statistically significant distinction was observed either. Therefore, according to our analyses, no aspects of adaptation to cold temperatures were identified within the Antarctic Group I.1a amino acid composition compared to the reference genomes.

In conclusion, our comparative genomic analyses revealed a high percentage of homologous genes shared between Antarctic Group I.1a and the genome references (79.4% to 83.5%) in a rearranged genomic structure. Simultaneously, we identified two highly conserved regions surrounding the ammonia monooxygenase operon and chaperone operon within all the thaumarchaeal genomes – indicating that these genomic regions are highly conserved within the phylum. Pairwise distance analysis between R

proteins together with the ANI results suggested that Antarctic Group I.1a represents a very low diversity Group I.1a of a novel *Nitrosopumilus* species more closely affiliated with the strain "*Ca.* Nitrosopumilus sedimentis" AR2. Therefore, our analyses are contributing to characterization of genomic features and phylogenetic distinctiveness of the Thaumarchaeota phylum (Spang *et al*., 2010). Finally, these results support the necessity of enriching and isolating a Thaumarchaeota Group I.1a strain from Antarctic oceans for futures analyses of physiological adaptation of this microbial assemblage to their environment.

## 4.4. Acknowledgements

This research was supported by NSF, International Polar Year 2009-2010 award ANT-0632389 (to JJG and AEM). EK was supported by Fulbright/CAPES-BRAZIL grant 2163-08-8. An award from the DOE Joint Genome Institute to AEM and JJG supported the environmental genome cloning and sequencing. Computational resources to produce the environmental genome data set available at the IMG-M/DOE Joint Genome Institute (JGI; https://img.jgi.doe.gov/) were provided by NSF EPSCoR grant (EPS-0447416). Thanks are extended to Jim Bristow, Kerrie Barry, and Susannah Tringe at JGI for support during genome sequencing. Special thanks to K.C. King, from the Desert Research Institute, for reviewing the chapter and providing valuable feedback.

## 4.5. References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215,* 403-410.

Amano-Sato, C., Akiyama, S., Uchida, M., Shimada, K., & Utsumi, M. (2013). Archaeal distribution and abundance in water masses of the Arctic Ocean, Pacific sector. *Aquatic Microbial Ecology, 69*(2), 101-112.

Aravind, L., Makarova, K. S., & Koonin, E. V. (2000). Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Research*, *28*(18), 3417–3432.

Béjà, O., Koonin, E. V., Aravind, L., Taylor, L. T., Seitz, H., Stein, J. L., Bensen, D. C., et al. (2002). Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Applied and Environmental Microbiology*, *68*(1), 335–345.

Bartossek, R., Spang, A., Weidler, G., Lanzen, A., & Schleper, C. (2012). Metagenomic analysis of ammonia-oxidizing archaea affiliated with the soil group. *Frontiers in Microbiology*, *3*(208), 1-14.

Berg, I. A., Kockelkorn, D., Vera, W. H. R., Say, R. F., Zarzycki, J., Hügler, M., et al. (2010). Autotrophic carbon fixation in archaea. *Nature Review in Microbiology*, *8*(6), 447-460.

Blainey, P. C., Mosier, A. C., Potanina, A., Francis, C. A., & Quake, S. R. (2011). Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS ONE*, *6*(2), e16626.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., & Forterre, P. (2008). Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Reviews in Microbiology*, *6*(3), 245–252.

Brochier-Armanet, C., Deschamps, P., López-García, P., Zivanovic, Y., Rodríguez-Valera, F., & Moreira, D. (2011). Complete-fosmid and fosmid-end sequences reveal frequent horizontal gene transfers in marine uncultured planktonic archaea. *The ISME Journal*, *5*(8), 1291–1302.

Brochier-Armanet, C., Gribaldo, S., & Forterre, P. (2012). Spotlight on the Thaumarchaeota. *The ISME Journal*, *6*(2), 227–230.

Brudno, M., Malde, S., Poliakov, A., Do, C. B., Couronne, O., Dubchak, I., & Batzoglou, S. (2003). Glocal alignment: finding rearrangements during alignment. *Bioinformatics, 19*(1), 54–62.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., & Parkhill, J. (2005). ACT: the Artemis Comparison Tool. *Bioinformatics*, *21*(16), 3422–3423.

Chiuri, R., Maiorano, G., Rizzello, a, del Mercato, L. L., Cingolani, R., Rinaldi, R., et al. (2009). Exploring local flexibility/rigidity in psychrophilic and mesophilic carbonic anhydrases. *Biophysical Journal*, *96*(4), 1586–1596.

Church, M. J., Wai, B., Karl, D. M., & Delong, E. F. (2010). Abundances of crenarchaeal *amoA* genes and transcripts in the Pacific Ocean. *Environmental Microbiology,12*(3), 679–688.

Copley, R. R., Russell, R. B., & Ponting, C. P. (2001). Sialidase-like Asp-boxes: sequence-similar structures within different protein folds. *Protein Science*, *10*, 285–292.

