Table of contents

Declaration III

Acknowledgement IV

Abstract V

Table of contents VII

List of figures XII

List of tables XV

List of appendices XVI

Chapter 1 1

1 Extremophiles in the 'omics' era 2

1.1 Introduction 2

1.2 Extremophiles 3

1.2.1 Psychrophilic (Cold-adapted) microorganisms 5

1.2.2 Halophilic (Salt Adapted) Microorganisms 8

1.2.3 Alkaliphiles (High pH-Adapted) Microorganisms 10

1.2.4 Industrial applications of extremophiles 11

1.3 Genomics and 'omic technologies 12

1.3.1 Genomics 12

1.3.1.1 Sequencing technologies 14

1.3.1.2 Genome assembly 16

1.3.1.3 Genome annotation 18

1.3.2 Comparative genomics 19

1.3.3 'Omic' technologies 22
1.3.3.1 Transcriptomics 22
1.3.3.2 Proteomics and other 'omics technologies 24
1.4 Extremophiles in the 'omics' era 25
1.4.1 Psychrophiles in the 'omics' era 26
1.4.2 Halophiles in the 'omics' era 29
1.4.3 Alkaliphiles in the 'omics' era 31
1.5 Conclusions 32
1.6 References 33
Chapter 2 50
2 Sequencing, assembly, annotation of the draft genome of Antarctic polyextremophile *Nesterenkonia* sp. AN1 and identification of putative molecular determinants underlying its polyextremophily 51
2.1 Introduction 51
2.2 Material and Methods 54
2.3 Genome Sequence of *Nesterenkonia* sp. AN1 54
2.3.1 Culturing and genomic DNA extraction 54
2.3.2 Genome Sequencing and assembly 55
2.3.3 Genome Annotation 56
2.3.4 Identification of genomic islands, prophages and extrachromosomal elements 57
2.3.5 Identification of stress response mechanisms 57
2.4 Results and Discussion 58
2.4.1 Genome sequencing and assembly 58

2.4.2 Genome annotation 59
2.4.3 Genomic islands, prophages and other extrachromosomal elements 63
2.5 Stress response mechanisms 67
2.5.1 Cold stress response 68
2.5.1.1 Cold shock proteins and cold acclimation proteins 68
2.5.1.2 Membrane adaptations 70
2.5.2 Oxidative stress response 71
2.5.3 Protection against osmotic stress 72
2.5.4 General stress response 73
2.6 Conclusion 74
2.7 References 75
Chapter 3 85
3 Comparative genomic analyses of four *Nesterenkonia* species 86
3.1 Introduction 86
3.2 Materials and Methods 88
3.2.1 Genome sequences and structural annotation 88
3.2.2 Phylo-genomic analyses 88
3.2.3 “Pan-genome” analysis 89
3.2.4 Detection of positive selection 90
3.2.4.1 Determination of Orthologous gene sets 90
3.2.4.2 Detection of recombination and positive selection 90
3.3 Results and Discussion 92
3.3.1 General genome metrics 92

3.3.2 Phylo-genomics of *Nesterenkonia* spp 96
3.3.3 The 'pan-genome' of four *Nesterenkonia* strains 100
3.3.3.1 Comparative genomics identifies unique stress response proteins in Nesterenkonia sp. AN1 105
3.3.4 Signatures of positive selection are prevalent in the core genome of *Nesterenkonia* ... 107
3.3.4.1 Orthologs sets and alignment strategy 107
3.3.4.2 Recombination and positive selection 108
3.4 Conclusion 115
3.5 References 117
Chapter 4 126
4 Transcriptome analysis reveals several key adaptive features for the survival of *Nesterenkonia* sp. AN1 under cold conditions 127
4.1 Introduction 127
4.2 Material and Methods 128
4.2.1 Growth conditions 128
4.2.2 RNA preparation 129
4.2.3 RNA sequencing 130
4.2.4 RNA-Seq analysis 130
4.3 Results and Discussion 131
4.3.1 RNA-Seq read mappings 131
4.3.2 Analysis of gene expression 132
4.3.2.1 Most stress response genes were expressed under both cold and optimal growth conditions 135

4.3.2.2 Induction of genes encoded on the core genome 137

4.3.2.3 Enrichment of biological processes associated with growth at 5 ºC 139

4.4 Conclusions 141

4.5 References 143

Chapter 5 147

5 General conclusions 148

5.1 Introduction 148

5.2 Adaptations of the Antarctic *Nesterenkonia* 148

5.3 Concluding remarks and future perspectives 151

5.4 References 152

Appendices 154

List of figures

Figure 1-1: Thermal life limits on Earth. 6

Figure 1-2 GOLD genome projects statistics on the number of projects related to extremophilic organisms. 13

Figure 2-1: Circular representation of the genome structure in *Nesterenkonia* sp. AN1 plotted on the CG View Server. 60

Figure 2-2: Proportion of CDSs in the *Nesterenkonia* sp. AN1 genome characterised using the Actinobacteria non-supervised orthologous groups (actNOG)..... 61

Figure 2-3: Functional categories of CDSs of *Nesterenkonia* sp. AN1. 62

Figure 2-4: Putative localisation of *Nesterenkonia* sp. AN1 proteins predicted using the CELLO2GO server. 63

Figure 2-5. Prophage proteins encoded on the *Nesterenkonia* sp. AN1 genome identified and annotated using PHAST. 64

Figure 2-6: Circular plot showing the positions of the genomic islands (GIs) predicted on the genome of *Nesterenkonia* sp. AN1.. 65

Figure 2-7: Comparison of the functional annotation of predicted proteins encoded on the genome of *Nesterenkonia* sp. AN1 and the genomic islands identified in the genome. 66

Figure 2-8: Important stress response features of *Nesterenkonia* sp. AN1. 69

Figure 3-1: Heat maps showing (A) Comparison of amino acids proportions in proteins of four strains of *Nesterenkonia*. (B) Abundance of tRNA genes in the four compared *Nesterenkonia* strains. 93

Figure 3-2: Whole genome phylogeny of the four *Nesterenkonia* strains for which genomes are available and the genomes of other members of the family Micrococcaceae. 98

Figure 3-3: Phylogeny constructed on the basis of the average amino acid identity (AAI) values for the core proteins of the four *Nesterenkonia* strains and selected members *Micrococceae* using MEGA6 program. 98

Figure 3-4: ANI and dDDH among the four *Nesterenkonia* strains and *Arthrobacter* sp. FB24. 99

Figure 3-5: POCP of four *Nesterenkonia* strains and *Arthrobacter* sp. FB24. 99

Figure 3-6: Venn diagram representation of the comparison between sequenced genomes of the genus *Nesterenkonia* based on orthologous relationships among the predicted CDSs.... 101

Figure 3-7: Proportion of proteins CDSs in the *Nesterenkonia* core and accessory genome fractions involved in the super functional categories in the COG and EGGNOG databases. 102

Figure 3-8: Functional categories of core and accessory proteins in the four compared *Nesterenkonia* strains. 103

Figure 3-9: Proportion of proteins in the core and accessory genomes of four *Nesterenkonia* strains characterised based using COG and EGGNOG. 104

Figure 3-10: Number of adaptation genes associated with different pan-genome fractions. 105

Figure 3-11: Phylogeny of four *Nesterenkonia* strains. The maximum likelihood tree was computed from 1,108 core genes of *Nesterenkonia* spp and *Arthrobacter* sp. FB24 trimmed using Gblocks0.9b..... 107

Figure 3-12: Proportions of recombination detection of 159 gene families using the three methods included in PhiPack program. 109

Figure 3-13: Distribution of gene families under positive selection based on branch-site test (codeml) in *Nesterenkonia* spp. 110

Figure 3-14: COG features of all gene families under positive selection based on branch-site test (Codeml) in *Nesterenkonia* spp. 111

Figure 3-15: The TCA and glyoxylate pathways highlighting specific *Nesterenkonia sp.* AN1 enzymes subjected to positive selection. 114

Figure 4-1: Growth profiles of cultures of *Nesterenkonia* sp. AN1 grown at 5 ºC and 21 ºC. 129

Figure 4-2: Track lists showing RNA-Seq read coverage for a section of contig 1 of the draft genome of *Nesterenkonia* sp. AN1 visualised using the track list function of the CLC Genomics Workbench 7.5. 132

Figure 4-3: Volcano plot showing transcripts expression profile in *Nesterenkonia* sp. AN1 grown at 5 ºC and 21 ºC. 132

Figure 4-4: Hierarchical clustering of the differentially expressed transcripts in *Nesterenkonia* sp. AN1 at 5°C (cold) relative to 21°C.(control). 133

Figure 4-5*:* EggNOG functional categorization of the differentially expressed genes of *Nesterenkonia* sp. AN1 at 5°C relative to 21°C based on EGGNOG. 134

Figure 4-6: EGGNOG classification of differentially expressed genes in *Nesterenkonia* sp. AN1 grown at 5°C. 135

Figure 4-7: Distribution of differentially expressed genes in the pan genome of *Nesterenkonia*. 138

Figure 4-8: GEOEAST graph of significantly enriched molecular functions for differentially expressed genes in *Nesterenkonia* sp. AN1 grown at 5°C and 21°C. 141

List of tables

Table 1-1: Extremophiles and their growth conditions. 4

Table 2-1: Statistics for the assemblies obtained using different assemblers. 58

Table 3-1: General genome features of the compared *Nesterenkonia* strains. 95

Table 3-2: GO terms enriched for all positively selected genes identified using branch-site test in the four strains of *Nesterenkonia*. 112

Table 4-1: Mapping statistics for *Nesterenkonia* sp. AN1 RNA-Seq paired-end (PE) reads. 131

List of appendices

Appendix F 1: Circular representation of the genome of *Nesterenkonia* sp. AN1 and three mesophilic *Nesterenkonia* strains plotted using the CG View Server. 154

Appendix F 2: Overview of the transcriptome of *Nesterenkonia* sp. AN1 grown at 5°C and 21°C mapped against KEGG metabolic pathways using Ipath2.0. 155

Appendix F 3: REVIGO interactive graph showing gene ontology (GO) terms enriched by the overexpressed genes at 5°C relative to 21°C in *Nesterenkonia* sp. AN1. 156

Appendix F 4: REVIGO (Supek *et al.*, 2011) interactive graph showing gene ontology (GO) terms enriched by the repressed genes at 5°C relative to 21°C in *Nesterenkonia* sp. AN1. 157

Appendix F 5: Proportion of differentially expressed genes associated with different KEGG pathways for *Nesterenkonia* sp. AN1grown at 5°C and 21°C. 158

Appendix F 6: Number of differentially expressed genes at 5°C relative to 21°C enriched in specific KEGG pathways. 159

Appendix T 1: The genome islands (GIs) identified in *Nesterenkonia* sp. AN1. 160

Appendix T 2. Genome features of *Nesterenkonia* sp. AN1 involved in adaptation to different stresses. 162

Appendix T 3. COG functional classification of positively selected core gene families in *Nesterenkonia* sp. AN1. 176

Appendix T 4. List of differentially expressed genes at 5 ºC relative to 21 ºC in *Nesterenkonia* sp. AN1 206

Appendix T 5. GOEAST annotation of differentially expressed genes associated with molecular functions (GO-term) in *Nesterenkonia* sp. AN1 grown at 5°C and 21°C. 223

Chapter 1

1 Extremophiles in the 'omics' era

1.1 Introduction

Over the last five decades, a large number of microorganisms have been discovered living within ecological niches where they are exposed to various environmental extremes (Canganella and Wiegel, 2011, Rothschild and Mancinelli, 2001). These microorganisms are continuously challenging the boundaries in what we consider to be the extreme limits of life. For instance, a hyper-thermophilic methanogen, *Methanopyrus kandleri* strain 116, has been shown to proliferate at the upper life thermal limit of 122 ºC and under an immense pressure of 20 MPa (Takai *et al.*, 2008). *Geogemma barossii* strain 121 has also been reported to proliferate at 121 ºC and could tolerate temperatures as high as 130 ºC (Kashefi and Lovley, 2003). *Thermococcus gammatolerans,* which was isolated from hydrothermal vent at a depth of 2,616 m can be exposed to radiation levels that are 100 times greater than at the surface of the Earth (Jolivet *et al.*, 2003). The archaeon *Pyrococcus* CH1 was collected from a hydrothermal site on the mid Atlantic ridge at the depth of 4, 100 m under severe heat (108 ºC) and pressure of (120 MPa) (Zeng *et al.*, 2009). At the lower thermal limit, the Gram-negative anaerobe *Serratia marcescens* has been reported to maintain normal metabolic activity at -20 ºC but the doubling time at this temperature was 160 days (Rivkina *et al.*, 2000).

The discovery of Taq polymerase from the extremophilic bacterium *Thermus aquaticus* (Brock and Freeze, 1969), which has found universal use in molecular biology, set the stage for the exploration of extreme environments for the discovery of novel extremophiles and the vast array of industrially and biotechnologically applicable biomolecules they produce. Notable extremophile products include extremozymes (e.g. lipases and proteases) and biomedicals (e. g. antibiotics and anti-oxidants) (Horikoshi and Bull, 2011). Furthermore, the study of these organisms has provided tremendous insight into the origin of life and the likelihood of its

existence on other planets (Canganella and Wiegel, 2011, Canganella and Wiegel, 2014, Rothschild and Mancinelli, 2001).

Studying the genome sequence of extremophiles holds the key to unravelling the genetics, evolution and mechanisms of their adaptations to harsh environments. Additionally, these studies could fast-track the discovery and development of novel extremozymes and other biologically useful products (DeLong, 2000, Majhi *et al.*, 2013). The development of cost-effective, massively parallel sequencing technologies (Logares *et al.*, 2012) and a wide range of user-friendly and open source analysis software and databases (Stothard and Wishart, 2006, Zhang *et al.*, 2011) has also made genome sequencing and analysis of extremophilic genomes both feasible and facile.

This thesis focuses on an analysis of the molecular determinants of adaptation and the survival strategies employed by the Antarctic polyextremophilic bacterium, *Nesterenkonia* sp. AN1.

1.2 Extremophiles

Extremophiles are found in environments where they are subjected to severe physiological and geochemical stresses (examples, extremes of temperature and pH, hyper saline conditions, high pressure and elevated levels of radiation) (Table 1-1) (MacElroy, 1974, Rothschild and Mancinelli, 2001). These extremophiles have evolved in all three domains of life, although an overwhelming majority of extremophiles are microorganisms (Horikoshi and Bull, 2011). Most of the well-known extremophiles belong to the domain Archaea (Van Den Burg, 2003), but some genera of Bacteria are also known to be well adapted to inhabit extreme environments (Horikoshi and Bull, 2011). Extremophiles adapted to withstand a combination of different environmental stresses are termed polyextremophiles (Mesbah and Wiegel, 2008). Extremophile studies have been encouraged, in part, by the endeavours of researchers to understand the origins of life and the potential presence of life on other planets (Dartnell, 2011).

Table 1-1: Extremophiles and their growth conditions. (*) Rerecord growth conditions only. (#) Anhydrobiosis.

Environmental factor	Growth conditions			Extremophile
	Minimum	maximum	Optimum	
aridity	-	-	#	Xerophile
pH	pH < 7.0	-	pH 3.0	Acidophile
	pH > 7.0	pH > 9.0	pH < 8.5	Alkaliphile
	pH ~ 7.0	≥ 10	pH > 8.5	Alkalitolerant
Pressure	-	120 Mpa*	-	Piezophile
Radiation	-	< 15 kGy*	-	Radioresistant
Salinity	0.2 M	-	1.5 M	Halophile
	-	2.5 M	> 0.25 M	Halotolerant
Temperature	< 0°C	20°C	15°C	Psychrophile
	< 15°C	25°C	> 15°C	Psychrotolerant
	40°C	> 100°C	80°C	Hyperthermophiles
	< 40°C	80°C	60°C	Thermophile

In the next sections, examples of extremophiles, along with their adaptive strategies, are highlighted, with particular emphasis on those relevant to the current study, namely those extremophiles adapted to extreme cold, high pH and high salt concentration, and their respective adaptation mechanisms.

1.2.1 Psychrophilic (Cold-adapted) microorganisms

Cold environments form the greater proportion of the Earth's biosphere, including most of the world's oceans (which constitute approximately 70 % of the Earth's surface), the polar regions, mountainous regions, as well as man-made cooling systems (Cowan *et al.*, 2007). Thus, cold-adapted organisms might be the most abundant extremophiles on Earth (De Maayer *et al.*, 2014a).

Cold-adapted organisms are classified into two groups, psychrophiles and psychrotrophs, based on their requirements for growth and development (Morita, 1975). Psychrophiles grow optimally at 15°C, and are capable of activity below 0°C with a maximum tolerable temperature of 20°C. The second group, psychrotrophs (also psychrotolerant microorganisms) have optimum growth temperatures greater than 20°C (Cowan *et al.,* 2007). Furthermore, organisms with strict cold requirements that do not tolerate elevated temperature are described as stenopsychrophiles. By contrast, psychrotolerant organisms having wider growth temperature ranges and capable of surviving higher temperatures (usually < 30°C) are termed eurypsychrophiles (Cavicchioli, 2006). Generally, the term ‘psychrophiles’ is used loosely to describe both strict psychrophiles and psychrotolerant organisms (De Maayer *et al.*, 2014a).

It is important to note that there is distinction between the temperature at which an organism survives (limit to survival, T_S) in one form or another (for example, as spores or cysts) and the temperature at which an organism can carry out normal metabolic activities and complete its life cycle (limit to life, T_L) (Figure 1-1).

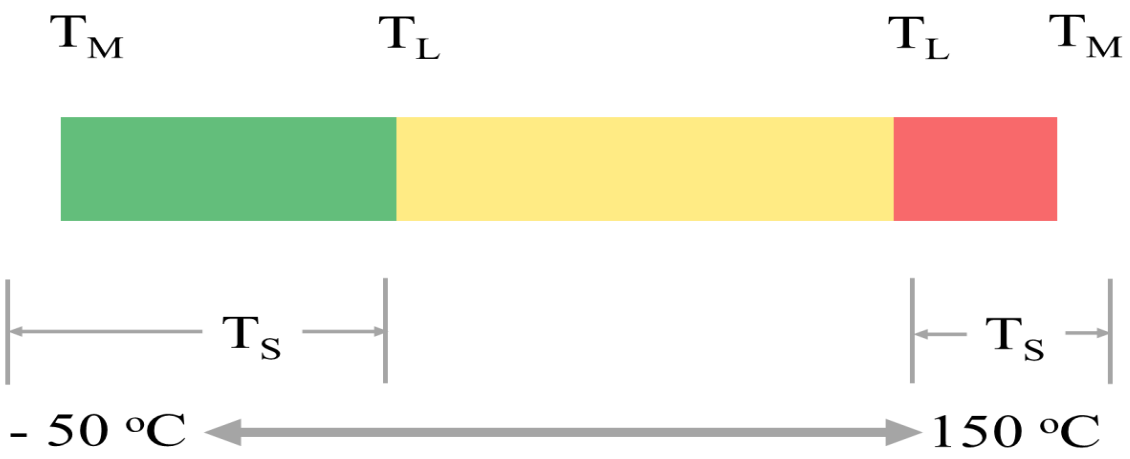


Figure 1-1: Thermal life limits on Earth. TL: temperature limits of completing life cycle; TM: temperature limits of metabolism; TS: temperature limits of survival. The yellow portion represents the temperature window in which an organism could proliferate. The figure was modified from Clarke et al. (2014).

The latter is generally not clearly defined because of limitations associated with environmental measurements (Clarke, 2014). Beyond T_L, an organism may remain viable and continue to metabolise until the temperature limit for metabolism (T_M) is exceeded (Clarke, 2014). The sea-ice bacterium, *Psychromonas ingrahamii*, isolated from the Arctic, has been demonstrated to maintain exponential growth at -12 ºC (T_L) (Riley *et al.*, 2008). Similarly, the psychrotrophic yeast, *Rhodotorulaglutinis*FMT157 has been reported to cause spoilage of peas at -18 ºC (Collins and Buick, 1989). Aside from these examples, *Psychrobacter urativorans*, *Colwellia psychrerythraea* and *Psychrobacter frigidicola* have been reported, on the basis of circumstantial evidence, to thrive at temperatures between -10 and -18 ºC (Chin *et al.*, 2010). However, there is also evidence that biotic activity at temperatures between -20 ºC and -40 ºC may occur (Clarke, 2014, De Maayer *et al.*, 2014a, Stevenson *et al.*, 2015).

Psychrophilic microorganisms, especially those inhabiting natural cold environments, may be subject to the combined effects of reduced water activity, cold and raised solute concentrations. In addition, due to their small cellular size, they are not capable of self-insulation and thus depend on a wide range of molecular and physiological adaptations for their survival (Cowan

et al., 2007, Russell, 2007). Cell membranes of psychrophilic microbes are adapted to cold temperatures through various mechanisms (Russell, 2007). These include thermal regulation of membrane fluidity through the production of hopanoids and adaptations of membrane lipid composition (Russell, 2000, Russell, 2007). Lipid composition modifications include enrichment of branched chain fatty acids, shortening the length of fatty acyl chains and unsaturation of acyl chains (Feller, 2003). Another important mechanism is the regulation of membrane passive permeability to ensure normal transmembrane uptake of dissolved solutes and gases, as well as proton permeability (Russell, 2007). Proton permeability was demonstrated to be dependent on growth temperature and critical for the survival of all microorganisms (van de Vossenberg *et al.*, 1999). Microorganism, therefore, possess molecular mechanisms that ensure constant and low proton permeability values. These operate via the adjustments of the lipid composition of the cytoplasmic membrane and hence the observation that proton permeability remains fairly constant irrespective of thermal adaptations (Van de Vossenberg *et al.*, 1995, van de Vossenberg *et al.*, 1999). The ability to rapidly adjust lipid composition is particularly crucial for psychrotrophic bacteria inhabiting Antarctic soils where wide daily and seasonal temperature fluctuations have been reported (Aislabie *et al.*, 2006, Dreesens *et al.*, 2014).

Passive permeability is largely determined by the composition of membrane lipids and could be regulated by cold shock proteins (CSPs) (D'Amico *et al.*, 2006, Russell, 2007). CSPs also enhance transcription and also function as RNA-binding proteins (Satyanarayana *et al.*, 2005). Microorganisms produce a range of CSPs such as CspA, GroEL and DnaK in response to sudden exposure to low temperature. In psychrophiles, production of CSPs occurs concurrently with that of other housekeeping proteins (Margesin and Miteva, 2011). Furthermore, the expression of CSPs in psychrophiles is constitutive and the rate of expression depends on temperature (D'Amico *et al.*, 2006, Margesin and Miteva, 2011).

Psychrophiles can also produce cryoprotectants and anti-freeze proteins. Cryoprotectants prevent the denaturation of proteins and include exopolysaccharides, polyamines, polyols and amino acids. The anti-freeze proteins are ice-binding proteins that interfere with formation and recrystallization of ice (D'Amico *et al.*, 2006, Margesin and Miteva, 2011). Psychrophiles also produce cold adapted enzyme that are efficiently active at low temperatures but inactivated at higher temperatures (Collins *et al.*, 2007, Feller, 2003).

1.2.2 Halophilic (Salt Adapted) Microorganisms

Saline environments are diverse and widely distributed across the Earth surface. The oceans represent the largest contiguous saline environment. Other saline environments include natural salt lakes, artificial salt lakes, underground deposits of rock salt, saline soils, salted foods, etc. (Oren, 2007). Salt-adapted microorganisms grow at salt concentrations greater than 0.2 M. However, there is distinction between halophilic and halotolerant microbes. Halophiles grow at a minimal salt concentration of 1 M, with an optimal concentration of 1.5 M and no upper limit. Halotolerant microorganisms, on the other hand, have no lower salt concentration limit but require 0.25 - 1.5 M of salt for optimum growth and will not survive at concentrations above 2.5 M (Bowers *et al.*, 2009, Mesbah and Wiegel, 2008).