Couronne, O., Poliakov, A., Bray, N., Ishkhanov, T., Ryaboy, D., Rubin, E., Pachter, L., et al. (2003). Strategies and tools for whole-genome alignments. *Genome Research, 13*(1), 73–80.

DeLong, E. F., Wu, K. Y., Prézelin, B. B., & Jovine, R. V. (1994). High abundance of Archaea in Antarctic marine picoplankton. *Nature*, *371*(6499), 695–697.

Frazer, K. A, Pachter, L., Poliakov, A., Rubin, E. M., & Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Research, 32* (Web Server issue), W273–W279.

Goris, J., Konstantinidis, K. T., Klappenbach, J. A, Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, *57*(1), 81–91.

Gribaldo, S., & Brochier-Armanet, C. (2006). The origin and evolution of Archaea: a state of the art. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1470), 1007–22.

Grzymski, J. J., Carter, B. J., Delong, E. F., Feldman, R. A., Ghadiri, A., & Murray, A. E. (2006). Comparative genomics of DNA fragments from six Antarctic marine planktonic Bacteria. *Applied and Environmental Microbiology*, *72*(2), 1532–1541.

Grzymski, J. J., Riesenfeld, C. S., Williams, T. J., Dussaq, A. M., Ducklow, H., Erickson, M., Cavicchioli, R., et al. (2012). A metagenomic assessment of winter and summer bacterioplankton from Antarctica Peninsula coastal surface waters. *The ISME Journal*, *6*(10), 1901–1915.

Hallam, S. J., Konstantinidis, K. T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., Preston, C., et al. (2006). Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum. Proceedings of the National Academy of Sciences, 103*(48), 18296–18301.

Könneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., & Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*, *437*(7058), 543–546.

Konstantinidis, K. T., Ramette, A., & Tiedje, J. M. (2006). The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1475), 1929–1940.

Labrenz, M., Sintes, E., Toetzke, F., Zumsteg, A., Herndl, G. J., Seidler, M., et al. (2010). Relevance of a crenarchaeotal subcluster related to *Candidatus* Nitrosopumilus maritimus to ammonia oxidation in the suboxic zone of the central Baltic Sea. *The ISME Journal, 4*(12), 1496–1508.

López-García, P., Brochier, C., Moreira, D., & Rodríguez-Valera, F. (2003). Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. *Environmental Microbiology*, *6*(1), 19–34.

Markowitz, V. M., Chen, I.M. A., Chu, K., Szeto, E., Palaniappan, K., Grechkin, Y., Ratner, A., et al. (2011). IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Research*, *40*, 123–129.

Markowitz, V. M., Ivanova, N. N., Szeto, E., Palaniappan, K., Chu, K., Dalevi, D., et al. (2008). IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Research*, *36* (Database issue), D534–D538.

Massana, R., Taylor, L. T., Murray, A. E.,Wu, K. Y., Jeffrey, W. H., & DeLong, E. F. (1998). Vertical distribution and temporal variation of marine planktonic archaea in the Gerlache Strait, Antarctica, during early spring. *Limnology and Oceanography, 43*(4), 607–617.

Molina, V., Belmar, L., & Ulloa, O. (2010). High diversity of ammonia-oxidizing archaea in permanent and seasonal oxygen-deficient waters of the eastern South Pacific. *Environmental Microbiology, 12*, 2450–2465.

Murray, A. E., Preston, C. M., Massana, R., Taylor, L. T., Blakis, A, Wu, K., & DeLong, E. F. (1998). Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. *Applied and Environmental Microbiology*, *64*(7), 2585–2595.

Murray, A. E., Wu, K. Y., Moyer, C. L., Karl, D. M., & DeLong, E. F. (1999). Evidence for circumpolar distribution of planktonic Archaea in the Southern Ocean. *Aquatic Microbial Ecology, 18*, 263–273.

Park, S. J., Kim, J. G., Jung, M. Y., Kim, S. J., Cha, I. T., Ghai, R., Martín-Cuadrado, A. B., et al. (2012a). Draft genome sequence of an ammonia-oxidizing archaeon, "*Candidatus*

Nitrosopumilus sediminis" AR2, from Svalbard in the Arctic Circle. *Journal of Bacteriology, 194*(24), 6948–6949.

Park, S. J., Kim, J. G., Jung, M. Y., Kim, S. J., Cha, I. T., Kwon, K., Lee, J. H., et al. (2012b). Draft genome sequence of an ammonia-oxidizing archaeon, "*Candidatus* Nitrosopumilus koreensis" AR1, from marine sediment. *Journal of Bacteriology*, *194*(24), 6940–6941.

Pester, M., Schleper, C., & Wagner, M. (2011). The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Current Opinion in Microbiology*, *14*(3), 300–306.

Philippe, H., & Douady, C. J. (2003). Horizontal gene transfer and phylogenetics. *Current Opinion in Microbiology*, *6*(5), 498–505.

Preston, C. M., Wu, K. Y., Molinski, T. F. & DeLong, E. F. (1996). A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proceedings of the National Academy of Sciences, 93*(13), 6241–6246.

Purkhold, U., Pommerening-Röser, A., Schmid, M. C., Koops, H., Juretschko, S., Pommerening-Ro, A., & Wagner, M. (2000). Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and *amoA* sequence analysis: implications for molecular diversity surveys. *Applied and Environmental Microbiology, 66*(12), 5368–5382.

Ruvabc, C., Sharples, G. J., Ingleston, S. M., & Lloyd, R. G. (1999). Holliday Junction processing in Bacteria: insights from the evolutionary conservation of RuvABC, RecG, and RusA. *Journal of Bacteriology*, *181*(18), 5543–5550.

Santoro, A. E., Casciotti, K. L., & Francis, C.A. (2010). Activity, abundance and diversity of nitrifying archaea and bacteria in the central California current. *Environmental Microbiology, 12*(7), 1989–2006.

Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit, W., et al. (2010). Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends in Microbiology, 8*(18), 331-340.

Srinivas, U. K., & Ray, M. K. (2006). Cold-stress response of low temperature adapted bacteria. *Molecular Biology*, *661*(2), 1–23.

Swan, B. K., Martinez-Garcia, M., Preston, C. M., Sczyrba, A., Woyke, T., Lamy, D., Reinthaler, T., et al. (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science*, *333*(6047), 1296–300.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution, 28*(10), 2731-2739.

Walker, C. B., De La Torre, J. R., Klotz, M. G., Urakawa, H., Pinel, N., Arp, D. J., Brochier-Armanet, C., et al. (2010). *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proceedings of the National Academy of Sciences, 107*(19), 8818–8823.

Williams, T. J., Long, E., Evans, F., Demaere, M. Z., Lauro, F. M., Raftery, M. J., et al. (2012). A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *The ISME Journal*, *6*(10), 1883–900.

## 4.6. Supplemental material

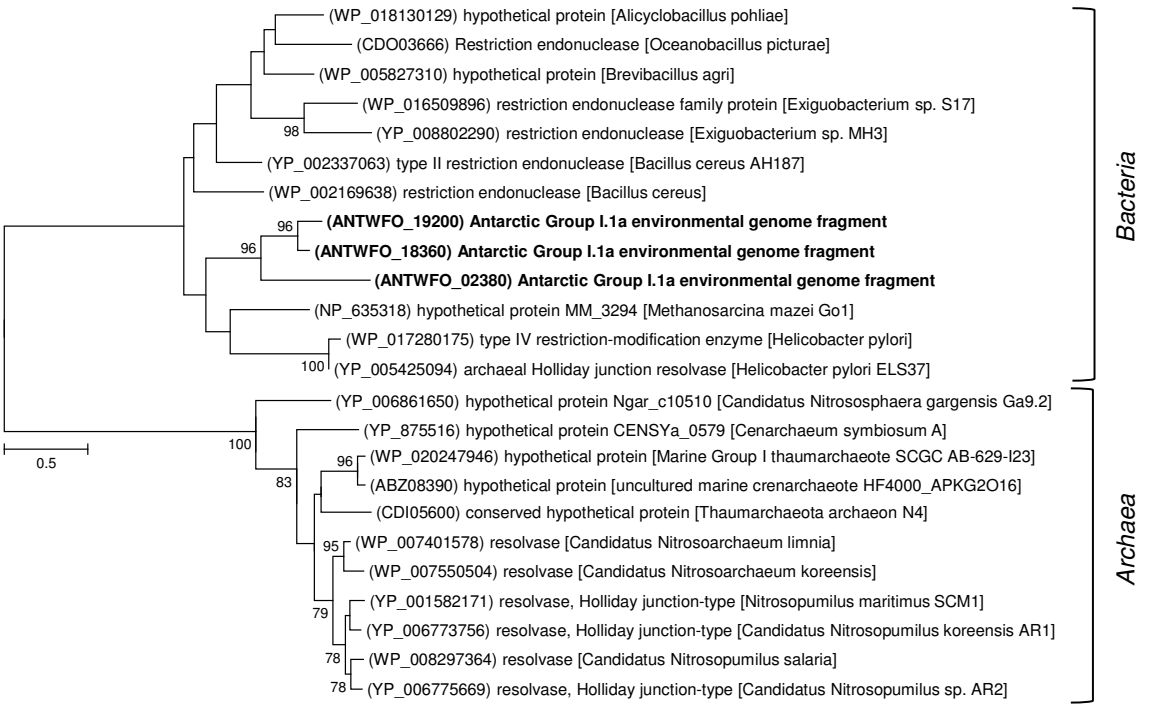



**Figure S4.1.** Phylogenetic affiliation of Antarctic Group I.1a Holliday junction resolvase and Mrr-like restriction enzymes. The maximum likelihood tree was based on the alignment of 154 amino acid position. The scale bar represents 0.5 changes per amino acid position. Bootstrap values are based on 1000 repetitions.