Salt-adapted organisms are highly diverse and occur across the all domains of life (Archaea, Bacteria and Eukarya) (Ma *et al.*, 2010). The most specialized halophilic microorganisms are found in the archaeal family *Halobacteriaceae* (order *Halobacteriales*). Members of this family are entirely halophilic and proliferate well at salt concentrations as high as 300 g/l (Oren, 2008). Among the bacteria, the anaerobe *Salinibacter ruber* represents the most specialised halophilic bacteria with a minimum salt requirement ≥ 150 g/l. *S. ruber* strains grow optimally at salt concentrations between 200 and 300 g/l (Antón *et al.*, 2002, Oren, 2007, Oren, 2008).

Microorganisms inhabiting saline environments are subject to losing water as a result of the osmotic pressure exerted on their cytoplasm and therefore must adopt strategies to counteract this stress (Mesbah and Wiegel, 2008). Salt-adapted microorganisms employ two main strategies to withstand osmotic stress; the salt-in-cytoplasm strategy and the accumulation of osmotic solutes. The salt-in-cytoplasm strategy involves the influx of inorganic ions (K^+, Cl^-) into the cytoplasm to attain internal salt concentrations equivalent to those of the external environment. This strategy is mainly employed by extreme halophiles (Oren, 2008). The salt-in strategy is coupled with adaptation of enzymes for optimum activity in the highly saline intracellular environment (Oren, 2008). The proteins in these organisms are tightly packed to prevent protein damage by salt (Hoff, 2009, Tadeo *et al.*, 2009). The proteomes of extreme halophiles show an overrepresentation of hydrophilic amino acid residues with a large number of aspartic acids and glutamic acid residues, and also substitution of lysine by arginine (Hoff, 2009). This specialisation, however, precludes survival of extreme halophiles at low salt concentrations (Hoff, 2009, Oren, 2008).

The second strategy, accumulation of osmotic solutes, is nearly universal among all halophiles, and involves the accumulation of compatible or osmotic solutes (osmolytes) while maintaining low salt concentrations in the cytoplasm (Oren, 2008). Accumulation of these osmolytes occurs either through biosynthesis or their uptake from the surrounding environment and is induced by high salt stress at both the genetic and protein levels (Averhoff and Müller, 2010). The solutes are water soluble organic compounds and include amino acids and their derivatives (e.g. glutamine, glycine, betaine, ectoine and hydroxyectoine) or polyols and sugars (e.g. glycerol, gluco-glycerols and sulfotrehalose (Saum and Müller, 2008). Although accumulation of solutes is energetically costly (Hoffmann *et al.*, 2012, Maskow and Babel, 2001), the strategy is widespread among archaea and bacteria, partly because osmolytes do not interfere with the metabolic activities of the organism even at high concentrations (Roeßler and Müller, 2001).

Additionally, the accumulation of osmolytes allows the organism to tolerate wider range of salt concentrations (Oren, 2008, Roeßler and Müller, 2001).

1.2.3 Alkaliphiles (High pH-Adapted) Microorganisms

Alkaliphiles are microorganisms adapted to high pH conditions. These organisms are adapted to a wide range of habitats. They occur naturally in alkaline hot springs, natural soda lakes, different kind of soils and in deep-sea trenches (Krulwich and Ito, 2013, Yumoto *et al.*, 2011). They are also often found associated with industrial activities such as indigo dye processing, the pulp and paper industries, the cement industry and food processing (Krulwich and Ito, 2013, Yumoto *et al.*, 2011). Alkaliphilic extremophiles are broadly categorised into two classes (Mesbah and Wiegel, 2008). The alkalitolerant organisms can survive slightly acidic to neutral pH and have a pH optimum and maximum of < 8.5 and > 9.0, respectively. Strict alkaliphilic organisms have a pH optimum of ~ 8.5. The minimum pH requirement is > 7.5 and the upper pH limit is ≥ 10.0. The latter group of organisms is further sub-divided as obligate (those that cannot thrive in pH lower than 8) and facultative (capable of surviving at near neutral pH) alkaliphiles (Yumoto, 2007). For example, the obligate alkaliphile *Alkalibacterium psychrotolerans* does not grow at pH < 9.0 (Yumoto *et al.*, 2004). *Alkaliphilus transvaalensis* isolated from South African gold mines also grows actively at a wide pH range between 8.5 and 12.5 (Takai *et al.*, 2001). The facultative alkaliphile *Clostridium thermoalcaliphilum* sp. has a pH optimum of between 9.6 and 10.1, but also grows well at a pH between 7.0 and 11.0 (Li *et al.*, 1994).

To survive in high pH environments, alkaliphiles must be able to maintain a cytoplasmic pH lower than that of the external environment (Horikoshi, 1999, Krulwich *et al.*, 2011). Alkaliphiles employ several structural and molecular adaptations to cope with this challenge. The peptidoglycan component of the cell wall in alkaliphilic bacteria contains acidic polymers

with negatively charged residues like aspartic, galacturonic and gluconic acid which enable the adsorption of Na^+ and H^+ and repulsion of OH^-, thereby lowering the pH at the cell surface (Horikoshi, 1999). The plasma membranes in alkaliphiles use an Na^+/H^+ antiporter system (both transmembrane pH and electrochemical gradient-dependent) as a key mechanism in internal pH homeostasis (Padan *et al.*, 2005). This involves the exchange of sodium from the cytoplasm into the surrounding with protons through the antiporters. However, sodium is also required for proper metabolism and internal Na^+ balance is maintained by the combined actions of symporter proteins and the flagellar motor (Horikoshi, 1999, Padan *et al.*, 2005, Satyanarayana *et al.*, 2005).

1.2.4 Industrial applications of extremophiles

Extremophiles have developed various means to survive or even thrive under extreme conditions. These include the production of unique biomolecules that are capable of carrying out complex biological processes under extreme environmental conditions (Niehaus *et al.*, 1999). A number of extremophile enzymes have found a wide range of applications in various industries (Karan *et al.*, 2012). Recent advances in next-generation sequencing technology have resulted in the discovery of a variety of novel bioactive products for various industrial applications (De Maayer *et al.*, 2014c). This has been enhanced by the development and application of novel metagenomics strategies (Cowan *et al.*, 2015) for bioactive product discovery from extreme environments (Barone *et al.*, 2014, Dalmaso *et al.*, 2015, López-López *et al.*, 2014). Key extremophile molecules currently in use include various enzymes, cryoprotectants and lipids (Podar and Reysenbach, 2006). Furthermore extremophile themselves have been incorporated into certain whole-cell industrial bioprocesses including bio-mining of metals like uranium and gold (Elleuche *et al.*, 2014, Podar and Reysenbach, 2006).

In recent years, there has been particular interest in cold-stable enzymes produced by psychrophilic organisms. Because the bioactive products derived from psychrophiles are adapted to function optimally at low temperatures (Cavicchioli *et al.*, 2002, Struvay and Feller, 2012), these 'psychrozymes' could serve in the energy-efficient production of raw materials in the detergent, pulp and paper, food and textile industries (Adrio and Demain, 2014). Enzymes and other bioproducts obtained from halophiles have potential applications in a wide range of industries including food, leather, pharmaceuticals and textiles (de Lourdes Moreno *et al.*, 2013). Alkaliphiles also produce enzymes with wide range of applications, particularly as additives in detergents or bio-detergents (Neifar *et al.*, 2015). Alkali-stable enzymes have been used in the production foods and pharmaceuticals and also in the paper and pulp industries (Horikoshi, 1999).

1.3 Genomics and 'omic technologies

1.3.1 Genomics

The biological functions within an organism are to a large extent determined by their genetic make-up. Extremophiles, for example, carry genes coding for mechanisms that allow them to survive under extreme conditions (Clarke, 2013). By deciphering the nucleotide sequences of genes in these organisms, one can study the molecular mechanisms underlying various genotypes and phenotypes (Oh *et al.*, 2012). This has been greatly facilitated by the development of genome sequencing technologies. Since the first complete genome sequence of *Haemophilus influenzae* became available in 1995 (Hutchison, 2007, Fleischmann *et al.*, 1995), the potential of genome sequences as a resource for gaining an understanding of the biology of organisms and to assess various phenotypes, such as pathogenicity, symbiosis and metabolism, has been effectively exploited (Henson *et al.*, 2012). Recently, new sequencing technologies, both next-generation sequencing (NGS) and third generation (TGS) technologies,

have been developed which have greatly advanced genome sequencing, making large-scale genome-sequencing feasible and cost-effective (van Dijk *et al.*, 2014). As of April 2015, 63,008 genome projects have been registered on the Genomes Online Database (Figure 1-2).

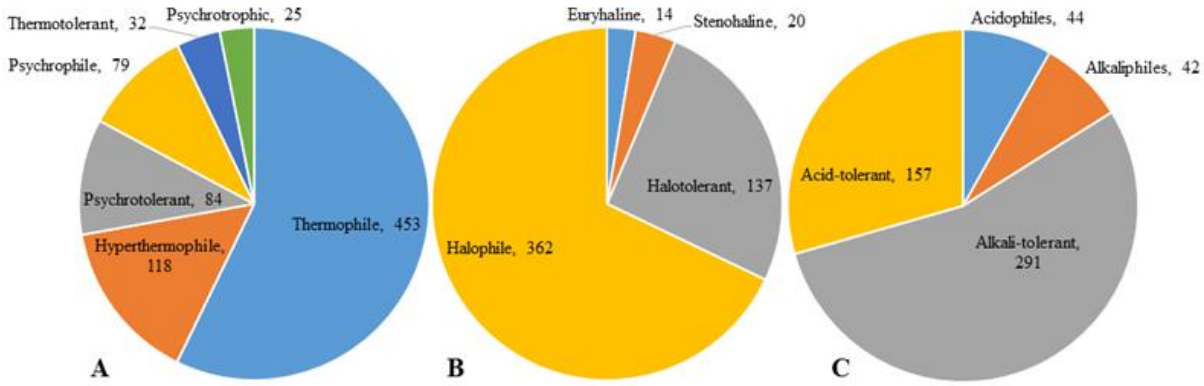


Figure 1-2 GOLD genome projects statistics on the number of projects related to extremophilic organisms. A, B and C represents the number of registered projects linked to temperature, salinity and pH adaptations.

These include 6,652 complete genome sequences, while a further 23,543 genomes have been sequenced to high quality draft status (Reddy *et al.*, 2015). The impact of genome sequencing and in particular NGS technologies has been enormous. For example, these technologies have been applied in microbial diversity exploration, ancient DNA studies, elucidation of genomes of bacteria and viruses, mutation and polymorphism studies and functional genomics (Marguerat *et al.*, 2008, Morozova and Marra, 2008).

1.3.1.1 Sequencing technologies

Sequencing technologies are variously described as first generation, next generation (NGS) and third generation sequencing technologies (Morey *et al.*, 2013). The advent of modern day sequencing was triggered by the introduction of Sanger sequencing chemistry in the late 1970s (Sanger *et al.*, 1977).

Sanger's automated sequencing method is the most reliable of the first generation methods and produces long (up to 1,000 bp) and highly accurate reads (Shendure and Ji, 2008). Using this technology, the first complete human genome sequence was generated at a staggering cost of over $1 billion, taking ten years to complete (Human Genome Sequencing, 2004, Metzker, 2010). Major drawbacks of the first generation Sanger sequencing technology are the low throughput (< 96 kilobases per 3 hour run compared to up to 170 gigabases per 3 hours in the most recent NGS technologies) and the high cost when applied in large-scale projects such as whole genome sequencing (Perkel, 2009).

Second generation sequencing technology, generally referred to as NGS technologies, generally involve fragmenting of genomic DNA and subsequent sequencing by one of two methods, sequencing-by-synthesis and sequencing-by-ligation (Mardis, 2008, Niedringhaus *et al.*, 2011, Zhang *et al.*, 2011). In sequencing-by-synthesis, the fragmented DNA is ligated to adapter sequences. This is followed by immobilisation of the fragments onto beads or solid surfaces and PCR amplification. Depending on the chemistry involved, a stepwise incorporation of nucleotides occurs by a reversible terminator polymerase reaction (e.g. in the Illumina platforms) or pyrosequencing (e.g. in the Roche 454 platforms) and these nucleotides are subsequently detected by means of fluorescence or electro-chemical signals (Niedringhaus *et al.*, 2011, Shokralla *et al.*, 2012). In sequencing by ligation, DNA ligase replaces the DNA polymerase. Oligonucleotide probes of specific sizes, each labelled with fluorescent tags, are selectively hybridised to complimentary genomic templates in presence of DNA ligase. This is

accompanied by fluorescence imaging of the ligated probe and chemical cleavage. Non-ligated probes are then washed away (Metzker, 2010). The method was applied in the DNA Nanoball, Polonator and SOLiD platforms (Morey *et al.*, 2013).

The NGS protocols offer massively parallel sequencing which generates a much greater throughput than first generation methods, at a fraction of both the time and economic costs associated with first generation sequencing (Hutchison, 2007). For example, the Illumina HiSeq 2000 can generate up to 3 billion reads and yield 600 Gigabases per run at an unprecedented cost of $0.03 per million bases compared to Sanger's 3730xl 96 reads/run, yield of 84 Kilobases and at the cost of $2400 per million bases (Liu *et al.*, 2012). At present, the high throughput achieved with NGS technologies is at the expense of read length (50 - 500 bases), but this is compensated for by the high degree of sequence coverage (Hutchison, 2007, Zhang *et al.*, 2011).

The present era of sequencing is an interface of the accomplishments of the NGS chemistries and the promise of the third generation technologies. Currently, two technologies at different stages of development are at the forefront of the third generation revolution; single molecule real time (SMRT) sequencing (first incorporated in the Pacific Biosciences Pacbio platform; http://www.pacificbiosciences.com/) and nanopore sequencing technology (incorporated in the Oxford Nanopore MinION and GridION sequencing platforms; https://www.nanoporetech.com/) and may offer great improvement in read length, cost and time of sequencing run (Land *et al.*, 2015, Reuter *et al.*, 2015).

The SMRT chemistry in PacBio RS II machine produces up to 40 Kb reads and, according to the manufacturer, approximately 50 % of the reads are > 14 Kb in length, although the read length is affected by the sequencing library chemistry. Experimental data has, however, revealed an average read length of between 5.5 and 8.5 Kb at a maximum cost of 150 US dollars per Megabase and maximum of three hours per run of 0.05 million reads (Carneiro *et*

al., 2012, Kremkow and Lee, 2015). The portable MINION™ machine recently commercialised by Oxford Nanopore (Quick *et al.*, 2014) has been reported to produce reads that range between 6 and 60 Kb (Ashton *et al.*, 2014, Jain *et al.*, 2015).

These new techniques are currently disadvantaged by high levels of sequencing error (Reuter *et al.*, 2015). For instance, overall error rates of approximately 18 % have been observed when sequencing using the MINION™ platform (Reuter *et al.*, 2015, Quick *et al.*, 2014) and PacBio RSII has an error rate of approximately 15 % (Carneiro *et al.*, 2012, Kremkow and Lee, 2015).

1.3.1.2 Genome assembly

Genome assembly is the process of developing correctly ordered and properly oriented long contiguous sequences (contigs) from overlapping sequencing reads and ordering of these contigs into scaffolds (Earl *et al.*, 2011, Edwards and Holt, 2013). The ultimate objective is to reproduce a complete nucleotide sequence identical to the original genomic DNA template from which the reads were produced (Paszkiewicz and Studholme, 2010). Depending on the availability of reference genomes, genome assembly can be undertaken using *de novo* or reference-based assembly methodologies (Zhang *et al.*, 2011).

Two types of algorithms for the *de novo* assembly of sequencing reads have been developed and are widely applied for the assembly of reads into contigs. These are the Overlap Layout Consensus (OLC) and De Bruijn Graph (DBG) methods (Schatz *et al.*, 2010). In the former, overlaps are scored between all possible pair of reads to form a consensus layout. A long string is formed in which reads represent vertices (nodes) connected by edges which represents overlaps and a Hamiltonian circuit is constructed to resolve the graph (Paszkiewicz and Studholme, 2010). The major drawback of the OLC method is the difficulty associated with the computational resolution of the generated graphs which are complex and require expensive computational resources to resolve the graphs (Georganas *et al.*, 2014). The De Bruijn Graph

method solves this problem by finding a path for which a cycle touches each edge only once following the shortest possible path (Eulerian path) across all edges (Compeau *et al.*, 2011). The reads are split into a number of base substrings called k-mers from which a continuous string is formed for which the K-mers overlap by k-1 bases (Compeau *et al.*, 2011). These algorithms have been incorporated into a number of different assembly software packages. For example, ABySS (Simpson *et al.*, 2009), AllPath (Butler *et al.*, 2008), Euler (Chaisson and Pevzner, 2008) and Velvet (Zerbino and Birney, 2008) all employ DBG algorithms, while the Celera Assembler (Myers *et al.*, 2000), Edena (Hernandez *et al.*, 2008) and Newbler (Roche Diagnostic Corporation, Switzerland) apply OLC algorithms. The application of any particular algorithm depends on the nature of the reads. DBG algorithms have been successfully applied to produce good assemblies from short NGS sequence reads, while OLC methods are efficient with all form of reads including those produced through Sanger technology (Li *et al.*, 2010, Schatz *et al.*, 2010).

Many methods are used to verify the quality of an assembly. These measure quality with reference to accuracy, contiguity, computing time and memory usage, coverage and assembly error rate (Lin *et al.*, 2011, Paszkiewicz and Studholme, 2010). Euler-SR (Chaisson and Pevzner, 2008) and AllPath (Butler *et al.*, 2008) correct errors prior to assembly while Velvet (Zerbino and Birney, 2008) and Fragment Gluer (Pevzner *et al.*, 2004) edit the graph produced during the assembly process. Commonly used quality metrics include metrics of length such as minimum, maximum, median and, most notably, N50. N50 is computed by first arranging all contigs by length and then summing up from the longest down until the total is equal to or greater than 50 % of the length of the genome (or total length of contigs). The length of the contig on which this value is attained is taken as N50. Another method of confirming assembly accuracy is comparison to a reference genome of a closely related organism (Earl *et al.*, 2011). Assembled contigs are ordered against reference genomes, which are genetically closely

related to the sequenced genome, using programs such as Artemis comparison tool (ACT) and Mauve (Darling *et al.*, 2010, Edwards and Holt, 2013).

Even with increasing read lengths achieved with novel NGS chemistries, repeat regions continue to present a major problem when it comes to genome assembly. To address this issue, gaps between contigs in scaffolds can be closed by PCR amplification and Sanger sequencing (Nagarajan *et al.*, 2010) to produce complete or high quality draft genome sequences. Recent advances in the development of third generation sequencing chemistries hold the potential for circumventing the genome finishing stages by producing reads long enough to produce complete and accurate bacterial genome sequences (Land *et al.*, 2015).

1.3.1.3 Genome annotation

Annotation involves the delineation and identification of genes in the assembled genome and assigning functions to these genes (Pareja-Tobes *et al.*, 2012, Zhang *et al.*, 2011). Annotation is accomplished through the use of bioinformatic tools which work on the basis of one of two approaches: homology-based search or algorithm-based prediction (Logares *et al.*, 2012), relying on the use of annotated reference genomes (Richardson and Watson, 2013) and *de novo* gene predictions, respectively (Edwards and Holt, 2013, Stothard and Wishart, 2006). Once gene prediction is completed, the resultant coding sequences are searched against specific reference genomes or genetic or genomic databases using various available computational tools (Richardson and Watson, 2013). Predicted coding sequences which significantly differ from those deposited in the databases are considered as putative novel genes (Stothard and Wishart, 2006).

A wide range of software packages and web-servers for genome annotation are available (Kisand and Lettieri, 2013, Logares *et al.*, 2012, Richardson and Watson, 2013). Some of the commonly used annotation pipelines include the Rapid Annotation using Subsystem

Technology (RAST) server (Overbeek *et al.*, 2014), the NCBI based Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (Tatusova *et al.*, 2013) and the Bacterial Annotation System (BASYS) (Van Domselaar *et al.*, 2005).

1.3.2 Comparative genomics

Genome wide comparative analysis of the protein encoding genes, the corresponding proteins and other genetic features in a group of microorganisms is a powerful strategy for elucidating the determinants of adaptation and diversification of the organisms in different environments (MacLean *et al.*, 2009, Qin *et al.*, 2014b). Comparing genomes of closely related organisms may reveal the genetic differences which underlie different adaptation strategies and can be used to ascertain the phylogenetic history of the compared organisms (Abby and Daubin, 2007, Alcaraz *et al.*, 2010, Sivashankari and Shanmughavel, 2007). Several automated web-tools and stand-alone programme have been developed recently for this purpose (Guimarães *et al.*, 2015). Prominent among these are PanGP (Zhao *et al.*, 2014), EDGAR (Blom *et al.*, 2009), PanCake (Ernst and Rahmann, 2013) and PGAT (Brittnacher *et al.*, 2011).

An important approach in the whole-genome comparison of closely related organisms is the determination of the gene compliment encoded by all the compared organisms; the pan-genome of the group (Guimarães *et al.*, 2015). The pan-genome is typically categorised into the core and the accessory genomes (Tettelin *et al.*, 2005, Vernikos *et al.*, 2015). The core genome comprises the orthologous gene sets that are conserved in all members of a taxonomic group. In contrast, the accessory (also dispensable or adaptive) genome consists of the genes/proteins that are not common to all members. These include singletons, which are genes or protein coding sequences unique to a single organism among those strains compared, as well as those that are shared by some, but not all, of the strains included in the analysis (Lapierre and Gogarten, 2009, Medini *et al.*, 2005).

Whereas the core genome is thought to encompass the minimum gene sets required for the survival of any member of the group, the accessory genome constitutes the gene sets associated with specialised survival strategies (niche specialization) exclusive to the individuals carrying the genes (Medini *et al.*, 2005, Mira *et al.*, 2010, Polz *et al.*, 2013). Genes making up the core genome have most likely originated via speciation (i.e. vertically acquired), while the genes of the accessory fraction are subjected to either gene loss or lateral acquisition (Mira *et al.*, 2001, Tettelin *et al.*, 2008). Consequently, to understand genetic imprints of adaptation in microorganisms, both the core and accessory genomes are assessed. Differences in adaptations across temperature, salinity or pH spectra among microorganisms can potentially be inferred based on presence or absence of genes and other genetic features in both the accessory and core genomes (De Maayer *et al.*, 2014b).

To compile a pan-genome, the orthologous relationships (Gabaldón and Koonin, 2013, Koonin, 2005) of all the genes/proteins that are encoded on the genomes of all the organisms included in the study are identified. Accurate determination of the orthologous relationships is an essential step in comparative genomics (Kristensen *et al.*, 2011). Sets of genes are assumed to be orthologs if they originated vertically from a common ancestor (Koonin, 2005). Some of the widely applied methods for the determination of orthologous genes include heuristic best-match (HBM), phylogenetic tree-based (PTB) and synteny methods (Kristensen *et al.*, 2011). In the synteny method, orthologous relationships are assigned by identifying conserved gene order in the genomes of organisms under consideration (Koonin, 2009). This co-localisation of genes in closely related organism is seen as one of the major signals for common ancestry (Koonin, 2009). In the PTB method, orthology is inferred via the reconciliation of gene family trees with a species tree (Lechner *et al.*, 2014). The tree-based methods is disadvantaged by high computational cost while the synteny approach is susceptible to incidences of horizontal gene transfer, gene duplication and loss, especially in prokaryotes (Lechner *et al.*, 2014).

The most widely applied method is the HBM, which relies on the computation of sequence similarity for the assignment of orthologs (Koonin, 2009, Lechner *et al.*, 2014). Similarity between sequences has been best estimated using approaches such as reciprocal best blast hit (RBBH)(Moreno-Hagelsieb and Latimer, 2008). Variants of the HBM have been implemented in many methods developed for the identification of orthologous relationships. The most widely used methods include COG (Tatusov *et al.*, 2000, Tatusov *et al.*, 2001, Tatusov *et al.*, 2003), eggNOG (Powell *et al.*, 2013, Powell *et al.*, 2012), OrthoMCL (Fischer *et al.*, 2011, Li *et al.*, 2003) and Proteinortho (Lechner *et al.*, 2011). These heuristic approaches identify orthologs with the same efficiency as the tree based methods, but using less computational resources (Kristensen *et al.*, 2011). The heuristic method is, however, inefficient at distinguishing between orthologs (which evolved by vertical descent) and paralogs (which evolved by duplication) (Koonin, 2005, Kristensen *et al.*, 2011).