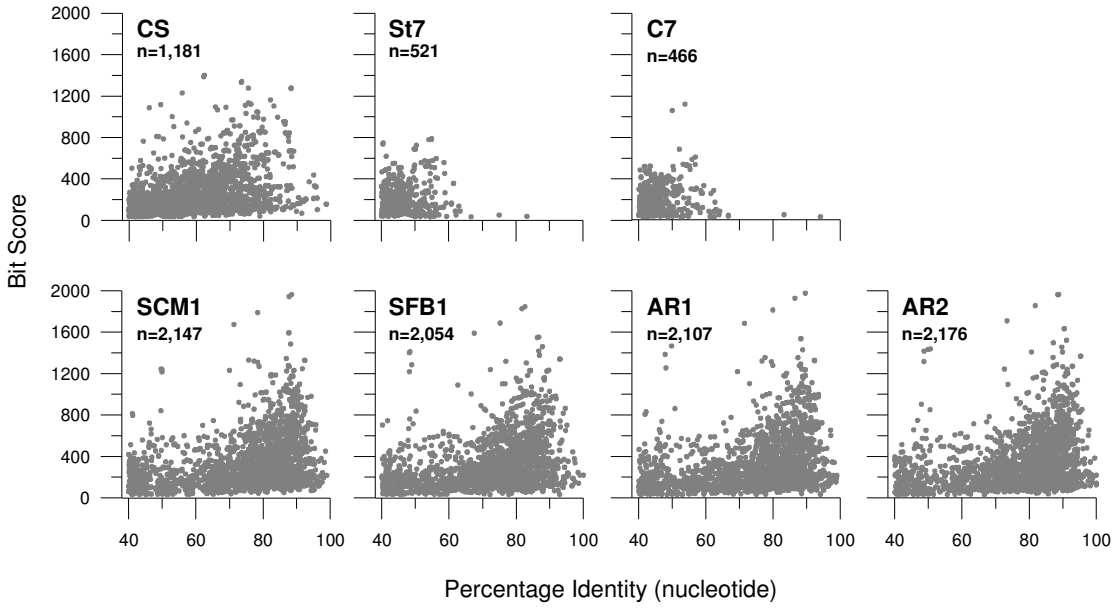



**Figure S4.2.** Best homologous hit within Antarctic Group I.1a ORFs and genome reference ORFS. CS: *C. symbiosum* A; St7: Crenarchaeota S*ulfulobus tokodaii* strain 7; C7: *Methanococcus maripaludis* C7; SCM1: *N. maritimus* SCM1; SFB1: "*Ca.* Nitrosoarchaeon limnia" SFB1; AR1: "*Ca.* Nitrosopumilus koreensis" AR1; AR2: "*Ca.* Nitrosopumilus sedimentis" AR2. N indicates the number of matches between the environmental ORFs and genome reference.

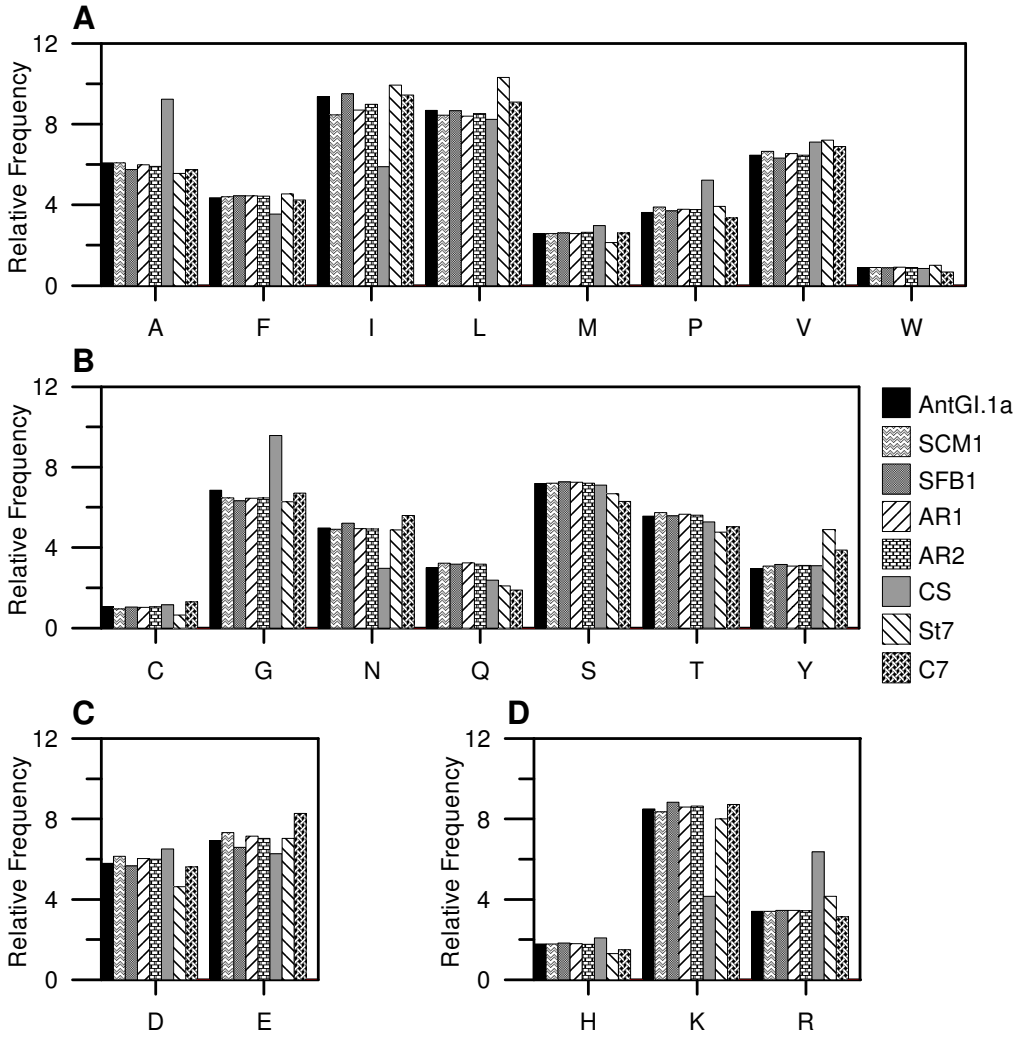



**Figure S4.3.** Amino acid distribution within the genomes analyzed. AntarGI.1a: Antarctic Group I.1a genome fragments; SCM1: *N. maritimus* SCM1; SFB1: "*Ca.* Nitrosoarchaeon limnia" SFB1; AR1: "*Ca.* N. koreensis" AR1; AR2: "*Ca.* N. sedimentis" AR2; CS: *C. symbiosum* A; St7: S*ulfolobus tokodaii* strain 7; C7: *Methanococcus maripaludis* C7. (A) Amino acids: Polar, hydrophobic: A- alanine, F– phenylalanine, I– isoleucine, L– leucine, M– methionine, P– proline, V– valine, W– tryptophan. (B) Polar, uncharged: C- cysteine, G- glycine, N- asparagine, Q- glutamine, S- serine, T- threonine, Y- tyrosine. (C) Polar, acid: D- aspartic acid, E- glutamine acid. (D) Polar basic: K- lysine, R- arginine, H- histidine.

**Table S4.1.** List of the Antarctic GroupI.1a fosmids recruited to the SCM1 genome. Scaffold size in bp, GC content, and start and stop positions of scaffold alignment in the *N. maritimus* SCM1genome (Figure 4.1) are specified.

| ID | IMG fosmid ID | Scaffold Size (bp) | Avg % GC | Position in SCM1 Start | Position in SCM1 Stop |
|---|---|---|---|---|---|
| 1 | ANTWF_4084609__Contig11<br>ANTWF_4084609__Contig12<br>ANTWF_4084651__Contig6 | 48,910 | 28 | 64,870<br>1,580,684 | 80,130<br>1,622,477 |
| 2 | ANTWF_4084617__Contig6<br>ANTWF_4084617__Contig7 | 39,506 | 31 | 161,291<br>1,456,797 | 168,882<br>1,507,677 |
| 3 | ANTWF_4084641__Contig2<br>ANTWF_4084652__Contig2<br>ANTWF_4084653__Contig7<br>ANTWF_4084645__Contig6 | 113,139 | 30 | 441,743 | 590,924 |
| 4 | ANTWF_4084608__Contig8<br>ANTWF_4084616__Contig5 | 68,217 | 29 | 568,719 | 641,547 |
| 5 | ANTWF_4084622__Contig6<br>ANTWF_4084654__Contig1 | 41,288 | 28 | 250,980 | 288,924 |
| 6 | ANTWF_4084649_Contig8 | 39,161 | 30 | 662,218<br>831,483<br>623,737 | 673,014<br>846,876<br>661,637 |
| 7 | ANTWF_4084648_Contig4 | 44,827 | 31 | 1,456,796<br>161,294 | 1,464,460<br>205,974 |
| 8 | ANTWF_4084647_Contig4 | 41,130 | 30 | 161,291<br>1,456,797 | 173,588<br>1,487,578 |
| 9 | ANTWF_4084646_Contig4 | 44,282 | 31 | 161,270<br>1,456,797 | 167,612<br>1,492,384 |
| 10 | ANTWF_4084638_Contig2 | 42,121 | 31 | 198,248 | 240,674 |
| 11 | ANTWF_4084627_Contig4 | 36,189 | 31 | 1,314,280 | 1,367,733 |
| 12 | ANTWF_4084614_Contig5 | 39,084 | 30 | 1,017,407 | 1,122,394 |
| 13 | ANTWF_4084609_Contig10 | 16,099 | 30 | 79,942 | 94,539 |
| 14 | AWFS_FIUX9419_contig02202 | 22,748 | 31 | 1,382,615 | 1,410,501 |
| 15 | AWFS_FIUX9419_contig02203 | 17,407 | 33 | 1,411,431 | 1,428,668 |
| 16 | AWFS_FIUX1598_contig01546 | 19,573 | 31 | 83,211 | 103,870 |
| 17 | AWFS_FIUX2292_contig01039 | 24,805 | 29 | 1,580,678<br>64,833 | 1,589,493<br>83,116 |
| 18 | AWFS_FIUX2292_contig01040 | 19,614 | | 83,139 | 103,797 |
| 19 | AWFS_FIUX5018_contig02086 | 44,128 | 31 | 1,073,458 | 1,140,381 |
| 20 | AWFS_FIUX6710_contig02206 | 13,336 | 32 | 1,394,389 | 1,407,582 |
| 21 | AWFS_FIUX1083<br>AWFS_FIUX5878 | 17,895 | 30 | 1,345,215<br>1,335,755 | 1,360,824<br>1,336,534 |
| 22 | AWFS_FIUX1370<br>AWFS_FIUX2477 | 16,408 | 29 | 1,302,972 | 1,335,637 |
| 23 | AWFS_FIUX2394<br>AWFS_FIUX6710 | 23,186 | 31 | 1,424,312 | 1,447,411 |