1.3.3 **'Omic' technologies**

The power of genomics for studying the biology of microorganisms has been strengthened by the development of complementary 'omics' techniques. These techniques have also been used, in combination with genome sequencing, to study the biology of extremophiles, and for the discovery and analyses of their bioactive products (Handley and Lloyd, 2013). Pertinent 'omics' technologies include proteomics, metabolomics, fluxomics and transcriptomics (Dalmaso *et al.*, 2015).

1.3.3.1 Transcriptomics

The transcriptome of an organism represents the total transcripts produced under prevailing conditions or specific stages of development (Kogenaru *et al.*, 2012, Wang *et al.*, 2009). The compilation of the transcriptome profiles has facilitated the efficient evaluation of the correlation between genotype and phenotype in many organisms (Gillings and Westoby, 2014). For example, transcriptome analyses have been used to study the dynamics of life processes through the evaluation of gene expression under different environmental conditions (Payne, 2015, Pinto *et al.*, 2011). Furthermore, transcriptomics has also been applied to improve genome annotation, including delineating the transcription start sites and the regulatory (intra and intergenic) elements (Raghavan *et al.*, 2011) and for the discovery of novel genes (Mazin *et al.*, 2014)

Previously, two limitations hindered analyses of the transcriptomes of prokaryotes, the general instability of the transcripts and lack of poly (A) tail which is utilized for the immobilization of mRNA during enrichment and reverse transcription (Siezen *et al.*, 2010). Recent advances in this field, including new strategies to minimize degradation of the extracted transcripts and different priming methods such as the use of oligo (dT) and random hexamers as well as mRNA

specific probes, have brought about improvements in the application of transcriptomics and the deciphering of gene regulation in prokaryotes (Sorek and Cossart, 2010).

Early analyses of the transcriptome of microorganisms relied on DNA microarray chips. Microarrays involve the hybridization of a complementary DNA (cDNA) synthesised from mRNA to oligonucleotide probes that were designed on the basis of known genes in a sequenced genome (Shendure, 2008, Van Vliet, 2010). This approach is, however, limited by low transcript detection efficiency associated with background noise, its reliance on predetermined genomes and challenges of quantification and reproducibility (Shendure, 2008, Siezen *et al.*, 2010). Oligonucleotide arrays technology is now being superseded and replaced by RNA sequencing as the method of choice for transcriptome analysis (Van Vliet, 2010, Mutz *et al.*, 2013).

Essentially, RNA-Seq involves sequencing of cDNA obtained by reverse transcription of the extracted RNA (Sorek and Cossart, 2010). Because rRNA constitutes > 95 % of the total RNA and to enhance coverage, particularly when the experimental objective is to identify differentially expressed genes, rRNA is depleted from total RNA prior to cDNA synthesis (Siezen *et al.*, 2010, Sorek and Cossart, 2010). The reads generated using any of the next generation sequencing platforms (e.g. Illumina, Ion proton, SOLiD and Roche GS-FLX) are then mapped against a reference genome to assess both identity and quantity of the mapped transcripts (Pinto *et al.*, 2011).

Compared with microarrays, RNA-Seq offers a highly efficient and reproducible method of quantifying gene expression in bacteria (Croucher and Thomson, 2010, Haas *et al.*, 2012, Van Vliet, 2010, Nookaew *et al.*, 2012). It has been demonstrated that RNA-seq provides better detection and quantification of transcripts when compared to microarrays (Fu *et al.*, 2009). RNA-Seq methods have facilitated the global transcriptome analysis of many more organism than microarrays by eliminating the requirements for *a priori* knowledge of gene contents

(McGettigan, 2013). This is particularly important for uncovering novel genes, delineation of operons and the characterisation of regulatory regions on the genome (Cho *et al.*, 2013, Croucher and Thomson, 2010, McGettigan, 2013).The major drawback to RNA-Seq still remains the various mRNA enrichment and cDNA synthesis procedures which are capable of introducing bias (Siezen *et al.*, 2010).

1.3.3.2 Proteomics and other 'omics technologies

Transcriptomics approaches have allowed the identification and quantification of the responses of organisms to environmental perturbations (Cho *et al.*, 2013, Croucher and Thomson, 2010, Sorek and Cossart, 2010). Genomic and transcriptomic studies, however, do not take into account post-transcriptional and -translational activities which modify the end products of mRNA translation and perhaps the regulation of cellular function (Payne, 2015, Vogel and Marcotte, 2012). Thus, analysing the whole proteome (proteomics) of an organism has extended the understanding of how organisms interact with their environment (Banks *et al.*, 2000, Chandrasekhar *et al.*, 2014). Proteomics provides insight into the functional roles of each protein and enables the reconstruction of metabolic pathways (Keller and Hettich, 2009). Proteomic analysis also allows the identification of post-transcriptional events, protein interactions and responses of organisms to perturbations (Keller and Hettich, 2009, Wilmes and Bond, 2006). Several methods have been developed for the studies of expression and structure of the protein compliment of organisms (Chandrasekhar *et al.*, 2014, Zhang *et al.*, 2010). These include gel based assays, e.g., "two-dimensional polyacrylamide gel electrophoresis" (2D-PAGE) (May *et al.*, 2012, Rabilloud *et al.*, 2010) and several mass spectrometry (MS) approaches, such as electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) (Catherman *et al.*, 2014, Jungblut, 2014, Zhang *et al.*, 2010). These MS

approaches are often combined with different analysers such as ion trap, time of flight and triple quad analysers (Kelleher, 2004, De Maayer *et al.*, 2014c).

Other important 'omic' approaches with wide range of application in the post-genomic era include, metabolomics, which is a set of techniques which can be used to profile the total metabolite content (metabolome) of an organism (Dettmer and Hammock, 2004, Cascante and Marin, 2008, Cortassa *et al.*, 2015) and fluxomics, which assesses the sum of metabolic fluxes in a cell under specific condition or over a given time (Cascante and Marin, 2008, Cortassa *et al.*, 2015, Lien *et al.*, 2015). These approaches are based on the premise that the ultimate response of organisms to perturbations cannot be determined solely based on genomics, transcriptomics or proteomics (Cascante and Marin, 2008). Differences in the metabolome and/or fluxome are likely to provide a better prediction of the phenotype than transcriptomics and proteomics (Cortassa *et al.*, 2015, Dettmer and Hammock, 2004).

1.4 Extremophiles in the 'omics' era

Given the recent interest in extremophiles, from the perspective of their evolution as well as their production of industrially relevant biomolecules, it is not surprising that extremophile research has fully embraced genomic and other 'omic technologies (Casanueva *et al.*, 2010). As a result a large number of genomes of extremophiles which are adapted to survive under a wide range of environmental extremes have been sequenced (DeLong, 2000, Majhi *et al.*, 2013, Logares *et al.*, 2012). A metadata query search for "salt tolerant" against the GOLD databases, for example, returned 533 organisms for which genome sequences are available, while the genomes of 791 organisms with different thermal adaptations and 898 organisms showing different pH requirements (acidic or alkaline) have also been sequenced (Reddy *et al.*, 2015). Despite the predominance of cold habitats on Earth (Rodrigues and Tiedje, 2008), only ~ 23 % of the organisms matching these metadata criteria were recorded as cold-adapted.

1.4.1 Psychrophiles in the 'omics' era

Several 'omic' strategies have been applied to elucidate the mechanisms of psychrophily in microorganisms. The complete genome of the psychrophilic gammaproteobacterium *Colwellia psychrerythraea* 34H (Methe *et al.*, 2005) is one classic example. The genome sequence revealed cold adaptation strategies such as the modifications of membrane fluidity, synthesis and accumulation of osmolytes, psychrophiles-specific amino acid composition and gene content as well as various modification to metabolic pathways underlying its psychrophilic adaptation (Methe *et al.*, 2005). Similar modifications have been reported in the Antarctic marine bacterium *Pseudoalteromonas haloplanktis* TAC125 (Medigue *et al.*, 2005). The genome of this psychrophile consist of two chromosomes of approximately, 3. 8 Mb in size. The strain exhibited one of the most versatile adaptations to oxidative stress. This is evident by the presence of multiple loci encoding several dioxygenases and the absence, altogether, of metabolic pathways that depend on molybdenum cofactor (Medigue *et al.*, 2005).

Comparison of the genome of the sea-ice bacterium, *Psychromonas ingrahamii* to those of other psychrophiles and mesophiles revealed no significant differences in the gross protein properties such as amino acid composition and length distribution. However, there was a distinct clustering of proteins in the psychrophiles which was attributed to the distribution of hydrophobic residues. There was also relative abundance of asparagine and underrepresentation of oxygen sensitive residues in the psychrophile *P. ingrahamii* (Riley *et al.*, 2008).

Recently, genomic analysis was employed to compare Antarctic and temperate strains (seven each) of *Arthrobacter* spp (Dsouza *et al.*, 2014). The Antarctic strains showed a repertoire of features reported in several other psychrophiles. These include features linked to cold responses, such as cold shock and cold acclimation features; features involved in modulation of membrane fluidity, detoxification of reactive oxygen species and osmo-protection (Dsouza *et al.*, 2014).

Comparison with the temperate strains showed that the majority these features were on their genomes regardless of the thermal preferences of the strains (Dsouza *et al.*, 2015). Similar investigations comparing three Antarctic and nine temperate *Paenibacillus* strains showed that the majority of the stress adaptive mechanisms were common to both groups (Dsouza *et al.*, 2014).

However, comparison of the proteins sets encoded on the genomes of a large number of psychrophilic and mesophilic bacterial species revealed marked differences in the amino acid usage profile between the groups (Metpally and Reddy, 2009). Preference for and avoidance of particular amino acids in the proteins of the psychrophiles were linked to protein level molecular adaptive strategies to cope with cold. For instance, amino acids with propensity to impose conformational restrictions, such as glutamic acid (E) and Leucine (L) were underrepresented in the proteome of psychrophiles. There was a corresponding preference for amino acids such as alanine (A), aspartic acid (D), serine (S) and threonine (T), which conferred the conformational flexibility essential for proper functioning of the proteins under cold condition (Metpally and Reddy, 2009). Similarly, cold adaptation has been linked to differences in amino acid substitution pattern in Antarctic halophilic Archaeon, *Halorubrum lacusprofundi*. Comparison of the Antarctic Archaeon with 12 temperate Haloarchaea revealed the presence of amino acid substitutions in 5,541 out of the 70,589 invariant residues in core proteins. The substitutions were linked to small changes in charge, hydrophobicity and size of the amino acids. For instance, glutamic acid (E) was frequently substituted by aspartic acid (D) or alanine (A) (DasSarma *et al.*, 2013).

Our understanding of the lifestyles of microorganisms surviving under extreme cold condition has been improved by the use of other ‘omics’ technologies, such as transcriptomics and proteomics in combination with genome sequencing (Casanueva *et al.*, 2010). For example, a combination of transcriptomic and proteomic analyses has been used to unravel the thermal

adaptation strategies employed by the psychrotrophic bacterium, *Exiguobacterium antarcticum* B7 which was isolated from Antarctica (Dall'Agnol *et al.*, 2014). The combined application of these strategies, in addition to the genome data, have provided greater insight into the gene content as well as some of the transcriptional and posttranscriptional events that follow growth at different temperatures in the Antarctic strain. The data suggested that four of the six CPSs encoded on the E. *antarcticum* B7 genome were induced under extreme cold. The expression of these and other cold inducible proteins have been associated with preserving transcriptional and translational activities and proper folding of proteins at low temperature (Dall'Agnol *et al.*, 2014). Previous transcriptome analyses of the psychrotrophic bacterium, *Exiguobacterium sibiricum* 255-15 (Rodrigues *et al.*, 2008) which has a similar growth temperature range, showed no differential gene expression of 'temperature stress' genes at 10°C and 28°C. There was, however, differential expression of several genes related to different stress responses when temperature extremes of -2.5°C and 39°C were compared (Rodrigues *et al.*, 2008).

Proteomic analysis was also employed to compare the protein contents of *Pseudoalteromonas haloplanktis* TAC125 grown at 4°C and 18°C using two-dimensional gel electrophoresis (Piette *et al.*, 2011). The results revealed that there was downregulation of a number of cellular functions at the lower temperatures and induction of proteins associated with oxidative stress response at 18°C (Piette *et al.*, 2011). Similarly, metabolomic approaches have been used to study cold acclimation in the Artic rhizobial species *Mesorhizobium* sp. N_{33} (Ghobakhlou *et al.*, 2013). There was overproduction of fatty acids associated with modulation of membrane fluidity at low temperature as well as of metabolites with probable roles as cyroprotectants when the organism was grown at 4°C, relative to 10°C and 21°C (Ghobakhlou *et al.*, 2013).

1.4.2 Halophiles in the 'omics' era

The biology of halophilic microorganisms has been extensively studied using genomic and other 'omic' strategies. For example, the genome sequence of *Salinibacter ruber* (Antón *et al.*, 2002, Oren, 2007, Oren, 2008) revealed adaptive signatures that defined the obligate salt requirement of the bacterium (Mongodin *et al.*, 2005). The genome of *S. ruber* codes for a number of hypersaline adaptations, including pathways for the intracellular accumulation of potassium (via K^+ uptake systems), enrichment of acidic amino acid residues (e.g. asparagine) and underrepresentation of basic amino acid residues (e.g. lysine) in the proteins encoded on the genome (Mongodin *et al.*, 2005). Similar adaptive strategies have been identified in the genomes of many other halophiles (Kennedy *et al.*, 2001, Ng *et al.*, 2000).

Comparative genome analysis has also been used to enhance the understanding of halo-adaptation (Paul *et al.*, 2008). Generally, the genomes of halophiles encode protein with fewer hydrophobic and basic amino acid residues, but enriched for acidic residues. The proteomes are also enriched with amino acids that encourage the formation of coil structures in cellular proteins (Paul *et al.*, 2008). A comparison of the genomes of sixty-six organisms, comprising 24, 27 and 15 haloarchaea, halophilic bacteria and non-halophilic bacteria, respectively, revealed that halophiles have evolved different adaptive strategies (Chen *et al.*, 2014). The haloarchaeal genomes had higher G+C contents when compared to those from halophilic and non-halophilic bacteria.

Furthermore, the mechanisms of adaptation to salt stress in cyanobacteria have been elucidated using various 'omic' strategies (Pade and Hagemann, 2015). Salt stress transcriptome data from both microarrays and RNA-Seq for *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 7002 showed the overexpression of genes coding for proteins associated with accumulation of compatible solutes, transport, general stress response and proteins involved in response to reactive oxygen species (Ludwig and Bryant, 2012, Marin *et al.*, 2004). Microarray studies of

Bacillus sp. N16-5 grown under high salt concentrations (8 % and 15 % NaCl) also revealed overexpression of genes coding for chaperones and general stress response proteins (Yin *et al.*, 2015). RNA-Seq data from *Synechococcus* sp. PCC 6803, however, revealed that the non-coding RNAs (ncRNAs) were expressed at equal levels to mRNA (Mitschke *et al.*, 2011), which indicates that ncRNAs also play significant regulatory roles and may direct posttranscriptional activities under stress conditions (Eddy, 2001). RNA-Seq analysis of the halophilic bacterium, *Tetragenococcus halophilus* also revealed a number of salt active/responsive genes with potential novelty in addition to those coding for well-characterised proteins (Liu *et al.*, 2015).

A combined transcriptomic and proteomic analysis of salt stress response in *Bacillus licheniformis* grown at high salt concentrations (1 M NaCl) showed no correlation between levels of gene expression and fold change in the corresponding protein. The overall result did, however, reveal an upregulation of genes and proteins associated with the accumulation of compatible solutes, response to oxidative stress and general stress response (Schroeter *et al.*, 2013). Proteomic analysis of the halotolerant alpha-proteobacterium *Tistlia consotensis* grown under high salt concentration (4.0 % NaCl) also showed that there was an upregulation of proteins associated with compatible solutes transport. Conversely, under low salt concentration (0.5 % NaCl), there was predominance of proteins associated with oxidative stress and general stress response (Rubiano-Labrador *et al.*, 2014).

Metabolomics has also been applied to investigate salt stress response in halophilic bacteria. Analysis of metabolites from five halophilic bacteria showed that certain metabolites increase in response to increasing salt concentration. The compositions of these metabolites were highly dependent on the individual species and the types of growth media (Joghee and Jayaraman, 2014).

1.4.3 Alkaliphiles in the 'omics' era

The ability of microorganism to adapt to alkaline conditions has been studied using various genomic, comparative genomic and other post-genomic 'omic' approaches (Janto *et al.*, 2011, Padan *et al.*, 2005). For instance, the alkaliphile *Bacillus halodurans* C-125 revealed several membrane adaptations that are essential for growth under alkaline conditions (Takami *et al.*, 2000). These include five genes encoding putative Na^+/H^+ antiporters and several genes associated with ABC transporters. The genome also encoded an acidic polymer, teichuronopeptide, which has been implicated in alkaliphilicity (Padan *et al.*, 2005, Takami *et al.*, 2000). The alkaliphilic adaptation of *Bacillus pseudofirmus* OF4 has also been linked to the presence of the Na^+/H^+ antiporters system that supports pH homeostasis in response to increased pH (Janto *et al.*, 2011).

The transcriptome analyses of three haloarchaea subjected to extreme alkaline relative to neutral pH revealed the expression of several genes encoding proteins involved in central metabolic pathways, universal stress proteins and other chaperones common to the three archaea (Moran-Reyna and Coker, 2014). Similarly, transcriptome profiles of *Bacillus subtilis* AG174 grown at pH 9 (relative to pH 6 and 7) showed that there was upregulation of genes involved in central metabolic pathways including arginine catabolism and cytochrome reductases, in addition to regulons for sigma factors (Wilks *et al.*, 2009). Cytoplasmic and membrane proteome profiles of *Corynebacterium glutamicum* ATCC 13032 subjected to different pH conditions have been reported. In the membrane fraction, proteins of central metabolic processes such as succinate dehydrogenase A and B, and ATP synthetase were identified to be upregulated when the organism was grown at pH 9.0 (Barriuso-Iglesias *et al.*, 2008).

1.5 Conclusions

Extreme environments, taken together, constitute a substantial portion of the Earth's biosphere. It is evident that the microorganisms inhabiting these environments play crucial ecosystem functions as drivers of biogeochemical cycles. The survival of organisms under different environmental extremes is consistently linked to the evolution of specialised adaptation strategies. These specialised features have been exploited for the development of bio-industrial applications, either through direct utilization of the extremophiles in various industrial processes or via bioprospecting of bioactive products derived from them. The exploration of extremophiles has been enhanced, greatly, by the advent of next generation sequencing technologies and recent advances in various 'omic' technologies. The availability (cost-effectiveness) of these techniques, coupled with improvements of computational and bioinformatic capabilities, have yielded huge data sets which have undoubtedly expanded the understanding of the mechanistic basis of extremophilic adaptations. Given the enormous variations observed in the lifestyles of extremophiles, it seems reasonable to hypothesise that studying additional extremophilic genomes may provide new insights into the determinants of extremophily and, simultaneously facilitates the discovery of novel bioactive products.

1.6 References

ABBY, S. & DAUBIN, V. 2007. Comparative genomics and the evolution of prokaryotes. *Trends Microbiol.,* 15**,** 135-141.

ADRIO, J. L. & DEMAIN, A. L. 2014. Microbial enzymes: Tools for biotechnological processes. *Biomolecules,* 4**,** 117-139.

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 Biol. Biochem.,* 38**,** 3041-3056.

ALCARAZ, L. D., MORENO-HAGELSIEB, G., EGUIARTE, L. E., SOUZA, V., HERRERA-ESTRELLA, L. & OLMEDO, G. 2010. Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics,* 11**,** 332.

ANTÓN, J., OREN, A., BENLLOCH, S., RODRÍGUEZ-VALERA, F., AMANN, R. & ROSSELLÓ-MORA, R. 2002. *Salinibacter ruber* gen. nov., sp. nov., a novel, extremely halophilic member of the Bacteria from saltern crystallizer ponds. *Int. J. Syst. Evol. Microbiol.,* 52**,** 485-491.

ASHTON, P. M., NAIR, S., DALLMAN, T., RUBINO, S., RABSCH, W., MWAIGWISYA, S., WAIN, J. & O'GRADY, J. 2014. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat. Biotechnol.,* 33**,** 296-300.

AVERHOFF, B. & MÜLLER, V. 2010. Exploring research frontiers in microbiology: recent advances in halophilic and thermophilic extremophiles. *Res. Microbiol.,* 161**,** 506-514.

BANKS, R. E., DUNN, M. J., HOCHSTRASSER, D. F., SANCHEZ, J.-C., BLACKSTOCK, W., PAPPIN, D. J. & SELBY, P. J. 2000. Proteomics: new perspectives, new biomedical opportunities. *Lancet,* 356**,** 1749-1756.

BARONE, R., DE SANTI, C., PALMA ESPOSITO, F., TEDESCO, P., GALATI, F., VISONE, M., DI SCALA, A. & DE PASCALE, D. 2014. Marine metagenomics, a valuable tool for enzymes and bioactive compounds discovery. *Front Mar Sci,* 1**,** 38.

BARRIUSO-IGLESIAS, M., SCHLUESENER, D., BARREIRO, C., POETSCH, A. & MARTÍN, J. F. 2008. Response of the cytoplasmic and membrane proteome of *Corynebacterium glutamicum* ATCC 13032 to pH changes. *BMC Microbiol.,* 8**,** 225.

BLOM, J., ALBAUM, S. P., DOPPMEIER, D., PÜHLER, A., VORHÖLTER, F.-J., ZAKRZEWSKI, M. & GOESMANN, A. 2009. EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics,* 10**,** 154.

BOWERS, K., MESBAH, N. & WIEGEL, J. 2009. Biodiversity of poly-extremophilic Bacteria: Does combining the extremes of high salt, alkaline pH and elevated temperature approach a physico-chemical boundary for life? *Saline Systems,* 5**,** 1-8.

BRITTNACHER, M. J., FONG, C., HAYDEN, H., JACOBS, M., RADEY, M. & ROHMER, L. 2011. PGAT: a multistrain analysis resource for microbial genomes. *Bioinformatics,* 27**,** 2429-2430.

BROCK, T. D. & FREEZE, H. 1969. *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *J. Bacteriol.,* 98**,** 289-297.

BUTLER, J., MACCALLUM, I., KLEBER, M., SHLYAKHTER, I. A., BELMONTE, M. K., LANDER, E. S., NUSBAUM, C. & JAFFE, D. B. 2008. ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Res.,* 18**,** 810-20.

CANGANELLA, F. & WIEGEL, J. 2011. Extremophiles: from abyssal to terrestrial ecosystems and possibly beyond. *Naturwissenschaften,* 98**,** 253-279.

CANGANELLA, F. & WIEGEL, J. 2014. Anaerobic Thermophiles. *Life,* 4**,** 77-104.

CARNEIRO, M. O., RUSS, C., ROSS, M. G., GABRIEL, S. B., NUSBAUM, C. & DEPRISTO, M. A. 2012. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics,* 13**,** 375-375.

CASANUEVA, A., TUFFIN, M., CARY, C. & COWAN, D. A. 2010. Molecular adaptations to psychrophily: the impact of 'omic'technologies. *Trends Microbiol.,* 18**,** 374-381.

CASCANTE, M. & MARIN, S. 2008. Metabolomics and fluxomics approaches. *Essays Biochem,* 45**,** 67-82.

CATHERMAN, A. D., SKINNER, O. S. & KELLEHER, N. L. 2014. Top down proteomics: facts and perspectives. *Biochem. Biophys. Res. Commun.,* 445**,** 683-693.

CAVICCHIOLI, R. 2006. Cold-adapted archaea. *Nat Rev Micro,* 4**,** 331-343.

CAVICCHIOLI, R., SIDDIQUI, K., ANDREWS, D. & SOWERS, K. 2002. Low-temperature extremophiles and their applications. *Curr. Opin. Biotechnol.,* 13**,** 253 - 261.

CHAISSON, M. J. & PEVZNER, P. A. 2008. Short read fragment assembly of bacterial genomes. *Genome Res.,* 18**,** 324-30.