*(Table, continued)*

*(Table, continuation)*

| | | | | | |
|---|---|---|---|---|---|
| 24 | AWFS_FIUX9646_contig02159 | 30,023 | | 238,521 | 268,147 |
| 25 | AWFS_FIUX471_contig02189 | 32,334 | 31 | 1,476,526 | 1,516,056 |
| 26 | AWFS_FIUX6338<br>AWFS_FIUX7129<br>AWFS_FIUX7901<br>AWFS_FIUX8785 | 13,433 | 33 | 1,180,779 | 1,198,706 |
| 27 | AWFS_FIUX2394<br>AWFS_FIUX6710 | 13,131 | 35 | 1,410,366 | 1,424,042 |
| 28 | AWFS_FIUX1370<br>AWFS_FIUX7901 | 34,123 | 32 | 1,199,849 | 1,279,197 |
| 29 | AWFS_FIUX3938<br>AWFS_FIUX8562 | 38,937 | 32 | 163,318 | 207,055 |
| 30 | AWFS_FIUX3938<br>AWFS_FIUX8562<br>AWFS_FIUX9646 | 28,758 | 31 | 207,410 | 238,755 |
| 31 | AWFS_FIUX8091<br>AWFS_FIUX9987<br>AWFS_FIUX10228 | 13,922 | 33 | 1,121,013 | 1,136,897 |
| 32 | AWFS_FUIX6338<br>AWFS_FUIX7129<br>AWFS_FUIX8091<br>AWFS_FUIX9987 | 20,610 | 32 | 1,149,038 | 1,178,384 |
| 33 | AWFS_FIUX1830_contig02119 | 39,694 | 32 | 895,616 | 941,007 |

**Table S4.2.** Ribosomal proteins identified within Antarctic Group I.1a genome fragments (AntGI.1a). Subsequent columns show the detection of the ribosomal proteins in the archaeal genomes used as genome references. LSU: large subunit and SSU: small subunit. SCM1: *N. maritimus* SCM1; SFB1: "*Ca.* Nitrosoarchaeon limnia" SFB1; AR1: "*Ca.* N. koreensis" AR1; AR2: "*Ca.* N. sedimentis" AR2; CS: *C. symbiosum* A; St7: S*ulfolobus tokodaii* strain 7; C7: *Methanococcus maripaludis* C7. Color gradient correlates to the number of gene copies identified in each genome.

| Ribosomal protein | AntGI.1a | SCM1 | SFB1 | AR1 | AR2 | CS |
|---|---|---|---|---|---|---|
| LSU ribosomal protein L2P | 3 | 1 | 1 | 1 | 1 | 1 |
| LSU ribosomal protein L7AE | 4 | 1 | 1 | 1 | 1 | 1 |
| LSU ribosomal protein L15E | 3 | 2 | 2 | 2 | 2 | 1 |
| LSU ribosomal protein L21E | 2 | 1 | 1 | 1 | 1 | 1 |
| LSU ribosomal protein L24E | 4 | 1 | 1 | 1 | 1 | 1 |
| LSU ribosomal protein L31E | 7 | 1 | 1 | 1 | 1 | 1 |
| LSU ribosomal protein L37E | 2 | 1 | 1 | 1 | 1 | 1 |
| LSU ribosomal protein L39E | 3 | 1 | 1 | 1 | 1 | 1 |
| LSU ribosomal protein L40E | 3 | 1 | 1 | 1 | 1 | 1 |
| LSU ribosomal protein L44E | 1 | 1 | 1 | 1 | 1 | 1 |
| SSU ribosomal protein S3AE | 1 | 1 | 1 | 1 | 1 | 1 |
| SSU ribosomal protein S4P | 1 | 1 | 1 | 1 | 1 | 1 |
| SSU ribosomal protein S8E | 2 | 1 | 1 | 1 | 1 | 1 |
| SSU ribosomal protein S10P | 1 | 1 | 1 | 1 | 1 | 1 |
| SSU ribosomal protein S11P | 2 | 1 | 1 | 1 | 1 | 1 |
| SSU ribosomal protein S13P | 1 | 1 | 1 | 1 | 1 | 1 |
| SSU ribosomal protein S15P | 1 | 1 | 1 | 1 | 1 | 1 |
| SSU ribosomal protein S17E | 1 | 1 | 1 | 1 | 1 | 2 |
| SSU ribosomal protein S19E | 3 | 1 | 1 | 1 | 1 | 1 |
| SSU ribosomal protein S27E | 1 | 1 | 1 | 1 | 1 | 1 |
| SSU ribosomal protein S28E | 4 | 1 | 1 | 1 | 1 | 1 |
| SSU ribosomal protein S30E | 2 | 0 | 0 | 0 | 0 | 1 |

**Table S4.3.** Percentage of nucleotide identity of homologous genes identified in Antarctic Group I.1a genome fragments and fosmid 74A4 (Béjà *et al.*, 2002).

| Gene ID in Antarctic Group I.1a | Gene ID in fosmid 74A4 | % nucleotide identity | Aligment size in nucleotide positions | Predicted function |
|---|---|---|---|---|
| 2042522490 | AAK96067 | 84.60% | 594 | NAD_synthase |
| 2042522491 | AAK96068 | 84.70% | 1305 | cysteinyl-tRNA_synthetase |
| 2080470906 | AAK96100 | 85.40% | 378 | hypothetical_protein |
| 2080470900 | AAK96103 | 88.00% | 225 | tRNA_intron_endonuclease |
| 2080470920 | AAK96089 | 90.30% | 165 | mannose-6-phosphate_isomerase-like_protein |
| 2080470919 | AAK96090 | 91.00% | 672 | molecular_chaperone |
| 2080470923 | AAK96087 | 91.60% | 336 | double-stranded_beta-helix_fold_enzyme |
| 2080470925 | AAK96086 | 91.70% | 822 | TPR-repeat_protein |
| 2080470905 | AAK96101 | 92.10% | 840 | ATP-dependent_DNA_ligase |
| 2080470898 | AAK96104 | 93.20% | 237 | hypothetical_protein |

**Table S4.4.** Relative abundance (in percentage) of arginine (Arg), asparctic acid (Asp), glutamine acid (Glu), and proline (Pro) residues within the protein sets. AntGI.1a: Antarctic Group I.1a; SCM1: *N. maritimus* SCM1; SFB1: "*Ca.* Nitrosoarchaeon limnia" SFB1; AR1: "*Ca.* N. koreensis" AR1 AR2: "*Ca.* N. sedimentis" AR2; CS: *C. symbiosum* A.

| Set of proteins | Amino acid | AntGI.1a | SCM1 | SFB1 | AR1 | AR2 | CS |
|---|---|---|---|---|---|---|---|
| R proteins | Arg | 8.21 | 8.03 | 8.40 | 8.03 | 8.07 | 9.94 |
| | Asp | 3.97 | 4.09 | 3.90 | 4.05 | 4.09 | 4.71 |
| | Glu | 5.95 | 5.94 | 5.52 | 5.94 | 5.75 | 4.95 |
| | Pro | 4.76 | 4.68 | 4.85 | 4.53 | 4.57 | 4.63 |
| | Arg/(Arg+Lys) | 0.44 | 0.43 | 0.45 | 0.43 | 0.43 | 0.52 |
| Genes associated to 3HP/4HB | Arg | 3.48 | 3.25 | 3.38 | 3.42 | 3.48 | 5.23 |
| | Asp | 6.11 | 5.92 | 5.76 | 6.05 | 5.99 | 6.05 |
| | Glu | 6.26 | 6.75 | 6.07 | 6.6 | 6.51 | 6.36 |
| | Pro | 3.96 | 4.04 | 4.06 | 4.1 | 4 | 4.14 |
| | Arg/(Arg+Lys) | 0.31 | 0.28 | 0.28 | 0.29 | 0.30 | 0.44 |
| Ammonia monooxygenase opeson | Arg | 3.09 | 3.09 | 3.09 | 2.87 | 3.02 | 3.36 |
| | Asp | 2.53 | 2.39 | 2.53 | 2.27 | 2.57 | 2.52 |
| | Glu | 3.93 | 4.07 | 4.07 | 4.38 | 4.08 | 4.34 |
| | Pro | 5.06 | 5.06 | 5.06 | 5.44 | 5.29 | 5.04 |
| | Arg/(Arg+Lys) | 0.49 | 0.50 | 0.47 | 0.44 | 0.47 | 0.53 |
| Chaperone operon | Arg | 3.56 | 4.03 | 4.14 | 4.03 | 3.66 | 5.91 |
| | Asp | 7.32 | 7.96 | 7.17 | 8.17 | 7.83 | 7.94 |
| | Glu | 6.3 | 6.55 | 7.17 | 6.55 | 6.51 | 6.42 |
| | Pro | 4.27 | 3.73 | 4.24 | 4.03 | 4.48 | 4.38 |
| | Arg/(Arg+Lys) | 0.29 | 0.32 | 0.30 | 0.32 | 0.29 | 0.45 |