CHANDRASEKHAR, K., DILEEP, A. & LEBONAH, D. E. 2014. A short review on proteomics and its applications. *Int lett nat sci,* 12**,** 77-84.

CHEN, S., YANG, J., LIU, Y., WANG, C. & YANG, Z. 2014. Comparative Genomic Analysis of Halophiles Reveals New Clues to their Adaptation Strategies in Hypersaline Environments. *Austin J Proteomics Bioinform & Genomics,* 1**,** 11.

CHIN, J. P., MEGAW, J., MAGILL, C. L., NOWOTARSKI, K., WILLIAMS, J. P., BHAGANNA, P., LINTON, M., PATTERSON, M. F., UNDERWOOD, G. J. C., MSWAKA, A. Y. & HALLSWORTH, J. E. 2010. Solutes determine the temperature windows for microbial survival and growth. *Proc Natl Acad Sci U S A,* 107**,** 7835-7840.

CHO, S., CHO, Y., LEE, S., KIM, J., YUM, H., KIM, S. C. & CHO, B.-K. 2013. Current Challenges in Bacterial Transcriptomics. *Genomics Inform,* 11**,** 76-82.

CLARKE, A. 2013. Life at extremes: environments, organisms and strategies for survival *In:* BELL, E. (ed.) *Antarctic Science* Wallingford: CABI, pp. 470-470.

CLARKE, A. 2014. The thermal limits to life on Earth. *Int J Astrobiology,* 13**,** 141-154.

COLLINS, M. A. & BUICK, R. K. 1989. Effect of temperature on the spoilage of stored peas by *Rhodotorula glutinis*. *Food Microbiol.,* 6**,** 135-141.

COLLINS, T., D'AMICO, S., MARX, J., FELLER, G. & GERDAY, C. 2007. Cold-adapted enzymes. *In:* GERDAY, C. & GLANSDORFF, N. (eds.) *Physiology and biochemistry of extremophiles.* Washington DC: ASM Press, pp. 165-179.

COMPEAU, P. E., PEVZNER, P. A. & TESLER, G. 2011. How to apply de Bruijn graphs to genome assembly. *Nat. Biotechnol.,* 29**,** 987-991.

CORTASSA, S., CACERES, V., BELL, L. N., O'ROURKE, B., PAOLOCCI, N. & AON, M. A. 2015. From metabolomics to fluxomics: a computational procedure to translate metabolite profiles into metabolic fluxes. *Biophys. J.,* 108**,** 163-172.

COWAN, D., CASANUEVA, A. & STAFFORD, W. 2007. Ecology and biodiversity of cold-adapted microorganisms. *In:* GERDAY, C. & GLANSDORFF, N. (eds.) *Physiology and Biochemistry of Extremophiles.* Washington DC: ASM Press, pp. 119-132.

COWAN, D., RAMOND, J., MAKHALANYANE, T. & DE MAAYER, P. 2015. Metagenomics of extreme environments. *Curr. Opin. Microbiol.,* 25**,** 97-102.

CROUCHER, N. J. & THOMSON, N. R. 2010. Studying bacterial transcriptomes using RNA-seq. *Curr. Opin. Microbiol.,* 13**,** 619-624.

D'AMICO, S., COLLINS, T., MARX, J. C., FELLER, G. & GERDAY, C. 2006. Psychrophilic microorganisms: challenges for life. *EMBO Rep,* 7**,** 385-9.

DALL'AGNOL, H. P. M. B., BARAÚNA, R. A., DE SÁ, P. H. C. G., RAMOS, R. T. J., NÓBREGA, F., NUNES, C. I. P., DAS GRAÇAS, D. A., CARNEIRO, A. R., SANTOS, D. M., PIMENTA, A. M. C., CAREPO, M. S. P., AZEVEDO, V., PELLIZARI, V. H., SCHNEIDER, M. P. C. & SILVA, A. 2014. Omics profiles used to evaluate the gene expression of *Exiguobacterium antarcticum* B7 during cold adaptation. *BMC Genomics,* 15**,** 986.

DALMASO, G. Z. L., FERREIRA, D. & VERMELHO, A. B. 2015. Marine Extremophiles: A Source of Hydrolases for Biotechnological Applications. *Mar. Drugs,* 13**,** 1925-1965.

DARLING, A. E., MAU, B. & PERNA, N. T. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One,* 5**,** e11147.

DARTNELL, L. 2011. Biological constraints on habitability. *Astro Geophys,* 52**,** 1.25-1.28.

DASSARMA, S., CAPES, M. D., KARAN, R. & DASSARMA, P. 2013. Amino acid substitutions in cold-adapted proteins from *Halorubrum lacusprofundi*, an extremely halophilic microbe from antarctica. *PloS One,* 8**,** e58587.

DE LOURDES MORENO, M., PÉREZ, D., GARCÍA, M. T. & MELLADO, E. 2013. Halophilic bacteria as a source of novel hydrolytic enzymes. *Life,* 3**,** 38-51.

DE MAAYER, P., ANDERSON, D., CARY, C. & COWAN, D. A. 2014a. Some like it cold: understanding the survival strategies of psychrophiles. *EMBO Rep,* 15**,** 508-517.

DE MAAYER, P., CHAN, W., RUBAGOTTI, E., VENTER, S., TOTH, I., BIRCH, P. R. & COUTINHO, T. 2014b. Analysis of the *Pantoea ananatis* pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts. *BMC Genomics,* 15**,** 404.

DE MAAYER, P., VALVERDE, A. & COWAN, D. A. 2014c. The Current State of Metagenomic Analysis. *In:* POPTSOVA, M. S. (ed.) *Genome Analysis: Current Procedures and Applications.* Norfolk, UK: Caister Academic Press, pp. 183-220.

DELONG, E. F. 2000. Extreme genomes. *Genome Biol,* 1**,** 10.

DETTMER, K. & HAMMOCK, B. D. 2004. Metabolomics--a new exciting field within the "omics" sciences. *Environ. Health Perspect.,* 112**,** A396-A397.

DREESENS, L. L., LEE, C. K. & CARY, S. C. 2014. The Distribution and Identity of Edaphic Fungi in the McMurdo Dry Valleys. *Biology,* 3**,** 466-483.

DSOUZA, M., TAYLOR, M. W., TURNER, S. J. & AISLABIE, J. 2014. Genome-based comparative analyses of Antarctic and temperate species of *Paenibacillus*. *PLoS One,* 9**,** e108009.

DSOUZA, M., TAYLOR, M. W., TURNER, S. J. & AISLABIE, J. 2015. Genomic and phenotypic insights into the ecology of *Arthrobacter* from Antarctic soils. *BMC Genomics,* 16**,** 36.

EARL, D., BRADNAM, K., ST JOHN, J., DARLING, A., LIN, D., FASS, J., YU, H. O., BUFFALO, V., ZERBINO, D. R., DIEKHANS, M., NGUYEN, N., ARIYARATNE, P. N., SUNG, W. K., NING, Z., HAIMEL, M., SIMPSON, J. T., FONSECA, N. A., BIROL, I., DOCKING, T. R., HO, I. Y., ROKHSAR, D. S., CHIKHI, R., LAVENIER, D., CHAPUIS, G., NAQUIN, D., MAILLET, N., SCHATZ, M. C., KELLEY, D. R., PHILLIPPY, A. M., KOREN, S., YANG, S. P., WU, W., CHOU, W. C., SRIVASTAVA, A., SHAW, T. I., RUBY, J. G., SKEWES-COX, P., BETEGON, M., DIMON, M. T., SOLOVYEV, V., SELEDTSOV, I., KOSAREV, P., VOROBYEV, D., RAMIREZ-GONZALEZ, R., LEGGETT, R., MACLEAN, D., XIA, F., LUO, R., LI, Z., XIE, Y., LIU, B., GNERRE, S., MACCALLUM, I., PRZYBYLSKI, D., RIBEIRO, F. J., YIN, S., SHARPE, T., HALL, G., KERSEY, P. J., DURBIN, R., JACKMAN, S. D., CHAPMAN, J. A., HUANG, X., DERISI, J. L., CACCAMO, M., LI, Y., JAFFE, D. B., GREEN, R. E., HAUSSLER, D., KORF, I. & PATEN, B. 2011. Assemblathon 1: a competitive assessment of de novo short read assembly methods. *Genome Res.,* 21**,** 2224-41.

EDDY, S. R. 2001. Non–coding RNA genes and the modern RNA world. *Nat. Rev. Genet.,* 2**,** 919-929.

EDWARDS, D. J. & HOLT, K. E. 2013. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microb Inform Exp,* 3**,** 2.

ELLEUCHE, S., SCHRÖDER, C., SAHM, K. & ANTRANIKIAN, G. 2014. Extremozymes — biocatalysts with unique properties from extremophilic microorganisms. *Curr. Opin. Biotechnol.,* 29**,** 116-123.

ERNST, C. & RAHMANN, S. PanCake: A Data Structure for Pangenomes. *In:* BEIßBARTH, T., KOLLMAR, M., LEHA, A., MORGENSTERN, B., SCHULTZ, A.-K., WAACK, S. & WINGENDER, E., eds. German Conference on Bioinformatics, 2013 Germany. Schloss Dagstuhl--Leibniz-Zentrum fuer Informatik, 35-45.

FELLER, G. 2003. Molecular adaptations to cold in psychrophilic enzymes. *Cell. Mol. Life Sci.,* 60**,** 648-62.

FISCHER, S., BRUNK, B. P., CHEN, F., GAO, X., HARB, O. S., IODICE, J. B., SHANMUGAM, D., ROOS, D. S. & STOECKERT, C. J. 2011. Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes Into New Ortholog Groups. *Curr Protoc Bioinformatics***,** 6.12. 1-6.12. 19.

FLEISCHMANN, R. D., ADAMS, M. D., WHITE, O., CLAYTON, R. A., KIRKNESS, E. F., KERLAVAGE, A. R., BULT, C. J., TOMB, J.-F., DOUGHERTY, B. A. & MERRICK, J. M. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science,* 269**,** 496-512.

FU, X., FU, N., GUO, S., YAN, Z., XU, Y., HU, H., MENZEL, C., CHEN, W., LI, Y. & ZENG, R. 2009. Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics,* 10**,** 161.

GABALDÓN, T. & KOONIN, E. V. 2013. Functional and evolutionary implications of gene orthology. *Nat. Rev. Genet.,* 14**,** 360-366.

GEORGANAS, E., BULUC, A., CHAPMAN, J., OLIKER, L., ROKHSAR, D. & YELICK, K. Parallel De Bruijn graph construction and traversal for de novo genome assembly. High Performance Computing, Networking, Storage and Analysis, SC14: International Conference for, 16-21 Nov. 2014 2014. 437-448.

GHOBAKHLOU, A., LABERGE, S., ANTOUN, H., WISHART, D. S., XIA, J., KRISHNAMURTHY, R. & MANDAL, R. 2013. Metabolomic analysis of cold acclimation of Arctic *Mesorhizobium* sp. strain N(33). *PLoS One,* 8**,** e84801.

GILLINGS, M. R. & WESTOBY, M. 2014. DNA technology and evolution of the Central Dogma. *Trends Ecol. Evol.,* 29**,** 1.

GUIMARÃES, L. C., DE JESUS, L. B., CANARIO VIANA, M. V., SILVA, A., JUCA RAMOS, R. T., SOARES, S. D. C. & AZEVEDO, V. 2015. Inside the Pan-genome-methods and software overview. *Curr. Genomics,* 16**,** 245-252.

HAAS, B. J., CHIN, M., NUSBAUM, C., BIRREN, B. W. & LIVNY, J. 2012. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics,* 13**,** 734.

HANDLEY, K. M. & LLOYD, J. R. 2013. Biogeochemical implications of the ubiquitous colonization of marine habitats and redox gradients by *Marinobacter* species. *Front Microbiol,* 4**,** 136.

HENSON, J., TISCHLER, G. & NING, Z. 2012. Next-generation sequencing and large genome assemblies. *Pharmacogenomics,* 13**,** 901-915.

HERNANDEZ, D., FRANÇOIS, P., FARINELLI, L., ØSTERÅS, M. & SCHRENZEL, J. 2008. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Res.,* 18**,** 802-809.

HOFF, M. 2009. Surviving salt: how do extremophiles do it? *PLoS Biol.,* 7**,** e1000258.

HOFFMANN, T., VON BLOHN, C., STANEK, A., MOSES, S., BARZANTNY, H. & BREMER, E. 2012. Synthesis, release, and recapture of compatible solute proline by osmotically stressed Bacillus subtilis cells. *Appl. Environ. Microbiol.,* 78**,** 5753-62.

HORIKOSHI, K. 1999. Alkaliphiles: some applications of their products for biotechnology. *Microbiol. Mol. Biol. Rev.,* 63**,** 735-750.

HORIKOSHI, K. & BULL, A. T. 2011. Prologue: definition, categories, distribution, origin and evolution, pioneering studies, and emerging fields of extremophiles. *In:* HORIKOSHI, K., ANTRANIKIAN, G., BULL, A. T., ROBB, F. T. & STETTER, K. O. (eds.) *Extremophiles handbook.* Japan: Springer, pp. 3-15.

HUMAN GENOME SEQUENCING, C. 2004. Finishing the euchromatic sequence of the human genome. *Nature,* 431**,** 931-945.

HUTCHISON, C. A. 2007. DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res.,* 35**,** 6227-6237.

JAIN, M., FIDDES, I. T., MIGA, K. H., OLSEN, H. E., PATEN, B. & AKESON, M. 2015. Improved data analysis for the MinION nanopore sequencer. *Nat. Methods,* 12**,** 351-356.

JANTO, B., AHMED, A., ITO, M., LIU, J., HICKS, D. B., PAGNI, S., FACKELMAYER, O. J., SMITH, T. A., EARL, J. & ELBOURNE, L. D. 2011. Genome of alkaliphilic *Bacillus pseudofirmus* OF4 reveals adaptations that support the ability to grow in an external pH range from 7.5 to 11.4. *Environ. Microbiol.,* 13**,** 3289-3309.

JOGHEE, N. N. & JAYARAMAN, G. 2014. Metabolomic characterization of halophilic bacterial isolates reveals strains synthesizing rare diaminoacids under salt stress. *Biochimie,* 102**,** 102-111.

JOLIVET, E., L'HARIDON, S., CORRE, E., FORTERRE, P. & PRIEUR, D. 2003. *Thermococcus gammatolerans* sp. nov., a hyperthermophilic archaeon from a deep-sea hydrothermal vent that resists ionizing radiation. *Int. J. Syst. Evol. Microbiol.,* 53**,** 847-851.

JUNGBLUT, P. R. 2014. The proteomics quantification dilemma. *J proteomics,* 107**,** 98-102.

KARAN, R., CAPES, M. D. & DASSARMA, S. 2012. Function and biotechnology of extremophilic enzymes in low water activity. *Aquat Biosyst,* 8.

KASHEFI, K. & LOVLEY, D. R. 2003. Extending the upper temperature limit for life. *Science,* 301**,** 934-934.

KELLEHER, N. L. 2004. Peer reviewed: Top-down proteomics. *Anal Chem,* 76**,** 196 A-203 A.

KELLER, M. & HETTICH, R. 2009. Environmental Proteomics: a Paradigm Shift in Characterizing Microbial Activities at the Molecular Level. *Microbiol. Mol. Biol. Rev.,* 73**,** 62-70.

KENNEDY, S. P., NG, W. V., SALZBERG, S. L., HOOD, L. & DASSARMA, S. 2001. Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res.,* 11**,** 1641-1650.

KISAND, V. & LETTIERI, T. 2013. Genome sequencing of bacteria: sequencing, de novo assembly and rapid analysis using open source tools. *BMC Genomics,* 14**,** 211.

KOGENARU, S., YAN, Q., GUO, Y. & WANG, N. 2012. RNA-seq and microarray complement each other in transcriptome profiling. *BMC Genomics,* 13**,** 629.

KOONIN, E. V. 2005. Orthologs, paralogs, and evolutionary genomics 1. *Annu. Rev. Genet.,* 39**,** 309-338.

KOONIN, E. V. 2009. Evolution of Genome Architecture. *Int J Biochem Cell Biol,* 41**,** 298-306.

KREMKOW, B. & LEE, K. 2015. Sequencing technologies for animal cell culture research. *Biotechnol. Lett.,* 37**,** 55-65.

KRISTENSEN, D. M., WOLF, Y. I., MUSHEGIAN, A. R. & KOONIN, E. V. 2011. Computational methods for Gene Orthology inference. *Brief Bioinform,* 12**,** 379-391.

KRULWICH, T. & ITO, M. 2013. Alkaliphilic prokaryotes. *In:* ROSENBERG, E., DELONG, E., LORY, S., STACKEBRANDT, E. & THOMPSON, F. (eds.) *The Prokaryotes.* Springer Berlin Heidelberg, pp. 441-469.

KRULWICH, T. A., LIU, J., MORINO, M., FUJISAWA, M., ITO, M. & HICKS, D. B. 2011. Adaptive mechanisms of extreme alkaliphiles. *In:* HORIKOSHI, K., ANTRANIKIAN, G., BULL, A. T., ROBB, F. T. & STETTER, K. O. (eds.) *Extremophiles handbook.* Japan: Springer, pp. 119-139.

LAND, M., HAUSER, L., JUN, S.-R., NOOKAEW, I., LEUZE, M. R., AHN, T.-H., KARPINETS, T., LUND, O., KORA, G. & WASSENAAR, T. 2015. Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics,* 15**,** 141-161.

LAPIERRE, P. & GOGARTEN, J. P. 2009. Estimating the size of the bacterial pan-genome. *Trends Genet.,* 25**,** 107-110.

LECHNER, M., FINDEIß, S., STEINER, L., MARZ, M., STADLER, P. F. & PROHASKA, S. J. 2011. Proteinortho: Detection of (Co-) orthologs in large-scale analysis. *BMC Bioinformatics,* 12**,** 124.

LECHNER, M., HERNANDEZ-ROSALES, M., DOERR, D., WIESEKE, N., THÉVENIN, A., STOYE, J., HARTMANN, R. K., PROHASKA, S. J. & STADLER, P. F. 2014. Orthology detection combining clustering and synteny for very large datasets. *PLoS One,* 9**,** e105015.

LI, L., STOECKERT, C. J. & ROOS, D. S. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.,* 13**,** 2178-2189.

LI, R., FAN, W., TIAN, G., ZHU, H., HE, L., CAI, J., HUANG, Q., CAI, Q., LI, B., BAI, Y., ZHANG, Z., ZHANG, Y., WANG, W., LI, J., WEI, F., LI, H., JIAN, M., LI, J., ZHANG, Z., NIELSEN, R., LI, D., GU, W., YANG, Z., XUAN, Z., RYDER, O. A., LEUNG, F. C., ZHOU, Y., CAO, J., SUN, X., FU, Y., FANG, X., GUO, X., WANG, B., HOU, R., SHEN, F., MU, B., NI, P., LIN, R., QIAN, W., WANG, G., YU, C., NIE, W., WANG, J., WU, Z., LIANG, H., MIN, J., WU, Q., CHENG, S., RUAN, J., WANG, M., SHI, Z., WEN, M., LIU, B., REN, X., ZHENG, H., DONG, D., COOK, K., SHAN, G., ZHANG, H., KOSIOL, C., XIE, X., LU, Z., ZHENG, H., LI, Y., STEINER, C. C., LAM, T. T., LIN, S., ZHANG, Q., LI, G., TIAN, J., GONG, T., LIU, H., ZHANG, D., FANG, L., YE, C., ZHANG, J., HU, W., XU, A., REN, Y., ZHANG, G., BRUFORD, M. W., LI, Q., MA, L., GUO, Y., AN, N., HU, Y., ZHENG, Y., SHI, Y., LI, Z., LIU, Q., CHEN, Y., ZHAO, J., QU, N., ZHAO, S., TIAN, F., WANG, X., WANG, H., XU, L., LIU, X., VINAR, T., et al. 2010. The sequence and de novo assembly of the giant panda genome. *Nature,* 463**,** 311-7.

LI, Y., ENGLE, M., WEISS, N., MANDELCO, L. & WIEGEL, J. 1994. *Clostridium thermoalcaliphilum* sp. nov., an anaerobic and thermotolerant facultative alkaliphile. *Int. J. Syst. Bacteriol.,* 44**,** 111-118.

LIEN, S. K., NIEDENFÜHR, S., SLETTA, H., NÖH, K. & BRUHEIM, P. 2015. Fluxome study of *Pseudomonas fluorescens* reveals major reorganisation of carbon flux through central metabolic pathways in response to inactivation of the anti-sigma factor MucA. *BMC Syst Biol,* 9**,** 6.

LIN, Y., LI, J., SHEN, H., ZHANG, L., PAPASIAN, C. J. & DENG, H. W. 2011. Comparative studies of de novo assembly tools for next-generation sequencing technologies. *Bioinformatics,* 27**,** 2031-7.

LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L. & LAW, M. 2012. Comparison of next-generation sequencing systems. *Biomed Res Int,* 2012.

LIU, L., SI, L., MENG, X. & LUO, L. 2015. Comparative transcriptomic analysis reveals novel genes and regulatory mechanisms of *Tetragenococcus halophilus* in response to salt stress. *J. Ind. Microbiol. Biotechnol.,* 42**,** 601-616.

LOGARES, R., HAVERKAMP, T. H., KUMAR, S., LANZÉN, A., NEDERBRAGT, A. J., QUINCE, C. & KAUSERUD, H. 2012. Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. *J. Microbiol. Methods,* 91**,** 106-113.

LÓPEZ-LÓPEZ, O., CERDÁN, M. E. & SISO, M. I. G. 2014. New extremophilic lipases and esterases from metagenomics. *Curr. Protein Peptide Sci.,* 15**,** 445.

LUDWIG, M. & BRYANT, D. A. 2012. *Synechococcus* sp. strain PCC 7002 transcriptome: acclimation to temperature, salinity, oxidative stress, and mixotrophic growth conditions. *Front Microbio,* 3.

MA, Y., GALINSKI, E. A., GRANT, W. D., OREN, A. & VENTOSA, A. 2010. Halophiles 2010: life in saline environments. *Appl. Environ. Microbiol.,* 76**,** 6971-6981.

MACELROY, R. 1974. Some comments on the evolution of extremophiles. *BioSyst.,* 6**,** 74-75.

MACLEAN, D., JONES, J. D. & STUDHOLME, D. J. 2009. Application of'next-generation'sequencing technologies to microbial genetics. *Nat. Rev. Microbiol.,* 7**,** 287-296.

MAJHI, M. C., BEHERA, A. K., KULSHRESHTHA, N. M., KUMAR, R. & KUMAR, A. 2013. ExtremeDB: a unified web repository of extremophilic archaea and bacteria. *PloS One,* 8**,** e63083.

MARDIS, E. R. 2008. The impact of next-generation sequencing technology on genetics. *Trends Genet.,* 24**,** 133-41.

MARGESIN, R. & MITEVA, V. 2011. Diversity and ecology of psychrophilic microorganisms. *Res. Microbiol.,* 162**,** 346-61.

MARGUERAT, S., WILHELM, B. T. & BAHLER, J. 2008. Next-generation sequencing: applications beyond genomes. *Biochem. Soc. Trans.,* 36**,** 1091-6.

MARIN, K., KANESAKI, Y., LOS, D. A., MURATA, N., SUZUKI, I. & HAGEMANN, M. 2004. Gene expression profiling reflects physiological processes in salt acclimation of *Synechocystis* sp. strain PCC 6803. *Plant Physiol.,* 136**,** 3290-3300.

MASKOW, T. & BABEL, W. 2001. Calorimetrically obtained information about the efficiency of ectoine synthesis from glucose in Halomonas elongata. *Biochim. Biophys. Acta,* 1527**,** 4-10.

MAY, C., BROSSERON, F., PFEIFFER, K., MEYER, H. E. & MARCUS, K. 2012. Proteome analysis with classical 2D-PAGE. *In:* MARCUS, K. (ed.) *Quantitative Methods in Proteomics.* New York: Humana Press, pp. 37-46.