# Chapter V

## Final considerations and future work

The main goal of this dissertation was to expand knowledge of Antarctic microbial life. Antarctica is the coldest, driest, and windiest continent on Earth. Even with these extreme environmental conditions, the Antarctic's soils, ice, lakes, subglacial lakes, rocks, sediments, and surrounding seawaters have been successfully colonized by microorganisms that are capable of surviving and thriving at low temperatures and maintaining metabolic activities at subzero temperatures. Study of psychrophile's diversity and metabolism is of great application for biotechnology (through discovery of new resources such as anti-freezing proteins, antibiotics, and detergents), evolution (through discovery of new taxa and metabolic pathways), and astrobiology (as a model for life that may exist on icy bodies such as Europa, Mars, and Enceladus). Therefore, this dissertation seeks to contribute to these three fields by advancing the following:

1. Revealing the nature of the ultra-small microbial assemblage in Lake Vida encapsulated cryobrine – challenging the previously known limits of minimum

cell size. For a spherical bacterial with minimal metabolic capacity (250 genes), the minimum size of the cell would be 172 nm in diameter to accommodate DNA, ribosomes, and proteins (Adams, 1999). Lake Vida ultramicrocells – 0.192 ± 0.065 μm in diameter – are a consequence of stress response to environmental conditions or related to aging and are not phylogenetically affiliated with bacteria known to be ultramicrocells. Taking a step further in considering the existence of life in other celestial bodies, Lake Vida life forms are an example of how cells could look in icy worlds such as Europa, where life, if present, is probably encapsulated in freezing brine within ice, such as in Lake Vida.

2. Recovering cultivable bacteria from the cryobrine that now can be prospected for new bioproducts and tested for physiological adaptation to extreme conditions, such as cold temperature and high salinity. *Marinobacter* and *Psychrobacter*-affiliated strains were previously isolated from Lake Vida brine (Mondino *et al.,* 2005; Murray *et al*., 2012). Cultivating brine cells from different size fractions added five new species to Lake Vida culture collection, increasing the number of taxa for bioprospecting.
3. Identifying and describing the existence of a diverse microbial assemblage in the cold and dark deep ice of Lake Vida from18 to 27 m below the lake surface. By microscopy analyses and 16S rRNA screening, well preserved microorganisms related to organisms that carry the metabolic capacity to live chemolithotrophically, by reducing and oxidizing sulfur compounds, and heterotrophically, by degrading complex organic polymers were found. This investigation, like investigation of the ultramicrocells in the lake brine, provides

insight into the existence of life in icy worlds elsewhere in the solar system and universe.

4. Characterizing a metagenome from Antarctic surface seawaters and finding a very low diversity Thaumarchaeota Group I.1a that appears to be representative of archaeal cells not yet cultivated. These findings add knowledge to the diversity of the phylum in Antarctic oceans and contribute to general characterization of the recently described phylum Thaumarchaeota.

Despite the findings of these studies, future approaches could be useful to further increase this knowledge including the following:

(i) Testing the response of bacterial isolates cultivated from Lake Vida brine (Chapter II) to distinct conditions, such as cold temperatures and high salinity, and monitoring their cellular, biochemical, and transcriptional alterations to help establish a complete understanding of the Lake Vida brine ultra-small microbial assemblage. Pleomorphism, aging, and production of membrane vesicles are cellular mechanisms that remain to be explored in *Pseudoalteromonas, Herbaspirillum*, *Marinobacter,* and *Psychrobacter* genera.

(ii) Use of high-throughput next generation sequencing, such as iTag sequencing, to describe the Lake Vida brine microbial assemblage will be useful to compare, at the same level of data generated, microbial assemblages from the ice and sediment (investigated in Chapter III) with the microbial assemblage in the brine.

(iii) Development of rRNA (as cDNA) iTag libraries from the microbial assemblage within the ice and comparison of these new results with results in Chapter III to complement the description of the Lake Vida microbial assemblage.

(iv) Enriching and cultivating Antarctic archaea represented by metagenomic data to further study its physiological capacities, biogeography, and evolution. Access to recently sequenced Thaumarchaeota genomes helped with identification and characterization of a set of genome fragments inferred to be associated with a very low diversity Group I.1a assemblage, supporting the benefits of this approach.

Study of microbial diversity in environments such as those found in Antarctica is essential to the study of life on Earth and elsewhere. Therefore, this dissertation should serve as a resource of knowledge and approaches for students and scientists fascinated by microbial life in cold environments, Antarctic life, and life in similar environments elsewhere beyond Earth.

## 5.1. References

Adams, M. W.W. (1999). The influence of environment and metabolic capacity of the size of a microorganism. *Size Limits of Very Small Microorganisms Proceedings of a Workshop.* National Academy Press, Washington, D.C. pp74-80.

Mondino, L. J., Asao, M., & Madigan, M. T. (2009). Cold-active halophilic bacteria from the ice-sealed Lake Vida, Antarctica. *Archives of Microbiology*, *191*(10), 785–90.

Murray, A. E., Kenig, F., Fritsen, C. H., McKay, C. P., Cawley, K. M., Edwards, R., et al. (2012). Microbial life at -13°C in the brine of an ice-sealed Antarctic lake. *Proceedings of the National Academy of Sciences*, *109*(50), 20626–20631.

***"Most of the life on the planet is microbial.***

***You can certainly run a world without dinosaurs and humans,***

***but you can't do it without microbes."***

***Dr. Andrew H. Knoll***