MAZIN, P. V., FISUNOV, G. Y., GORBACHEV, A. Y., KAPITSKAYA, K. Y., ALTUKHOV, I. A., SEMASHKO, T. A., ALEXEEV, D. G. & GOVORUN, V. M. 2014. Transcriptome analysis reveals novel regulatory mechanisms in a genome-reduced bacterium. *Nucleic Acids Res.,* 42**,** 13254-13268.

MCGETTIGAN, P. A. 2013. Transcriptomics in the RNA-seq era. *Curr. Opin. Chem. Biol.,* 17**,** 4-11.

MEDIGUE, C., KRIN, E., PASCAL, G., BARBE, V., BERNSEL, A., BERTIN, P., CHEUNG, F., CRUVEILLER, S., D'AMICO, S. & DUILIO, A. 2005. Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res.,* 15**,** 1325 - 1335.

MEDINI, D., DONATI, C., TETTELIN, H., MASIGNANI, V. & RAPPUOLI, R. 2005. The microbial pan-genome. *Curr. Opin. Genet. Dev.,* 15**,** 589-594.

MESBAH, N. M. & WIEGEL, J. 2008. Life at Extreme Limits. *Ann. N. Y. Acad. Sci.,* 1125**,** 44-57.

METHE, B., NELSON, K., DEMING, J., MOMEN, B., MELAMUD, E., ZHANG, X., MOULT, J., MADUPU, R., NELSON, W. & DODSON, R. 2005. The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. *Proc Natl Acad Sci U S A,* 102**,** 10913 - 10918.

METPALLY, R. & REDDY, B. 2009. Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins. *BMC Genomics,* 10**,** 11.

METZKER, M. L. 2010. Sequencing technologies - the next generation. *Nat. Rev. Genet.,* 11**,** 31-46.

MIRA, A., MARTÍN-CUADRADO, A. B., D'AURIA, G. & RODRÍGUEZ-VALERA, F. 2010. The bacterial pan-genome: a new paradigm in microbiology. *Int. Microbiol.,* 13**,** 45-57.

MIRA, A., OCHMAN, H. & MORAN, N. A. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet.,* 17**,** 589-596.

MITSCHKE, J., GEORG, J., SCHOLZ, I., SHARMA, C. M., DIENST, D., BANTSCHEFF, J., VOß, B., STEGLICH, C., WILDE, A. & VOGEL, J. 2011. An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proceedings of the National Academy of Sciences,* 108**,** 2124-2129.

MONGODIN, E. F., NELSON, K. E., DAUGHERTY, S., DEBOY, R. T., WISTER, J., KHOURI, H., WEIDMAN, J., WALSH, D. A., PAPKE, R. T., SANCHEZ PEREZ, G., SHARMA, A. K., NESBO, C. L., MACLEOD, D., BAPTESTE, E., DOOLITTLE, W. F., CHARLEBOIS, R. L., LEGAULT, B. & RODRIGUEZ-VALERA, F. 2005. The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc Natl Acad Sci U S A,* 102**,** 18147-52.

MORAN-REYNA, A. & COKER, J. A. 2014. The effects of extremes of pH on the growth and transcriptomic profiles of three haloarchaea. *F1000Research,* 3**,** 168.

MORENO-HAGELSIEB, G. & LATIMER, K. 2008. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics,* 24**,** 319 - 324.

MOREY, M., FERNÁNDEZ-MARMIESSE, A., CASTIÑEIRAS, D., FRAGA, J. M., COUCE, M. L. & COCHO, J. A. 2013. A glimpse into past, present, and future DNA sequencing. *Mol. Genet. Metab.,* 110**,** 3-24.

MORITA, R. Y. 1975. Psychrophilic bacteria. *Bacteriol Rev,* 39**,** 144.

MOROZOVA, O. & MARRA, M. A. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics,* 92**,** 255-64.

MUTZ, K.-O., HEILKENBRINKER, A., LÖNNE, M., WALTER, J.-G. & STAHL, F. 2013. Transcriptome analysis using next-generation sequencing. *Curr. Opin. Biotechnol.,* 24**,** 22-30.

MYERS, E. W., SUTTON, G. G., DELCHER, A. L., DEW, I. M., FASULO, D. P., FLANIGAN, M. J., KRAVITZ, S. A., MOBARRY, C. M., REINERT, K. H., REMINGTON, K. A., ANSON, E. L., BOLANOS, R. A., CHOU, H. H., JORDAN, C. M., HALPERN, A. L.,

LONARDI, S., BEASLEY, E. M., BRANDON, R. C., CHEN, L., DUNN, P. J., LAI, Z., LIANG, Y., NUSSKERN, D. R., ZHAN, M., ZHANG, Q., ZHENG, X., RUBIN, G. M., ADAMS, M. D. & VENTER, J. C. 2000. A whole-genome assembly of *Drosophila*. *Science,* 287**,** 2196-204.

NAGARAJAN, N., COOK, C., DI BONAVENTURA, M., GE, H., RICHARDS, A., BISHOP-LILLY, K. A., DESALLE, R., READ, T. D. & POP, M. 2010. Finishing genomes with limited resources: lessons from an ensemble of microbial genomes. *BMC Genomics,* 11**,** 242.

NEIFAR, M., MAKTOUF, S., GHORBEL, R. E., JAOUANI, A. & CHERIF, A. 2015. Extremophiles as source of novel bioactive compounds with industrial potential. *In:* GUPTA, V. K. & TUOHY, M. G. (eds.) *Biotechnology of Bioactive Compounds: Sources and Applications.* Chichester: John Wiley & Sons, Ltd, pp. 245-267.

NG, W. V., KENNEDY, S. P., MAHAIRAS, G. G., BERQUIST, B., PAN, M., SHUKLA, H. D., LASKY, S. R., BALIGA, N. S., THORSSON, V. & SBROGNA, J. 2000. Genome sequence of *Halobacterium* species NRC-1. *Proc Natl Acad Sci U S A,* 97**,** 12176-12181.

NIEDRINGHAUS, T. P., MILANOVA, D., KERBY, M. B., SNYDER, M. P. & BARRON, A. E. 2011. Landscape of next-generation sequencing technologies. *Anal Chem,* 83**,** 4327-41.

NIEHAUS, F., BERTOLDO, C., KÄHLER, M. & ANTRANIKIAN, G. 1999. Extremophiles as a source of novel enzymes for industrial application. *Appl. Microbiol. Biotechnol.,* 51**,** 711-729.

NOOKAEW, I., PAPINI, M., PORNPUTTPONG, N., SCALCINATI, G., FAGERBERG, L., UHLÉN, M. & NIELSEN, J. 2012. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Res.***,** gks804.

OH, D.-H., DASSANAYAKE, M., BOHNERT, H. J. & CHEESEMAN, J. M. 2012. Life at the extreme: lessons from the genome. *Genome Biol,* 13**,** 241.

OREN, A. 2007. Biodiversity in highly saline environments. *In:* GLANSDORFF, N. & GERDAY, C. (eds.) *Physiology and biochemistry of extremophiles.* Washington: ASM Press, pp. 223-231.

OREN, A. 2008. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Systems,* 4**,** 13.

OVERBEEK, R., OLSON, R., PUSCH, G. D., OLSEN, G. J., DAVIS, J. J., DISZ, T., EDWARDS, R. A., GERDES, S., PARRELLO, B. & SHUKLA, M. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.,* 42**,** D206-D214.

PADAN, E., BIBI, E., ITO, M. & KRULWICH, T. A. 2005. Alkaline pH homeostasis in bacteria: new insights. *Biochim. Biophys. Acta,* 1717**,** 67-88.

PADE, N. & HAGEMANN, M. 2015. Salt acclimation of cyanobacteria and their application in biotechnology. *Life,* 5**,** 25-49.

PAREJA-TOBES, P., MANRIQUE, M., PAREJA-TOBES, E., PAREJA, E. & TOBES, R. 2012. BG7: a new approach for bacterial genome annotation designed for next generation sequencing data. *PLoS One,* 7**,** e49239.

PASZKIEWICZ, K. & STUDHOLME, D. J. 2010. De novo assembly of short sequence reads. *Brief Bioinform,* 11**,** 457-72.

PAUL, S., BAG, S. K., DAS, S., HARVILL, E. T. & DUTTA, C. 2008. Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol,* 9**,** R70.

PAYNE, S. H. 2015. The utility of protein and mRNA correlation. *Trends Biochem. Sci.,* 40**,** 1-3.

PERKEL, J. M. 2009. Life Science Technologies: Sanger Who? Sequencing the Next Generation. *Science,* 324**,** 275-279.

PEVZNER, P. A., TANG, H. & TESLER, G. 2004. De novo repeat classification and fragment assembly. *Genome Res.,* 14**,** 1786-1796.

PIETTE, F., STRUVAY, C. & FELLER, G. 2011. The protein folding challenge in psychrophiles: facts and current issues. *Environ. Microbiol.,* 13**,** 1924-33.

PINTO, A., MELO-BARBOSA, H., MIYOSHI, A., SILVA, A. & AZEVEDO, V. 2011. Application of RNA-seq to reveal the transcript profile in bacteria. *Genet Mol Res,* 10**,** 1707-1718.

PODAR, M. & REYSENBACH, A.-L. 2006. New opportunities revealed by biotechnological explorations of extremophiles. *Curr. Opin. Biotechnol.,* 17**,** 250-255.

POLZ, M. F., ALM, E. J. & HANAGE, W. P. 2013. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet.,* 29**,** 170-175.

POWELL, S., FORSLUND, K., SZKLARCZYK, D., TRACHANA, K., ROTH, A., HUERTA-CEPAS, J., GABALDÓN, T., RATTEI, T., CREEVEY, C. & KUHN, M. 2013. eggNOG v4. 0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.***,** gkt1253.

POWELL, S., SZKLARCZYK, D., TRACHANA, K., ROTH, A., KUHN, M., MULLER, J., ARNOLD, R., RATTEI, T., LETUNIC, I. & DOERKS, T. 2012. eggNOG v3. 0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.,* 40**,** D284-D289.

QIN, Q. L., XIE, B. B., YU, Y., SHU, Y. L., RONG, J. C., ZHANG, Y. J., ZHAO, D. L., CHEN, X. L., ZHANG, X. Y. & CHEN, B. 2014. Comparative genomics of the marine bacterial genus *Glaciecola* reveals the high degree of genomic diversity and genomic characteristic for cold adaptation. *Environ. Microbiol.,* 16**,** 1642-1653.

QUICK, J., QUINLAN, A. R. & LOMAN, N. J. 2014. A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer. *Gigascience,* 3**,** 22.

RABILLOUD, T., CHEVALLET, M., LUCHE, S. & LELONG, C. 2010. Two-dimensional gel electrophoresis in proteomics: past, present and future. *J proteomics,* 73**,** 2064-2077.

RAGHAVAN, R., GROISMAN, E. A. & OCHMAN, H. 2011. Genome-wide detection of novel regulatory RNAs in E. coli. *Genome Res.,* 21**,** 1487-1497.

REDDY, T. B., THOMAS, A. D., STAMATIS, D., BERTSCH, J., ISBANDI, M., JANSSON, J., MALLAJOSYULA, J., PAGANI, I., LOBOS, E. A. & KYRPIDES, N. C. 2015. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res.,* 43**,** D1099-106.

REUTER, J. A., SPACEK, D. V. & SNYDER, M. P. 2015. High-Throughput Sequencing Technologies. *Mol. Cell,* 58**,** 586-597.

RICHARDSON, E. J. & WATSON, M. 2013. The automatic annotation of bacterial genomes. *Brief Bioinform,* 14**,** 1-12.

RILEY, M., STALEY, J. T., DANCHIN, A., WANG, T. Z., BRETTIN, T. S., HAUSER, L. J., LAND, M. L. & THOMPSON, L. S. 2008. Genomics of an extreme psychrophile, *Psychromonas ingrahamii*. *BMC Genomics,* 9**,** 210.

RIVKINA, E., FRIEDMANN, E., MCKAY, C. & GILICHINSKY, D. 2000. Metabolic activity of permafrost bacteria below the freezing point. *Appl. Environ. Microbiol.,* 66**,** 3230-3233.

RODRIGUES, D. F., IVANOVA, N., HE, Z., HUEBNER, M., ZHOU, J. & TIEDJE, J. M. 2008. Architecture of thermal adaptation in an *Exiguobacterium sibiricum* strain isolated from 3 million year old permafrost: a genome and transcriptome approach. *BMC Genomics,* 9**,** 547.

RODRIGUES, D. F. & TIEDJE, J. M. 2008. Coping with our cold planet. *Appl. Environ. Microbiol.,* 74**,** 1677-1686.

ROEßLER, M. & MÜLLER, V. 2001. Osmoadaptation in bacteria and archaea: common principles and differences. *Environ. Microbiol.,* 3**,** 743-754.

ROTHSCHILD, L. J. & MANCINELLI, R. L. 2001. Life in extreme environments. *Nature,* 409**,** 1092-1101.

RUBIANO-LABRADOR, C., BLAND, C., MIOTELLO, G., GUÉRIN, P., PIBLE, O., BAENA, S. & ARMENGAUD, J. 2014. Proteogenomic insights into salt tolerance by a halotolerant alpha-proteobacterium isolated from an Andean saline spring. *J Proteomics,* 97**,** 36-47.

RUSSELL, N. 2000. Toward a molecular understanding of cold activity of enzymes from psychrophiles. *Extremophiles,* 4**,** 83 - 90.

RUSSELL, N. J. 2007. Psychrophiles: membrane adaptations. *In:* GERDAY, C. & GLANSDORFF, N. (eds.) *Physiology and Biochemistry of Extremophiles.* ASM Press: Washington DC, pp. 155-164.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-5467.

SATYANARAYANA, T., RAGHUKUMAR, C. & SHIVAJI, S. 2005. Extremophilic microbes: Diversity and perspectives. *Curr. Sci.,* 89**,** 78-90p.

SAUM, S. H. & MÜLLER, V. 2008. Regulation of osmoadaptation in the moderate halophile *Halobacillus halophilus*: chloride, glutamate and switching osmolyte strategies. *Saline Systems,* 4**,** 2014.

SCHATZ, M. C., DELCHER, A. L. & SALZBERG, S. L. 2010. Assembly of large genomes using second-generation sequencing. *Genome Res.,* 20**,** 1165-1173.

SCHROETER, R., HOFFMANN, T., VOIGT, B., MEYER, H., BLEISTEINER, M., MUNTEL, J., JÜRGEN, B., ALBRECHT, D., BECHER, D., LALK, M., EVERS, S., BONGAERTS, J., MAURER, K.-H., PUTZER, H., HECKER, M., SCHWEDER, T. & BREMER, E. 2013. Stress responses of the industrial workhorse *Bacillus licheniformis* to osmotic challenges. *PLoS One,* 8**,** e80956.

SHENDURE, J. 2008. The beginning of the end for microarrays? *Nat. Methods,* 5**,** 585-587.

SHENDURE, J. & JI, H. 2008. Next-generation DNA sequencing. *Nat. Biotechnol.,* 26**,** 1135-45.

SHOKRALLA, S., SPALL, J. L., GIBSON, J. F. & HAJIBABAEI, M. 2012. Next-generation sequencing technologies for environmental DNA research. *Mol. Ecol.,* 21**,** 1794-805.

SIEZEN, R. J., WILSON, G. & TODT, T. 2010. Prokaryotic whole-transcriptome analysis: deep sequencing and tiling arrays. *Microb Biotechnol,* 3**,** 125-130.

SIMPSON, J. T., WONG, K., JACKMAN, S. D., SCHEIN, J. E., JONES, S. J. & BIROL, I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res.,* 19**,** 1117-23.

SIVASHANKARI, S. & SHANMUGHAVEL, P. 2007. Comparative genomics - A perspective. *Bioinformation,* 1**,** 376-378.

SOREK, R. & COSSART, P. 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat. Rev. Genet.,* 11**,** 9-16.

STEVENSON, A., BURKHARDT, J., COCKELL, C. S., CRAY, J. A., DIJKSTERHUIS, J., FOX-POWELL, M., KEE, T. P., KMINEK, G., MCGENITY, T. J. & TIMMIS, K. N. 2015. Multiplication of microbes below 0.690 water activity: implications for terrestrial and extraterrestrial life. *Environ. Microbiol.,* 17**,** 257-277.

STOTHARD, P. & WISHART, D. S. 2006. Automated bacterial genome analysis and annotation. *Curr. Opin. Microbiol.,* 9**,** 505-510.

STRUVAY, C. & FELLER, G. 2012. Optimization to low temperature activity in psychrophilic enzymes. *Int J. Mol. Sci.,* 13**,** 11643-11665.

TADEO, X., LÓPEZ-MÉNDEZ, B., TRIGUEROS, T., LAÍN, A., CASTAÑO, D. & MILLET, O. 2009. Structural basis for the aminoacid composition of proteins from halophilic archea. *PLoS Biol.,* 7**,** 2821.

TAKAI, K., MOSER, D. P., ONSTOTT, T. C., SPOELSTRA, N., PFIFFNER, S. M., DOHNALKOVA, A. & FREDRICKSON, J. K. 2001. Alkaliphilus transvaalensis gen. nov., sp. nov., an extremely alkaliphilic bacterium isolated from a deep South African gold mine. *Int. J. Syst. Evol. Microbiol.,* 51**,** 1245-1256.

TAKAI, K., NAKAMURA, K., TOKI, T., TSUNOGAI, U., MIYAZAKI, M., MIYAZAKI, J., HIRAYAMA, H., NAKAGAWA, S., NUNOURA, T. & HORIKOSHI, K. 2008. Cell proliferation at 122 C and isotopically heavy CH4 production by a hyperthermophilic methanogen under high-pressure cultivation. *Proc Natl Acad Sci U S A,* 105**,** 10949-10954.

TAKAMI, H., NAKASONE, K., TAKAKI, Y., MAENO, G., SASAKI, R., MASUI, N., FUJI, F., HIRAMA, C., NAKAMURA, Y. & OGASAWARA, N. 2000. Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with Bacillus subtilis. *Nucleic Acids Res.,* 28**,** 4317-4331.

TATUSOV, R., GALPERIN, M., NATALE, D. & KOONIN, E. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucl Acid Res,* 28**,** 33 - 36.

TATUSOV, R., NATALE, D., GARKAVTSEV, I., TATUSOVA, T., SHANKAVARAM, U., RAO, B., KIRYUTIN, B., GALPERIN, M., FEDOROVA, N. & KOONIN, E. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.,* 29**,** 22 - 28.

TATUSOV, R. L., FEDOROVA, N. D., JACKSON, J. D., JACOBS, A. R., KIRYUTIN, B., KOONIN, E. V., KRYLOV, D. M., MAZUMDER, R., MEKHEDOV, S. L., NIKOLSKAYA, A. N., RAO, B. S., SMIRNOV, S., SVERDLOV, A. V., VASUDEVAN, S., WOLF, Y. I., YIN, J. J. & NATALE, D. A. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics,* 4**,** 41.

TATUSOVA, T., DICUCCIO, M., BADRETDIN, A., CHETVERNIN, V., CIUFO, S. & LI, W. 2013. *Prokaryotic genome annotation pipeline*. *In* The NCBI handbook [Internet]. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); 2013-. Available from: http://www.ncbi.nlm.nih.gov/books/NBK174280/

TETTELIN, H., MASIGNANI, V., CIESLEWICZ, M. J., DONATI, C., MEDINI, D., WARD, N. L., ANGIUOLI, S. V., CRABTREE, J., JONES, A. L., DURKIN, A. S., DEBOY, R. T., DAVIDSEN, T. M., MORA, M., SCARSELLI, M., MARGARIT Y ROS, I., PETERSON, J. D., HAUSER, C. R., SUNDARAM, J. P., NELSON, W. C., MADUPU, R., BRINKAC, L. M., DODSON, R. J., ROSOVITZ, M. J., SULLIVAN, S. A., DAUGHERTY, S. C., HAFT, D. H., SELENGUT, J., GWINN, M. L., ZHOU, L., ZAFAR, N., KHOURI, H., RADUNE, D., DIMITROV, G., WATKINS, K., O'CONNOR, K. J., SMITH, S., UTTERBACK, T. R., WHITE, O., RUBENS, C. E., GRANDI, G., MADOFF, L. C., KASPER, D. L., TELFORD, J. L., WESSELS, M. R., RAPPUOLI, R. & FRASER, C. M. 2005. Genome analysis of multiple pathogenic

isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A,* 102**,** 13950-5.

TETTELIN, H., RILEY, D., CATTUTO, C. & MEDINI, D. 2008. Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.,* 11**,** 472-477.

VAN DE VOSSENBERG, J., UBBINK-KOK, T., ELFERINK, M. G., DRIESSEN, A. J. & KONINGS, W. N. 1995. Ion permeability of the cytoplasmic membrane limits the maximum growth temperature of bacteria and archaea. *Mol. Microbiol.,* 18**,** 925-932.

VAN DE VOSSENBERG, J. L., DRIESSEN, A. J., DA COSTA, M. S. & KONINGS, W. N. 1999. Homeostasis of the membrane proton permeability in *Bacillus subtilis* grown at different temperatures. *Biochim. Biophys. Acta,* 1419**,** 97-104.

VAN DEN BURG, B. 2003. Extremophiles as a source for novel enzymes. *Curr. Opin. Microbiol.,* 6**,** 213-218.

VAN DIJK, E. L., AUGER, H., JASZCZYSZYN, Y. & THERMES, C. 2014. Ten years of next-generation sequencing technology. *Trends Genet.,* 30**,** 418-426.

VAN DOMSELAAR, G. H., STOTHARD, P., SHRIVASTAVA, S., CRUZ, J. A., GUO, A., DONG, X., LU, P., SZAFRON, D., GREINER, R. & WISHART, D. S. 2005. BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Res.,* 33**,** W455-W459.

VAN VLIET, A. H. 2010. Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiol. Lett.,* 302**,** 1-7.

VERNIKOS, G., MEDINI, D., RILEY, D. R. & TETTELIN, H. 2015. Ten years of pan-genome analyses. *Curr. Opin. Microbiol.,* 23**,** 148-154.

VOGEL, C. & MARCOTTE, E. M. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.,* 13**,** 227-232.

WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.,* 10**,** 57-63.

WILKS, J. C., KITKO, R. D., CLEETON, S. H., LEE, G. E., UGWU, C. S., JONES, B. D., BONDURANT, S. S. & SLONCZEWSKI, J. L. 2009. Acid and Base Stress and Transcriptomic Responses in *Bacillus subtilis*. *Appl. Environ. Microbiol.,* 75**,** 981-990.

WILMES, P. & BOND, P. L. 2006. Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol.,* 14**,** 92-97.

YIN, L., XUE, Y. & MA, Y. 2015. Global Microarray Analysis of Alkaliphilic Halotolerant *Bacterium Bacillus* sp. N16-5 Salt Stress Adaptation. *PloS One,* 10**,** e0128649.

YUMOTO, I. 2007. Environmental and Taxonomic Biodiversities of Gram-Positive Alkaliphiles. *In:* GLANSDORFF, C. G. N. & (eds.) *PHYSIOLOGY AND BIOCHEMISTRY OF EXTREMOPHILES.* Washington, D. C: ASM, pp. 295-310.

YUMOTO, I., HIROTA, K., NODASAKA, Y., YOKOTA, Y., HOSHINO, T. & NAKAJIMA, K. 2004. *Alkalibacterium psychrotolerans* sp. nov., a psychrotolerant obligate alkaliphile that reduces an indigo dye. *Int. J. Syst. Evol. Microbiol.,* 54**,** 2379-2383.

YUMOTO, I., HIROTA, K. & YOSHIMUNE, K. 2011. Environmental distribution and taxonomic diversity of alkaliphiles. *Extremophiles Handbook.* Springer, pp. 55-79.

ZENG, X., BIRRIEN, J.-L., FOUQUET, Y., CHERKASHOV, G., JEBBAR, M., QUERELLOU, J., OGER, P., CAMBON-BONAVITA, M.-A., XIAO, X. & PRIEUR, D. 2009. *Pyrococcus* CH1, an obligate piezophilic hyperthermophile: extending the upper pressure-temperature limits for life. *The ISME J,* 3**,** 873-876.

ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.,* 18**,** 821-9.

ZHANG, J., CHIODINI, R., BADR, A. & ZHANG, G. 2011. The impact of next-generation sequencing on genomics. *J Genet Genomics,* 38**,** 95-109.

ZHANG, W., LI, F. & NIE, L. 2010. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology,* 156**,** 287-301.

ZHAO, Y., JIA, X., YANG, J., LING, Y., ZHANG, Z., YU, J., WU, J. & XIAO, J. 2014. PanGP: a tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics,* 30**,** 1297-1299.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Chapter 2

2 Sequencing, assembly, annotation of the draft genome of Antarctic polyextremophile *Nesterenkonia* sp. AN1 and identification of putative molecular determinants underlying its polyextremophily

2.1 Introduction

The ice-free soils of terrestrial Antarctica constitute approximately 0.4 % of the continental land mass (Chan *et al.*, 2013, Hopkins *et al.*, 2006). These cold deserts represent the harshest dry terrestrial environments on Earth (Cary *et al.*, 2010, Cowan *et al.*, 2014). The dry soils are characterized by extreme cold, with frequent wide temperature fluctuations (Aislabie *et al.*, 2006, Dreesens *et al.*, 2014), low water availability, high alkalinity and salinity, elevated levels of UV irradiation and scarcity of nutrients (Aislabie *et al.*, 2006, De Maayer *et al.*, 2014). Despite the harsh physico-geochemical conditions, these soils harbour abundant microbial communities, particularly within refugia where the organisms are shielded from direct impact of some elements of the harsh conditions (Cowan *et al.*, 2014).

On the basis of fitness, microorganisms living in Antarctica are categorized as either specialists (e.g. obligate psychrophiles) or generalists (e.g. psychrotolerant microorganisms) (Vincent, 2000). The former are adapted to survive optimally under the Antarctic ambient environmental conditions. The generalists are further classified into two types, those surviving sub-optimally (Type I) and those capable of adjusting between optimal and suboptimal growth (Type II) depending on the prevailing environmental conditions (Vincent, 2000). Several bacterial isolates from the Antarctic terrestrial habitats exhibit psychrotrophic adaptations. Examples include strains of *Arthrobacter* and *Paenibacillus* isolated from soils in the Ross Sea region (Dsouza *et al.*, 2014, Dsouza *et al.*, 2015). The predominance of psychrotolerant bacteria in Antarctic dry soils has been attributed to the evolution of essential adaptive strategies which enable the organisms to cope with frequent and wide temperature fluctuations in the Dry Valleys (Kirby *et al.*, 2011). Very little is known regarding the precise mechanisms that

determine survival of these microorganisms (Chan *et al.*, 2013). However, functional and community ecology of the Antarctic dry soils have shown that Actinobacteria are among the most successful colonizers of these cold arid soils (Geyer *et al.*, 2014, Schwartz *et al.*, 2014).

The Gram-positive genus *Nesterenkonia* belongs to the family *Micrococcaceae* (Stackebrandt, 2012, Stackebrandt *et al.*, 1995) and can be easily distinguished from other members of the family by the morphological features, G +C content, isoprenoid quinones and composition of peptidoglycans of *Nesterenkonia* spp. (Goodfellow, 2012). *Nesterenkonia* spp. are generally aerobic, catalase positive, chemo-organotrophic and haloalkaliphilic (Collins *et al.*, 2002, Li *et al.*, 2005, Stackebrandt *et al.*, 1995). They are usually coccoid or rod-shaped, with or without branching (Stackebrandt *et al.*, 1995). They are non-spore forming, and non-encapsulated and the genomic DNA is characterized by a high G+C content of 64 % to 72 % (Li *et al.*, 2005). The isoprenoid quinones are mainly comprised of menaquinones with seven, eight and nine isoprene units which are completely unsaturated. The peptidoglycans in *Nesterenkonia* are of the A4α type (Stackebrandt, 2012, Stackebrandt *et al.*, 1995).

Stackebrandt and co workers proposed the emendation of the genus *Micrococcus* after a detailed phylogenetic and chemotaxonomic evaluation of its members (Stackebrandt *et al.*, 1995). Consequently, *Micrococcus halobius* was reclassified and named *Nesterenkonia halobia.* Further affiliation of new strains to the genus were recommended to be on the basis of mena-quinone composition, types of peptidoglycans, morphology and 16S ribosomal RNA (Stackebrandt, 2012). The genus *Nesterenkonia* is currently comprised of thirteen validly described species (Parte, 2013).

Members of the genus *Nesterenkonia* are isolated from a wide range of environmental sources. These include cotton and paper mills in China (Luo *et al.*, 2008, Luo *et al.*, 2009), fermented seafood in Korea (Yoon *et al.*, 2006) as well as faeces of an AIDS patient in France (Edouard *et al.*, 2014). Predominantly, however, *Nesterenkonia* spp. are isolated from extreme

environments. Several strains have been reported mainly from saline environments, such as desert and saline soils in Egypt and China, respectively (Li *et al.*, 2005, Li *et al.*, 2004, Li *et al.*, 2008), hyper-saline lake in east Antarctica (Collins *et al.*, 2002) and a soda lake in Ethiopia (Delgado *et al.*, 2006).

Nesterenkonia sp. AN1 was isolated from soil samples collected from the Miers Valley, which forms part of the McMurdo Dry Valleys of Antarctica (Nel *et al.*, 2011). *Nesterenkonia* sp. AN1 represents the first reported psychrophilic member of the genus. Like other species in the genus, *Nesterenkonia* sp. AN1 is an obligate alkaliphile, with pH requirement of 9 - 10 and halotolerant, capable of optimal growth at 0 to 15 % (wt/vol) salt concentration (Li *et al.*, 2005, Nel *et al.*, 2011). As for several other bacterial isolates from the cold desert soils, *Nesterenkonia* sp. AN1 survives temperatures well below the optimum growth temperature of psychrophiles (15°C) but has an optimal growth temperature of 21°C, and is hence classified as 'psychrotolerant' (Nel *et al.*, 2011). Furthermore, a novel aliphatic amidase (superfamily nitrilase) that is active on short chain amides has been isolated and characterized in *Nesterenkonia* sp. AN1 (Nel *et al.*, 2011).

Here we present the high-quality draft genome sequence of *Nesterenkonia* sp. AN1 and highlight the important adaptive features identified from the genome sequence, which are likely to underscore its polyextremophily.

2.2 Material and Methods

2.3 Genome Sequence of *Nesterenkonia* sp. AN1

2.3.1 Culturing and genomic DNA extraction

Nesterenkonia sp. AN1 was routinely cultured at 21°C in modified Castenholz media (Nel *et al.*, 2011). Cultures were streaked on modified Castenholz agar plates and maintained at 21°C. The colonies were scraped from the plates and transferred to three 50ml sterile tube containing Castenholz broth.

Genomic DNA was extracted using a modified bead beating phenol/chloroform extraction protocol (http://www.utoledo.edu/search.html?q=DNA_extraction_soil.pdf). Briefly, pooled colonies for the replicated cultures were centrifuged for 5 min at 14000 rpm to obtain strong pellets. The pellets were re-suspended in 1 ml of extraction buffer (50 mMNaCl, 50 mMTris-HCl, 50 mM EDTA and 5 % SDS; pH 8). Each suspension was transferred to a sterile 2 ml safe-locks tube containing 0.4 ml of 0.10 mm glass beads. The cells were homogenized using the PowerLyser™ 24 Bench Top Beed-Based Homogenizer at 2,000 rpm for five minutes as per the manufacturer's recommendations. The supernatant was transferred to a new 2ml eppendorf tube and 300 µl of phenol and chloroform/isoamyl alcohol were added. The preparation was mixed by vortexing for 10 seconds and centrifuged at 14000 rpm. The DNA contained in the upper aqueous phase was into a new 2 ml eppendorf tube. Further extraction of DNA was completed using 500 µl of chloroform only. The aqueous upper phase was collected into a 1.5 ml micro centrifuge tube. To improve the purity of the genomic DNA, an RNA removal step was incorporated by adding 1 µl of RNase A to the sample. The sample was incubated at 37°C for 60 minutes. The genomic DNA was precipitated by adding 0.1 and 0.7 volumes of 3 M sodium acetate and isopropanol, respectively and centrifuged at 14,000 rpm for 30 min at 10°C. The supernatant was aspirated carefully. The pellet was washed using 0.5 ml 70 % ice-cold ethanol and centrifuged at maximum speed for 5 min to re-pellet the DNA.

The pellet then re-suspended in 50 µl DNase/RNase-free water. The quality and quantity of the extracted DNA was assessed using a Nanodrop™ spectrometer, Qubit ® 2.0 fluorimeter and visualized by electrophoresis on a 1 % agarose gel.

2.3.2 Genome Sequencing and assembly

A first sequencing run was done using Illumina GA IIx chemistry at the University of Western Cape. Because of the challenges associated with assembling the GA IIx short reads, a second sequencing run was conducted using the Ion torrent PGM at the University of Pretoria sequencing facility. The quality of the reads was assessed using the sequencing QC report tool in CLC Genomics Workbench v. 6 (CLC BIO: http://www.clcbio.com/products/clc-genomics-workbench). The reads were assembled *de novo* using several commercial and publically available assemblers, including Velvet v1.2.10 (Zerbino and Birney, 2008), CLC Genomics Workbench v. 6 (CLC BIO: http://www.clcbio.com/products/clc-genomics-workbench) and the Seqman NGen assembler v11 (DNAStar - http://www.dnastar.com/t-nextgen-seqman-ngen.aspx). To generate highly accurate genome assemblies, the assembled contiguous sequences (contigs) from the individual assemblers were compared using Mauve (Darling *et al.*, 2010) and Bioedit v 7.2.5 (Hall, 1999, Hall, 2011). The contigs from CLC and NGen were assembled further into longer contigs ('scaffolds') using different *in silico* strategies including local BLAST search within and between contigs from different assemblers using Bioedit v 7.2.5 (Hall, 1999, Hall, 2011) and searching for open reading frames using NCBI-ORF finder (Sayers *et al.*, 2011). Final decision to extend contigs was based on the comparison of the *Nesterenkonia* sp. AN1 contigs against the draft genomes sequences of *Nesterenkonia* sp. F, *Nesterenkonia* sp. NP1 and *N. alba* DSM 19423, obtained from the NCBI database using Mauve (Darling *et al.*, 2004, Darling *et al.*, 2010).

2.3.3 Genome Annotation

The draft genome of *Nesterenkonia* sp. AN1 was uploaded into several structural annotation pipelines, including the Rapid Annotation using Subsystem Technology (RAST) server (Overbeek *et al.*, 2014), FGENESB (Solovyev and Salamov, 2011), GeneMark.hmm (prokaryotic version 2.8) (Borodovsky and Lomsadze, 2011, Borodovsky and Lomsadze, 2013), the NCBI prokaryotic genome annotation pipeline (Tatusova *et al.*, 2013) and the Bacterial Annotation SYStem (BASys) pipeline (Van Domselaar *et al.*, 2005). Protein coding sequences (CDSs) that were identified as hypothetical proteins were assessed using the NCBI conserved domain batch search tool (Marchler-Bauer *et al.*, 2015). To assign putative functions to the CDSs, the proteins were functionally annotated using eggNOG non-supervised orthologous groups (Powell *et al.*, 2013). The genes encoding tRNA were predicted using the tRNA prediction program ARAGORN v. 1.2 (Laslett and Canback, 2004). Cellular localization for all the predicted proteins of *Nesterenkonia* sp. AN1 was determined using CELLO2GO (Yu *et al.*, 2014). The tool uses a combination of CELLO (Yu *et al.*, 2006) localization methods and Blast analysis of characterised proteins with GO annotation. CELLO2GO has been reported to predict subcellular localization in Gram-positive bacteria with 99.4 % accuracy (Yu *et al.*, 2014).

The draft genome sequence was deposited and made public on the NCBI under the GenBank accession number NZ_JEMO00000000 (GI: 738529954) and NCBI Reference Sequence number NZ_JEMO00000000.1. The draft genome is also available via the genomes online database (GOLD) under the accession number Gp0085412.

2.3.4 Identification of genomic islands, prophages and extrachromosomal elements

The draft genome of *Nesterenkonia* sp. AN1 was queried for the presence of genomic islands (GIs) using IslandViewer 3 (Dhillon *et al.*, 2015). The predicted features in the various GIs were further assessed via BlastP searches against the NCBI RefSeq and non-redundant (nr) protein databases (Jenuth, 2000). The genome was also analysed for the presence of phage elements using the online PHAST tool (Zhou *et al.*, 2011).

2.3.5 Identification of stress response mechanisms

The proteins encoded on the genome of *Nesterenkonia* sp. AN1 were screened to identify proteins with putative adaptive functions which have been reported in the literature (Dsouza *et al.*, 2015, Medigue *et al.*, 2005). Protein functions were also determined on the basis of the *Nesterenkonia* sp. AN1 annotations obtained using RAST subsystems (Overbeek *et al.*, 2014), BASys bacterial annotation pipeline (Van Domselaar *et al.*, 2005), eggNOG classification (Powell *et al.*, 2013, Powell *et al.*, 2012) and via BLASTP analysis using the NCBI RefSeq and non-redundant (nr) protein databases (Jenuth, 2000).

2.4 Results and Discussion

2.4.1 Genome sequencing and assembly

The first sequencing run, using the Illumina GA IIx, yielded 5,177,635 reads of 44 bp average length, with an estimated genome coverage of ~ 36x. The assembly of these reads produced a large number of contigs, probably due to the presence of repeat regions in the genome. This prompted additional sequencing using the Ion Torrent chemistry, which yields longer reads. The Ion Torrent PGM yielded 3,842,066 reads of mean length and coverage of 324 bp and ~ 351x, respectively.

The reads from both Illumina GA IIx and Ion-torrent PGM platforms were assembled *de novo* using CLC Genomics Workbench v. 6, DNAStar Seqman NGen v11 and Velvet v1.2.10 (Table 2-1). By applying different *in silico* techniques, contigs derived from the Illumina GAIIx reads were assembled into 410 contigs of ~ 2.98 megabases (Mb), with a GC content of 67.48 % and a contig length range of between 133 and 62,284 Mb. Regions of overlap between these contigs and those produced from Ion-torrent PGM reads were identified using Mauve (Darling *et al.*, 2004, Darling *et al.*, 2010) and the information was utilised to improve the assembly of the Ion-torrent PGM reads. The final genome assembly from the Ion-torrent PGM yielded a genome of ~ 3.04 Mb assembled in 41 contigs ranging between 1,439 and 339,148 nucleotides in length with a mean G+C content 67.42 %.

Table 2-1: Statistics for the assemblies obtained using different assemblers. (I and T in brackets represent reads from to Illumina GA IIX and Ion-torrent PGM reads, respectively).

Assembler	CLC (I)	Velvet (I)	CLC (T)	Ngen (T)
Kmer	23	27	23	20
Number of Contigs	1,021	2,871	1,652	2,847
N50	6,626	2,503	22,528	5,972
Max contig length	29,152	13,761	77,390	44,663

2.4.2 Genome annotation

The draft genome sequence of *Nesterenkonia* sp. AN1 was annotated using different automatic annotation programs. It was noted that some of the predicted proteins coding sequences (CDSs) were split, probably due to sequencing error. To resolve these, the CDSs predicted using FGENESB were queried against the genomes of *Nesterenkonia* sp. F, NP1 and *N. alba* DSM 19423 (details of genomes in chapter 3) using BlastP and tBlastN analyses. Out of the 2,970 CDSs predicted on the *Nesterenkonia* sp. AN1 genome, 234 CDSs were split and intact open reading frames for all of the 234 CDSs were identified on the genome of at least one of the three *Nesterenkonia* strains. The final annotation of the *Nesterenkonia* sp. AN1 included 2,852 genes which code for 2,200 non-hypothetical and 574 hypothetical proteins. A total of 1,277 (44.78 %) were encoded on the positive strand while 1,575 (55.20 %) were encoded on the negative strand (Figure 2-1).

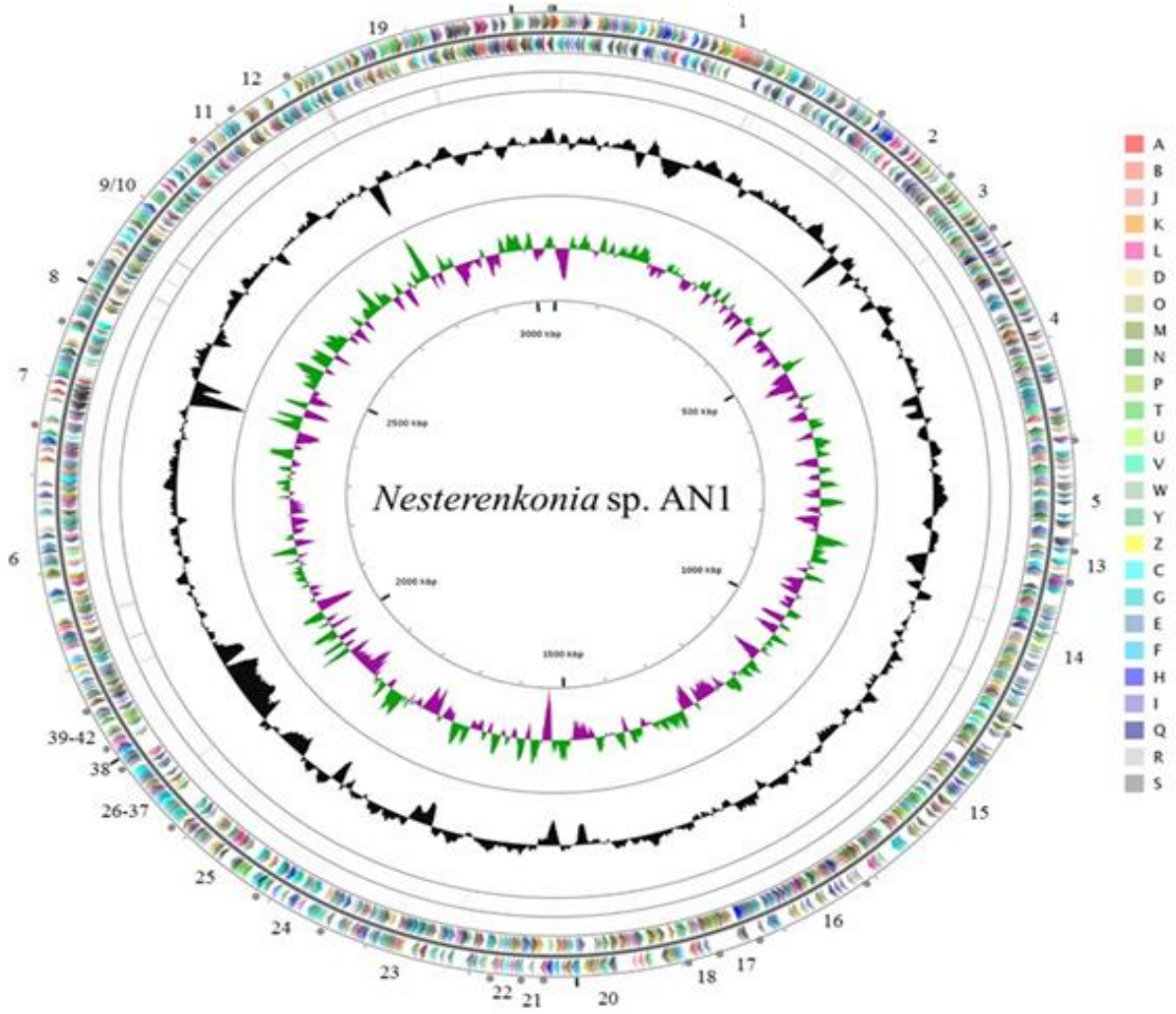


Figure 2-1: Circular representation of the genome structure in *Nesterenkonia* sp. AN1 plotted on the CG View Server (Grant and Stothard, 2008). The contigs of *Nesterenkonia* sp. AN1 were ordered using the contigs of *N. alba* DSM 19423. The first two circles from the outside show mapping CDS encoded on the positive (first circle) and negative (second strand) strands of the genome and the COG classification predicted using eggNOG (description of the COGs is given in Figure 2-3), the third circle shows the position of the genes coding for rRNAs (blast 1), the fourth circle depicts the tRNA encoding genes (blast 2). The fifth and sixth circles show the percentage G+C content plot and GC skew (green indicating GC skew + and purple GC skew -), respectively.

Functional annotation of the CDSs in *Nesterenkonia* sp. AN1 revealed that 2,561 (~ 90 %) of the 2,852 CDSs were assigned to different functional categories using eggNOG (Figures 2-2). About 10 % of the CDSs could not be assigned to any functional category. The functional distribution of the 2,561 CDSs showed that the majority of the proteins belong to the 'metabolism' (~ 41 %) and 'poorly categorised' (~ 27 %) super-functional categories. Further analyses of the CDSs revealed an over-representation of proteins assigned to the functional category 'general function prediction only (R)' which is a trend observed in the genomes of most bacteria (Monnet *et al.*, 2010). Aside from these, the CDSs encoded were enriched in the functional categories depicted in Figure 2-3. The overrepresentation of CDSs linked to the functional classes E and K has been reported for several species in the family Micrococcaceae (Edouard *et al.*, 2014). It is noteworthy that there were higher proportions of CDSs associated with the functions 'energy production and conversion' (C) and 'inorganic ion transport and metabolism; (P) in *Nesterenkonia* sp. AN1 when compared to other members of Micrococcaceae (Edouard *et al.*, 2014, Monnet *et al.*, 2010).

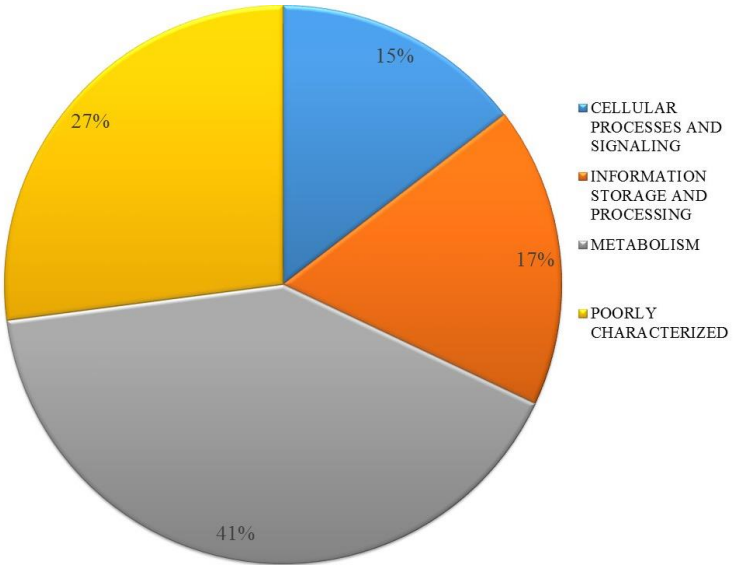


Figure 2-2: Proportion of CDSs in the *Nesterenkonia* sp. AN1 genome characterised using the Actinobacteria non-supervised orthologous groups (actNOG) obtained from the Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups (eggNOG v3.0) database (Powell et al., 2012).

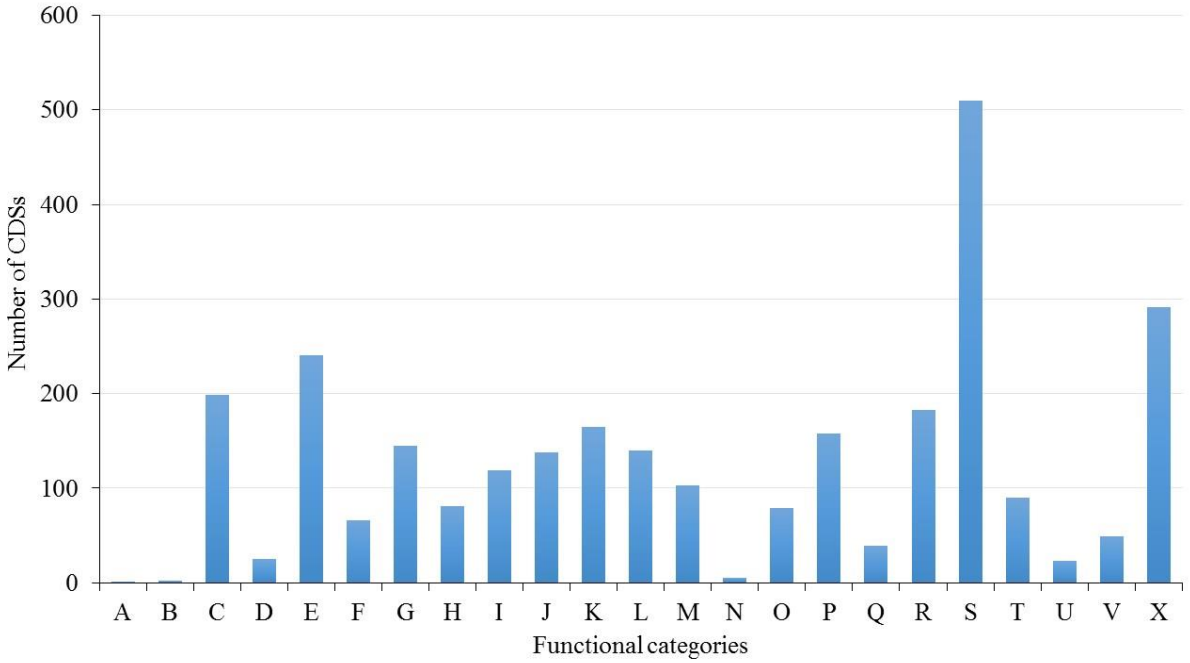


Figure 2-3: Functional categories of CDSs of *Nesterenkonia* sp. AN1. The COG categories are defined as follows: "INFORMATION STORAGE AND PROCESSING; translation, ribosomal structure and biogenesis (J), transcription (K), replication, recombination and repair (L) and chromatin structure and dynamics (B). CELLULAR PROCESSES AND SIGNALING; cell cycle control, cell division, chromosome partitioning (D), defense mechanisms (V), signal transduction mechanisms (T), cell wall/membrane/envelope biogenesis (M), intracellular trafficking, secretion, and vesicular transport (U) and posttranslational modification, protein turnover, chaperones (O). METABOLISM; energy production and conversion (C), carbohydrate transport and metabolism (G), amino acid transport and metabolism (E), nucleotide transport and metabolism (F), coenzyme transport and metabolism (H), lipid transport and metabolism (I), inorganic ion transport and metabolism (P) and secondary metabolites biosynthesis, transport and catabolism (Q). POORLY CHARACTERIZED; general function prediction only (R) and function unknown (S). And X represents the unknown proteins".

Subcellular localisation of proteins have been shown to correlate to the functions performed by the proteins (Rudner and Losick, 2010).The accurate prediction of the cellular location of proteins is therefore a crucial step in determining their functions (Dehzangi *et al.*, 2015, Real *et al.*, 2008). Of the 2,852 proteins encoded on the genome, 1,844 (~ 65 %) were predicted to be localised in the cytoplasm (Figure 2-4). Furthermore, sixteen (~ 1 %) and 275 (~ 10 %) of the proteins were predicted to be associated with cell wall or extracellular, respectively. A total of 717 (~ 25 %) were predicted to be localised in the cytoplasmic membrane (Figure 2-4).

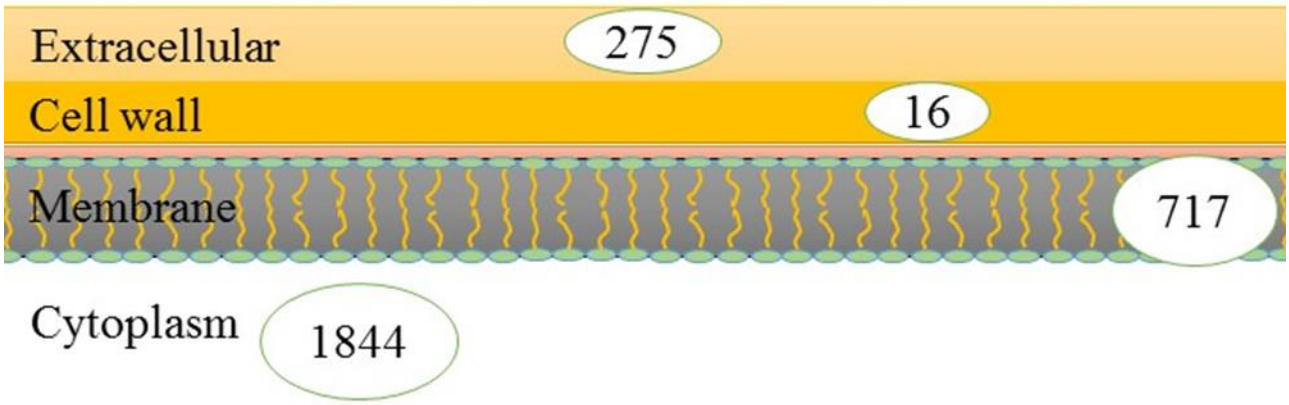


Figure 2-4: Putative localisation of *Nesterenkonia* sp. AN1 proteins predicted using the CELLO2GO server (Yu et al., 2006, Yu et al., 2014).

Three structural RNAs (16S, 23S and 5S rRNA) were predicted using RAST. The exact number of copies of ribosomal genes could not be determined from the draft genome as these are highly conserved and are hence difficult to completely assemble *in silico*. Fifty five tRNA coding genes were predicted using the ARAGORN annotation program. This included one copy of the pyrrolysine tRNA ($tRNA^{Pyl}$). Pyrrolysine is the 22nd proteinogenic amino acid encoded by amber stop codon (UAG) and utilised in the biosynthesis of methane (Ambrogelly *et al.*, 2007, Have *et al.*, 2013). This amino acid was first discovered in the archaeon *Methanosarcina* and the bacterium *Desulfitobacterium hafniense* (Ambrogelly *et al.*, 2007). Pyrrolysine has since been reported to occur in the genomes of at least twelve archaeal and bacterial species (Gaston *et al.*, 2011, Srinivasan *et al.*, 2002).

2.4.3 Genomic islands, prophages and other extrachromosomal elements

The draft genome of *Nesterenkonia* sp. AN1 was searched for the presence of plasmid replication elements (including *rep* and *mob* genes). The absence of orthologs of these genes from the genome sequence indicates that *Nesterenkonia* sp. AN1 does not contain any plasmids. Similarly, no plasmids were reported in the genome sequences of three other *Nesterenkonia*

species whose genomes are publicly available (Edouard *et al.*, 2014, Sarikhan *et al.*, 2011). The *Nesterenkonia* sp. AN1 genome contained one intact prophage. This prophage was identified using the PHAST pipeline, with a completeness score of 130 (Figure 2-5). The prophage element is a predicted 19.7Kb in size and codes for seventeen phage proteins and seven hypothetical proteins. The phage proteins had homologs in thirteen phage species with the largest number of hits (three proteins) associated with several mycobacteriophages. The average G+C content of the phage elements (66. 7 %) was slightly lower than that of the means G+C content for the genome (67.5 %) (Aliyu *et al.*, 2014).

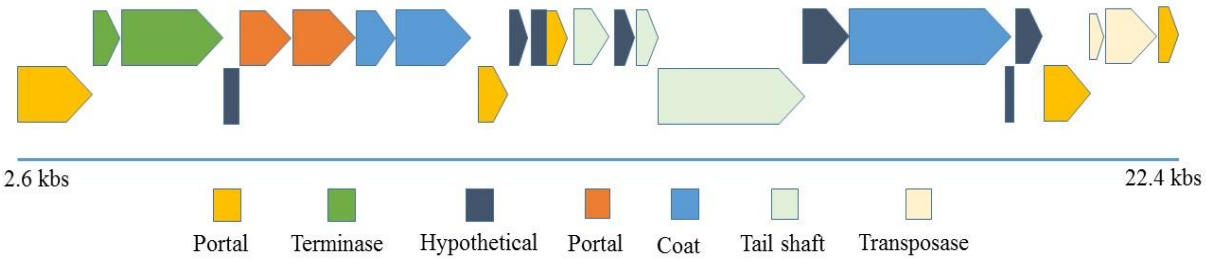


Figure 2-5. Prophage proteins encoded on the *Nesterenkonia* sp. AN1 genome identified and annotated using PHAST (Zhou et al., 2011).

Sixteen genomic islands (GIs) ranging between 4 and 21 Kb in size were predicted in the genome sequence of *Nesterenkonia* sp. AN1 using the Islandviewer platform (Figure 2-6, Appendix T 1). The GIs comprised of a total of 169 genes (Appendix T 1), of which 139 were annotated using the NCBI RefSeq database. Fifty (36 %) of these genes encoded hypothetical proteins. Approximately, 72 % of the predicted proteins were classified into eggNOG non-supervised orthologous groups (Powell *et al.*, 2013). The majority of the proteins (~ 32 %) were of unknown functions. Aside from the proteins with unknown function, there was a high proportion (~ 17 %) of proteins grouped in the replication, recombination and repair functional category. Other proteins were associated with cell wall/membrane/envelope biogenesis (~

11 %), transcription (~ 7 %), post-translational modification, protein turnover, chaperones (5 %) and signal transduction mechanisms (5 %) (Figure 2-7).

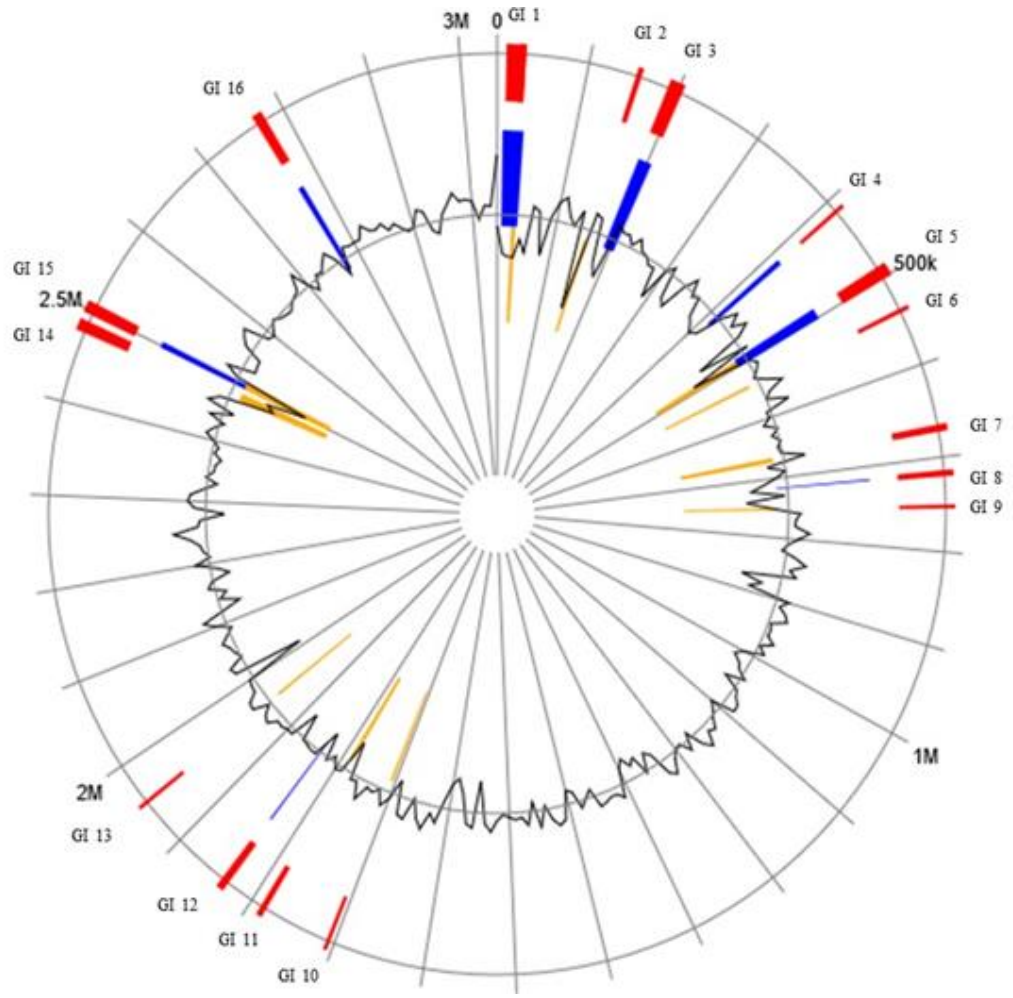


Figure 2-6: Circular plot showing the positions of the genomic islands (GIs) predicted on the genome of *Nesterenkonia* sp. AN1. The blue and yellow blocks represent the GIs predictions based on SIGI-HMM and IslandPath-DIMOB methods implemented by the IslandViewer 3 tool (Dhillon et al., 2015). The red blocks represents the final GIs predictions based on the integration of the two methods.

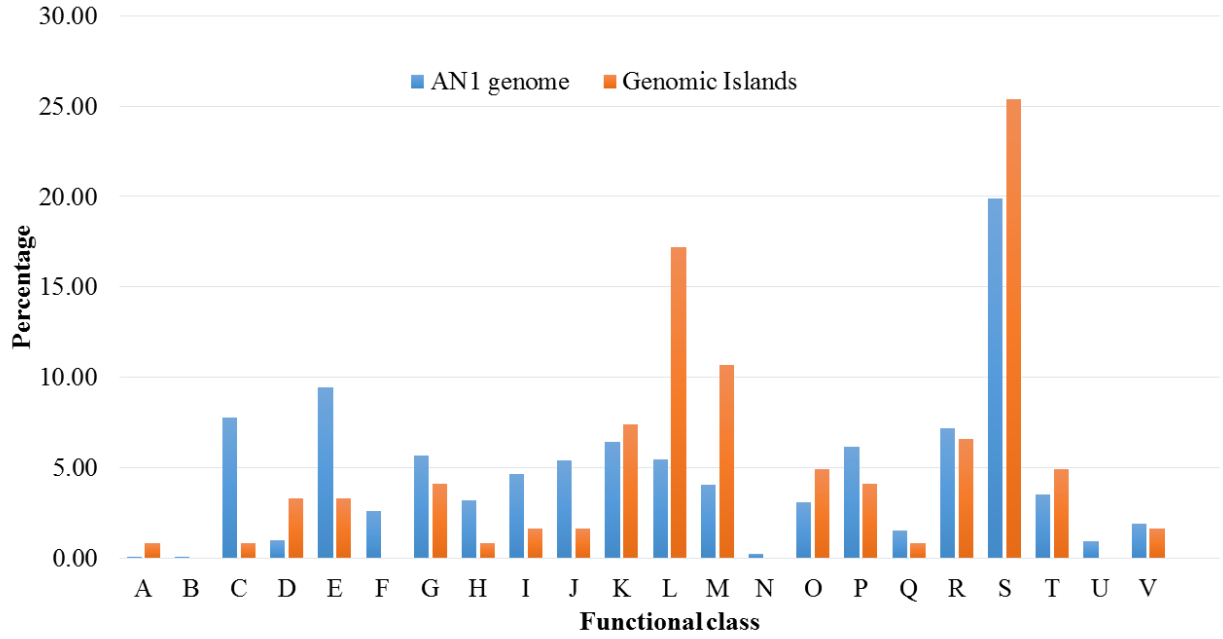


Figure 2-7: Comparison of the functional annotation of predicted proteins encoded on the genome of *Nesterenkonia* sp. AN1 and the genomic islands identified in the genome. A description of the EggNOG functional classes is given in Figure 2-3.

Notable among the proteins encoded on the GIs in the replication, recombination and repair category are orthologs of a DNA mismatch repair enzyme (GI2), DNA polymerase III subunit alpha (GI13), excinuclease ABC subunit A (GI14) and excinuclease ABC subunit C (GI14) (Appendix T 1). These protein have been reported to play a role in resistance to gamma radiation (Sghaier *et al.*, 2008, Leuko *et al.*, 2011, Sghaier, 2011). The rest of the genes in this group include six integrases (GI3, GI5, GI9, and GI12), a resolvase (GI1) and eight transposases (GI3, GI4, GI5, GI8 and GI15) (Appendix T 1). These proteins have been shown to be integral components of genomic islands as one of the principal signatures of horizontal gene transfer (HGT) (Zhang *et al.*, 2015, Dobrindt *et al.*, 2004).

The GIs also contain genes encoding proteins that were grouped in the "inorganic ion transport" class (Appendix T 1), including genes encoding potassium transporter proteins (GI11). These proteins have been implicated in the uptake of potassium (K^+) ions, the major mechanism for

protection against osmotic stress identified in various microorganisms (Corratgé-Faillie *et al.*, 2010). Other genes of the "inorganic ion transport" group include putative zinc (ABC transporter substrate-binding protein) (GI3) and cobalt (G5) transporters which could also be linked to protection against osmotic stress (Eitinger *et al.*, 2011, Wood, 2015).

Nesterenkonia sp. AN1 has the largest genome of the four *Nesterenkonia* species for which the genomes are publicly available (which range in size from 2.59 Mb to 3.04 Mb). The large genome size of *Nesterenkonia* sp. AN1 could be partly attributed to the acquisition of genetic elements via horizontal gene transfer evident by the presence of several distinct GIs (Juhas *et al.*, 2009). The size of bacterial genomes has been shown to be determined largely by events like horizontal gene transfer, gene decay and gene duplication (Mira *et al.*, 2001, Ventura *et al.*, 2007). However, microorganisms living in complex environments, e.g. edaphic microorganisms tend to show larger genome size and encode more features that are likely associated with different adaptation strategies (Bentley and Parkhill, 2004, Ventura *et al.*, 2007).

2.5 Stress response mechanisms

For microorganisms to withstand the thermal and related stresses inherent in the arid soils of Antarctica (De Maayer *et al.*, 2014, Tehei and Zaccai, 2005), they must possess adaptive features that ensure "normal" functioning of their cellular processes in the extreme environments (Thieringer *et al.*, 1998). To elucidate the stress response mechanisms of *Nesterenkonia* sp. AN1, the draft genome was queried for a range of known stress response features.

The CDSs encoded on the genome of *Nesterenkonia* sp. AN1 were compared to those known to play a role in stress response. One hundred and eighty-nine (189) genes that are putatively associated with different stress response mechanisms could be identified (Figure 2-8). The list

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

includes fifty genes with putative roles in different aspects of cold adaptations, forty-seven oxidative stress genes, fifty-nine genes involved in osmotic stress protection, nineteen genes linked to DNA repair mechanisms and fourteen genes classified in the general stress response category (Appendix T 2).

2.5.1 Cold stress response

2.5.1.1 Cold shock proteins and cold acclimation proteins

Microorganisms synthesise cold shock proteins (CSPs) and cold acclimation proteins (CAPs) that enable them to cope with cold stress (Berger *et al.*, 1996). CSPs are rapidly produced on temperature downshift in the external environment. The activity of the CSPs is usually transient in response to sudden decrease in temperature, compared to cold acclimation proteins whose activity is continuous under cold condition (Phadtare, 2004, Phadtare *et al.*, 1999). CSPs act as RNA chaperones which stabilise RNAs and enhance their translation at low temperature (Chaikam and Karlson, 2010, Rabus *et al.*, 2004). However, their activity varies with the system in which they are expressed. For instance, CspC is cold inducible in *Bacillus subtilis* but not associated with cold stress response in *Escherichia coli* (Graumann and Marahiel, 1999). CspA has also been shown to mediate transcription antitermination production of secondary cold shock proteins which are critical survival features in some bacteria (Phadtare *et al.*, 2007, Rodrigues and Tiedje, 2008). For example, pychrotrophic bacterium *Yersinia enterocolitica* W22703 mutant for the secondary cold shock protein, PNPase, was unable to survive at 5°C (Goverde *et al.*, 1998). The rapid synthesis and continuous expression of CapA and other CspA-like proteins is associated with prolonged cold exposure in psychrophilic bacteria. This feature is not observed in their mesophilic counterparts (Berger *et al.*, 1997, Phadtare *et al.*, 1999).

Cold acclimation proteins have been reported to act as RNA chaperones, where some are involved in adjusting membrane fluidity and the production of anti-freeze proteins (Berger *et*

al., 1996). The DeaD-box ATP-dependent RNA helicase and RbfA have been implicated in ribosomes biogenesis under cold condition. DeaD is also associated with other activities including RNA degradation and cold acclimation (Charollais *et al.*, 2004, Phadtare, 2011).

The genome of *Nesterenkonia* sp. AN1 encodes twenty predicted cold shock and cold acclimation proteins (Appendix T 2, Figure 2-8). Some of the CSPs and CAPs encoded on the *Nesterenkonia* sp. AN1 genome include, cold shock proteins CspA and CspC, cold-shock DEAD-box protein a homolog (DeaD), translation initiation factor IF-1 & 2 (Inf1 & Inf2), Transcription elongation protein (NusA), polyribonucleotide nucleotidyltransferase (PNPase) and ribosome-binding factor A (RbfA).

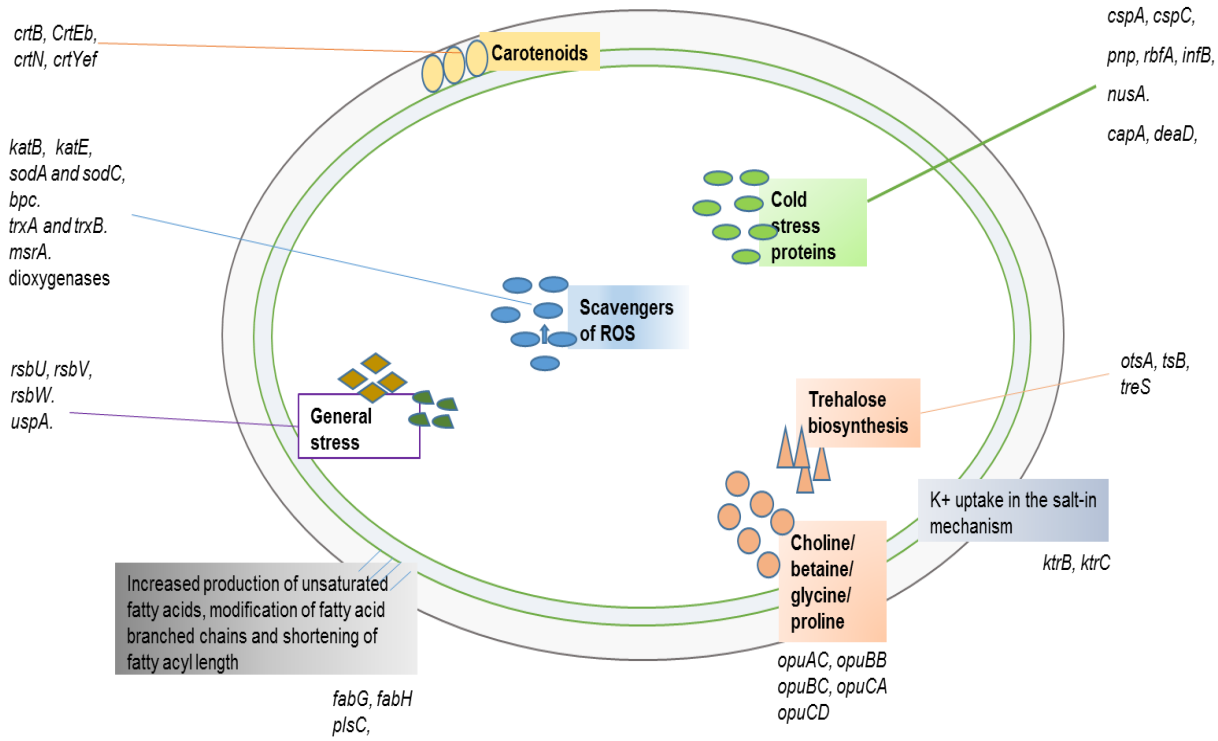


Figure 2-8: Important stress response features of *Nesterenkonia* sp. AN1.

2.5.1.2 Membrane adaptations

The *Nesterenkonia* sp. AN1 genome encodes multiple proteins that are important in modulation of membrane fluidity, a crucial cold adaptation strategy in microorganisms (D'Amico *et al.*, 2006, Feller, 2003). These include nineteen proteins with putative roles in fatty acid biosynthesis and eleven proteins associated with the production of carotenoid pigments. Some of the fatty acids biosynthetic elements include orthologs of the 3-oxoacyl-[acyl-carrier-protein] reductase (FabG) and 3-oxoacyl-[acyl-carrier-protein] synthases 2 & 3 (FabF and FabH) which catalyse the initial reaction in the biosynthesis and condensation of fatty acids (Hoang *et al.*, 2002, Methe *et al.*, 2005). Other *Nesterenkonia* sp. AN1 proteins with putative roles in cold adaptation of the membrane include orthologs of 1-acyl-sn-glycerol-3-phosphate acyltransferase (PlsC), which catalyses intermediary steps in the biosynthesis of phospholipids and 3-Ketoacyl-(Acyl-Carrier-Protein) reductase which enhances the synthesis of polyunsaturated lipids which are essential for proper functioning of the membrane at low temperatures (Methe *et al.*, 2005).

Some psychrophilic bacteria have pigments that are involved in many essential processes at low temperature, including photosynthesis, osmo-protection, UV screening and modulation of membrane fluidity (Chintalapati *et al.*, 2004, Dieser *et al.*, 2010, Morgan-Kiss *et al.*, 2006). The low temperature-dependent induction of C-50 carotenoid biosynthesis has been demonstrated in a psychrophilic strain of *Arthrobacter agilis* (Fong *et al.*, 2001). The majority of carotenoid biosynthetic features of *Nesterenkonia* sp. AN1 were encoded at two loci. The first locus contained six genes; *crtB, crtEb*, *crtN, crtYef* and *ipi*, which encode phytoene synthase, lycopene elongase, phytoene dehydrogenase, C50 carotenoid epsilon cyclase, and isopentenyl-diphosphate delta-isomerase. The second locus encoded two carotenoid genes; *crtEb* and *crtN* which code for lycopene elongase and phytoene dehydrogenase, respectively (Appendix T 2, Figure 2-8). Furthermore, the gene *crtE* which encodes geranylgeranyl

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

diphosphate synthase was also predicted on the genome of *Nesterenkonia* sp. AN1. Geranylgeranyl diphosphate synthase, phytoene synthase and phytoene dehydrogenase catalyse the production of lycopene (Klassen, 2010, Krubasik *et al.*, 2001). The enzyme isopentenyl-diphosphate delta-isomerase catalyses the conversion of isopentenyl-diphosphate to dimethylallyl pyrophosphate (DMPP) (Hagi *et al.*, 2015). Furthermore, lycopene elongase catalyses the extension of lycopene with DMMPP to produce flavuxanthin which is subsequently cyclised to decaprenoxanthin by C50 carotenoid epsilon cyclase (Hagi *et al.*, 2015, Krubasik *et al.*, 2001). Thus, it is evident from the genome data that two carotenoids, decaprenoxanthin and lycopene, are potentially synthesised in *Nesterenkonia* sp. AN1. In addition to their putative roles in different stress responses (Chintalapati *et al.*, 2004, Dieser *et al.*, 2010, Morgan-Kiss *et al.*, 2006), these carotenoids probably determine the orange coloration of *Nesterenkonia* sp. AN1 colonies.

2.5.2 Oxidative stress response

Normal metabolic activities result in the endogenous accumulation of hydrogen peroxide H_2O_2 and other types of reactive oxygen species (ROS) (Bayr, 2005). Furthermore, solubility of gasses increases with decreasing temperature, which implies that microorganisms surviving in cold environments are subject to toxic effects of reactive oxygen species (ROS) (Chattopadhyay, 2006, D'Amico *et al.*, 2006).

The mechanisms by which microorganisms cope with ROS have been described in many psychrophiles, e.g. the permafrost dwelling *Planococcus halocryophilus* Or1 and the cold Arctic sediment bacterium *Desulfotalea psychrophila* (Mykytczuk *et al.*, 2013, Rabus *et al.*, 2004). Generally, these involve production of factors that scavenge ROS by splitting them into less toxic derivatives that are easily processed (Baatout *et al.*, 2006, Nystrom, 2006). The antioxidants encoded on the genome of *Nesterenkonia* sp. AN1 include orthologs of the

thioredoxin (TrxA) and thioredoxin reductase (TrxB), thiol peroxidase (Bcp), superoxide dismutases (SodA and SodC), methionine sulfoxide reductase (MsrA) and the catalases (KatB and KatE) (Appendix T 2, Figure 2-8). SodA and SodC catalyse the conversion of superoxide anion radicals (O_2^-) to hydrogen peroxide (H_2O_2) and oxygen (O_2) (Fridovich, 1983, Molina-Heredia *et al.*, 2006). Other enzymes such as KatB, KatE and Bcp catalyse the disproportionation of H_2O_2 into H_2O and O_2 (Ebara and Shigemori, 2008, Moustafa *et al.*, 2010). TrxA and TrxB scavenge on ROS and also balance the intracellular levels of thiol-disulfide (Ballal and Manna, 2010). Furthermore, MsrA is implicated in shielding methionine from the toxic effects of ROS (Weissbach *et al.*, 2005).

The *Nesterenkonia* sp. AN1 genome also encodes seven dioxygenases. These enzymes are associated with antioxidant activity (Medigue *et al.*, 2005, Thannickal and Fanburg, 2000). The presence of multiple dioxygenases in the marine Antarctic *Pseudoalteromonas haloplanktis* TAC125, the genome of which encodes sixteen dioxygenases, has been predicted to play a role in resistance against ROS damage (Medigue *et al.*, 2005).

2.5.3 Protection against osmotic stress

The genome sequence of *Nesterenkonia* sp. AN1 revealed the presence of both major osmotic stress response strategies known in bacteria; the salt-in strategy and the accumulation of osmolytes (Mesbah and Wiegel, 2008, Oren, 2007, Oren, 2008). In total, the genome contained fifty-nine genes that code for proteins linked to different forms of osmotic stress protection (Appendix T 2, Figure 2-8). The salt-in system is made up of Ktr K^+ uptake proteins (KtrB and KtrC) involved in the influx of K^+ to counterbalance the salinity gradient created by the high salt concentration in the external environment. One of the two loci encoding KtrB and KtrC is localized on the genomic island GI11 (Appendix T 1), possibly acquired for protection against osmotic stress. The genome encodes two sets of proteins for the accumulation of compatible

solutes. The first set of proteins includes a number transporters for the uptake of glycine, betaine, choline and proline. The second set is linked to the endogenous biosynthesis of trehalose (Kappes *et al.*, 1999, Kuhlmann and Bremer, 2002, Wargo *et al.*, 2008). Trehalose biosynthetic genes include orthologs of *otsA, otsB and treS* which code for alpha-trehalose-phosphate synthase, trehalose-6-phosphate phosphatase and trehalose synthase, respectively. These enzymes were also significantly induced in *E. coli* in response to downshifts in temperature with a corresponding increase in the production of trehalose (Kandror *et al.*, 2002).

2.5.4 General stress response

Analysis of the *Nesterenkonia* sp. AN1 genome also revealed other adaptation features considered as general stress response elements. For instance, the DNA repair system in *Nesterenkonia* sp. AN1 includes putative genes encoding proteins such as RecA, RecN and UvrA-C. These genes have been reported to occur within loci regulated by the LexA repressor in *E. coli* (Kamenšek *et al.*, 2010, White *et al.*, 1999). The genome also encodes the sigma B regulatory cascade (RsbU, RsbV and RsbW) comprising proteins which regulate the sigma B. Sigma B is one of the alternative dissociable subunits of RNA polymerase that is essential for transcription initiation under stress (van Schaik and Abee, 2005). The regulator of sigma subunit RsbU, dephosphorylates the antagonistic factor RsbV, which in turn binds to the anti-sigma factor RsbW. This results in the sequestration of sigma B factor from RsbW and its subsequent association with RNA polymerase to initiate transcription under cold stress (Foster, 2005, Marles-Wright and Lewis, 2007, Utratna *et al.*, 2014). Sigma factors have also been associated with responses to a wide range of other environmental stresses (Mykytczuk *et al.*, 2013, van Schaik and Abee, 2005). Seven copies of the general stress response proteins (UspA) are encoded on the *Nesterenkonia* sp. AN1 genome. UspA proteins have been implicated in responses to several stresses including cold, osmotic and oxidative stresses and carbon

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

starvation (Garnier *et al.*, 2010, Phadtare, 2011, Seifart Gomes *et al.*, 2011). Similarly, seven copies of UspA encoding genes were predicted in the genome of the obligate psychrophile *Psychromonas ingrahamii* (Riley *et al.*, 2008).

2.6 Conclusion

Here we have present the draft genome sequence of *Nesterenkonia* sp. AN1 (Aliyu *et al.*, 2014), the first psychrotolerant member of the genus *Nesterenkonia*. The high quality draft genome sequence, which comprised forty one contigs, was generated using two sequencing runs. Further assembly was hampered by the presence of repetitive elements and highly conserved genes such as transposases and ribosomal RNA genes. However, the high quality draft status provided a good benchmark for the downstream analyses of the genome. The genome of the Antarctic *Nesterenkonia* sp. AN1 encoded a repertoire of features linked to the stresses typical of its Antarctic edaphic origin. These include proteins involved in responses to cold shock, osmotic and oxidative stresses, modulation of membrane fluidity and many others linked to general stress responses. Furthermore, features underlying the alkaliphilic and halophilic nature of members of the genus *Nesterenkonia* could be identified. The combination of these features are likely to determine the survival of *Nesterenkonia* sp.AN1 in the Antarctic Dry Valley soil.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

2.7 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 Biol. Biochem.,* 38**,** 3041-3056.

ALIYU, H., DE MAAYER, P., REES, J., TUFFIN, M. & COWAN, D. A. 2014. Draft Genome Sequence of the Antarctic Polyextremophile *Nesterenkonia* sp. Strain AN1. *Genome Announc,* 2.

AMBROGELLY, A., PALIOURA, S. & SÖLL, D. 2007. Natural expansion of the genetic code. *Nat. Chem. Biol.,* 3**,** 29-35.

BAATOUT, S., DE BOEVER, P. & MERGEAY, M. 2006. Physiological changes induced in four bacterial strains following oxidative stress. *Appl. Biochem. Microbiol.,* 42**,** 369-377.

BALLAL, A. & MANNA, A. C. 2010. Control of thioredoxin reductase gene (trxB) transcription by SarA in *Staphylococcus aureus*. *J. Bacteriol.,* 192**,** 336-345.

BAYR, H. 2005. Reactive oxygen species. *Crit Care Med,* 33**,** S498-S501.

BENTLEY, S. D. & PARKHILL, J. 2004. Comparative genomic structure of prokaryotes. *Annu. Rev. Genet.,* 38**,** 771-791.

BERGER, F., MORELLET, N., MENU, F. & POTIER, P. 1996. Cold shock and cold acclimation proteins in the psychrotrophic bacterium *Arthrobacter globiformis* SI55. *J. Bacteriol.,* 178**,** 2999-3007.

BERGER, F., NORMAND, P. & POTIER, P. 1997. capA, a cspA-like gene that encodes a cold acclimation protein in the psychrotrophic bacterium *Arthrobacter globiformis* SI55. *J. Bacteriol.,* 179**,** 5670-5676.

BORODOVSKY, M. & LOMSADZE, A. 2011. Gene identification in prokaryotic genomes, phages, metagenomes, and EST sequences with GeneMarkS suite. *Curr Protoc Bioinformatics***,** 4.5. 1-4.5. 17.

BORODOVSKY, M. & LOMSADZE, A. 2013. Gene Identification in Prokaryotic Genomes, Phages, Metagenomes, and EST Sequences with GeneMarkS Suite. *Curr Protoc Microbiol***,** 1E. 7.1-1E. 7.17.

CARY, S. C., MCDONALD, I. R., BARRETT, J. E. & COWAN, D. A. 2010. On the rocks: the microbiology of Antarctic Dry Valley soils. *Nat. Rev. Microbiol.,* 8**,** 129-138.

CHAIKAM, V. & KARLSON, D. T. 2010. Comparison of structure, function and regulation of plant cold shock domain proteins to bacterial and animal cold shock domain proteins. *BMB Rep,* 43**,** 1-8.

CHAN, Y., VAN NOSTRAND, J. D., ZHOU, J., POINTING, S. B. & FARRELL, R. L. 2013. Functional ecology of an Antarctic dry valley. *Proc Natl Acad Sci U S A,* 110**,** 8990-8995.

CHAROLLAIS, J., DREYFUS, M. & IOST, I. 2004. CsdA, a cold-shock RNA helicase from Escherichia coli, is involved in the biogenesis of 50S ribosomal subunit. *Nucleic Acids Res.,* 32**,** 2751-2759.

CHATTOPADHYAY, M. 2006. Mechanism of bacterial adaptation to low temperature. *J. Biosci. (Bangalore),* 31**,** 157-165.

CHINTALAPATI, S., KIRAN, M. & SHIVAJI, S. 2004. Role of membrane lipid fatty acids in cold adaptation. *Cell Mol Biol (Noisy-le-grand),* 50**,** 631-642.

COLLINS, M. D., LAWSON, P. A., LABRENZ, M., TINDALL, B. J., WEISS, N. & HIRSCH, P. 2002. *Nesterenkonia lacusekhoensis* sp. nov., isolated from hypersaline Ekho Lake, East Antarctica, and emended description of the genus *Nesterenkonia*. *Int. J. Syst. Evol. Microbiol.,* 52**,** 1145-1150.

CORRATGÉ-FAILLIE, C., JABNOUNE, M., ZIMMERMANN, S., VÉRY, A.-A., FIZAMES, C. & SENTENAC, H. 2010. Potassium and sodium transport in non-animal cells: the Trk/Ktr/HKT transporter family. *Cell. Mol. Life Sci.,* 67**,** 2511-2532.

COWAN, D. A., MAKHALANYANE, T. P., DENNIS, P. G. & HOPKINS, D. W. 2014. Microbial ecology and biogeochemistry of continental Antarctic soils. *Front Microbiol,* 5**,** 154.

D'AMICO, S., COLLINS, T., MARX, J. C., FELLER, G. & GERDAY, C. 2006. Psychrophilic microorganisms: challenges for life. *EMBO Rep,* 7**,** 385-9.

DARLING, A., MAU, B., BLATTNER, F. & PERNA, N. 2004. Mauve: multiple alignment of conserved genomic sequences with rearrangements. *Genome Res.,* 14**,** 1394 - 1403.

DARLING, A. E., MAU, B. & PERNA, N. T. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One,* 5**,** e11147.

DE MAAYER, P., ANDERSON, D., CARY, C. & COWAN, D. A. 2014. Some like it cold: understanding the survival strategies of psychrophiles. *EMBO Rep,* 15**,** 508-517.

DEHZANGI, A., SOHRABI, S., HEFFERNAN, R., SHARMA, A., LYONS, J., PALIWAL, K. & SATTAR, A. 2015. Gram-positive and gram-negative subcellular localization using rotation forest and physicochemical-based features. *BMC Bioinformatics,* 16**,** S1.

DELGADO, O., QUILLAGUAMÁN, J., BAKHTIAR, S., MATTIASSON, B., GESSESSE, A. & HATTI-KAUL, R. 2006. *Nesterenkonia aethiopica* sp. nov., an alkaliphilic, moderate halophile isolated from an Ethiopian soda lake. *Int. J. Syst. Evol. Microbiol.,* 56**,** 1229-1232.

DHILLON, B. K., LAIRD, M. R., SHAY, J. A., WINSOR, G. L., LO, R., NIZAM, F., PEREIRA, S. K., WAGLECHNER, N., MCARTHUR, A. G. & LANGILLE, M. G. 2015.

IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res.*, gkv401.

DIESER, M., GREENWOOD, M. & FOREMAN, C. M. 2010. Carotenoid pigmentation in Antarctic heterotrophic bacteria as a strategy to withstand environmental stresses. *Arct. Antarct. Alp. Res.,* 42, 396-405.

DOBRINDT, U., HOCHHUT, B., HENTSCHEL, U. & HACKER, J. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.,* 2, 414 - 424.

DREESENS, L. L., LEE, C. K. & CARY, S. C. 2014. The Distribution and Identity of Edaphic Fungi in the McMurdo Dry Valleys. *Biology,* 3, 466-483.

DSOUZA, M., TAYLOR, M. W., TURNER, S. J. & AISLABIE, J. 2014. Genome-based comparative analyses of Antarctic and temperate species of *Paenibacillus*. *PLoS One,* 9, e108009.

DSOUZA, M., TAYLOR, M. W., TURNER, S. J. & AISLABIE, J. 2015. Genomic and phenotypic insights into the ecology of *Arthrobacter* from Antarctic soils. *BMC Genomics,* 16, 36.

EBARA, S. & SHIGEMORI, Y. 2008. Alkali-tolerant high-activity catalase from a thermophilic bacterium and its overexpression in *Escherichia coli*. *Protein Expression Purif.,* 57, 255-260.

EDOUARD, S., SANKAR, S., DANGUI, N. P., LAGIER, J. C., MICHELLE, C., RAOULT, D. & FOURNIER, P. E. 2014. Genome sequence and description of *Nesterenkonia massiliensis* sp. nov. strain NP1(T.). *Stand Genomic Sci,* 9, 866-82.

EITINGER, T., RODIONOV, D. A., GROTE, M. & SCHNEIDER, E. 2011. Canonical and ECF-type ATP-binding cassette importers in prokaryotes: diversity in modular organization and cellular functions. *FEMS Microbiol. Rev.,* 35, 3-67.

FELLER, G. 2003. Molecular adaptations to cold in psychrophilic enzymes. *Cell. Mol. Life Sci.,* 60, 648-62.

FONG, N., BURGESS, M., BARROW, K. & GLENN, D. 2001. Carotenoid accumulation in the psychrotrophic bacterium *Arthrobacter agilis* in response to thermal and salt stress. *Appl. Microbiol. Biotechnol.,* 56, 750-756.

FOSTER, P. L. 2005. Stress responses and genetic variation in bacteria. *Mutat. Res.,* 569, 3-11.

FRIDOVICH, I. 1983. Superoxide radical: an endogenous toxicant. *Annu. Rev. Pharmacol. Toxicol.,* 23, 239-257.

GARNIER, M., MATAMOROS, S., CHEVRET, D., PILET, M.-F., LEROI, F. & TRESSE, O. 2010. Adaptation to cold and proteomic responses of the psychrotrophic biopreservative *Lactococcus piscium* strain CNCM I-4031. *Appl. Environ. Microbiol.,* 76, 8011-8018.

GASTON, M. A., JIANG, R. & KRZYCKI, J. A. 2011. Functional context, biosynthesis, and genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.,* 14**,** 342-349.

GEYER, K. M., ALTRICHTER, A. E., TAKACS-VESBACH, C. D., VAN HORN, D. J., GOOSEFF, M. N. & BARRETT, J. E. 2014. Bacterial community composition of divergent soil habitats in a polar desert. *FEMS Microbiol. Ecol.,* 89**,** 490-494.

GOODFELLOW, M. 2012. Phylum XXVI. Actinobacteria phyl. nov. *In:* WHITMAN, W., GOODFELLOW, M., KÄMPFER, P., BUSSE, H.-J., TRUJILLO, M., LUDWIG, W., SUZUKI, K.-I. & PARTE, A. (eds.) *Bergey's manual of systematic bacteriology.* New York: Springer, pp. 33-2028.

GOVERDE, R. L., HUIS IN'T VELD, J. H., KUSTERS, J. G. & MOOI, F. R. 1998. The psychrotrophic bacterium *Yersinia enterocolitica* requires expression of pnp, the gene for polynucleotide phosphorylase, for growth at low temperature (5 C). *Mol. Microbiol.,* 28**,** 555-569.

GRANT, J. R. & STOTHARD, P. 2008. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res.,* 36**,** W181-W184.

GRAUMANN, P. L. & MARAHIEL, M. A. 1999. Cold shock response in Bacillus subtilis. *J. Mol. Microbiol. Biotechnol.,* 1**,** 203-209.

HAGI, T., KOBAYASHI, M. & NOMURA, M. 2015. Aerobic conditions increase isoprenoid biosynthesis pathway gene expression levels for carotenoid production in *Enterococcus gilvus*. *FEMS Microbiol. Lett.*, 362.

HALL, T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser,* 41**,** 95 - 98.

HALL, T. 2011. BioEdit: An important software for molecular biology. *GERF Bulletin of Bioscience,* 2**,** 60-61.

HAVE, C. T., ZAMBACH, S. & CHRISTIANSEN, H. 2013. Effects of using coding potential, sequence conservation and mRNA structure conservation for predicting pyrrolysine containing genes. *BMC Bioinformatics,* 14**,** 118.

HOANG, T. T., SULLIVAN, S. A., CUSICK, J. K. & SCHWEIZER, H. P. 2002. β-Ketoacyl acyl carrier protein reductase (FabG) activity of the fatty acid biosynthetic pathway is a determining factor of 3-oxo-homoserine lactone acyl chain lengths. *Microbiology,* 148**,** 3849-3856.

HOPKINS, D., SPARROW, A., NOVIS, P., GREGORICH, E., ELBERLING, B. & GREENFIELD, L. 2006. Controls on the distribution of productivity and organic resources in Antarctic Dry Valley soils. *Proc Biol Sci,* 273**,** 2687-2695.

JENUTH, J. 2000. The NCBI. Publicly available tools and resources on the Web. *Methods Mol Biol,* 132**,** 301 - 312.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

JUHAS, M., VAN DER MEER, J. R., GAILLARD, M., HARDING, R. M., HOOD, D. W. & CROOK, D. W. 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.,* 33**,** 376-393.

KAMENŠEK, S., PODLESEK, Z., GILLOR, O. & ŽGUR-BERTOK, D. 2010. Genes regulated by the *Escherichia coli* SOS repressor LexA exhibit heterogenous expression. *BMC Microbiol.,* 10**,** 283.

KANDROR, O., DELEON, A. & GOLDBERG, A. L. 2002. Trehalose synthesis is induced upon exposure of *Escherichia coli* to cold and is essential for viability at low temperatures. *Proc Natl Acad Sci U S A,* 99**,** 9727-9732.

KAPPES, R. M., KEMPF, B., KNEIP, S., BOCH, J., GADE, J., MEIER-WAGNER, J. & BREMER, E. 1999. Two evolutionarily closely related ABC transporters mediate the uptake of choline for synthesis of the osmoprotectant glycine betaine in *Bacillus subtilis*. *Mol. Microbiol.,* 32**,** 203-216.

KIRBY, B., BARNARD, D., MARLA TUFFIN, I. & COWAN, D. 2011. Ecological Distribution of Microorganisms in Terrestrial, Psychrophilic Habitats. *In:* HORIKOSHI, K. (ed.) *Extremophiles Handbook.* Springer Japan, pp. 839-863.

KLASSEN, J. L. 2010. Phylogenetic and Evolutionary Patterns in Microbial Carotenoid Biosynthesis Are Revealed by Comparative Genomics. *PLoS ONE,* 5**,** e11257.

KRUBASIK, P., KOBAYASHI, M. & SANDMANN, G. 2001. Expression and functional analysis of a gene cluster involved in the synthesis of decaprenoxanthin reveals the mechanisms for C50 carotenoid formation. *Eur. J. Biochem.,* 268**,** 3702-3708.

KUHLMANN, A. U. & BREMER, E. 2002. Osmotically regulated synthesis of the compatible solute ectoine in Bacillus pasteurii and related Bacillus spp. *Appl. Environ. Microbiol.,* 68**,** 772-83.

LASLETT, D. & CANBACK, B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.,* 32**,** 11-16.

LEUKO, S., NEILAN, B., BURNS, B., WALTER, M. & ROTHSCHILD, L. 2011. Molecular assessment of UVC radiation-induced DNA damage repair in the stromatolitic halophilic archaeon, *Halococcus hamelinensis*. *J. Photochem. Photobiol. B: Biol.,* 102**,** 140-145.

LI, W.-J., ZHANG, Y.-Q., SCHUMANN, P., LIU, H.-Y., YU, L.-Y., ZHANG, Y.-Q., STACKEBRANDT, E., XU, L.-H. & JIANG, C.-L. 2008. *Nesterenkonia halophila* sp. nov., a moderately halophilic, alkalitolerant actinobacterium isolated from a saline soil. *Int. J. Syst. Evol. Microbiol.,* 58**,** 1359-1363.

LI, W. J., CHEN, H. H., KIM, C. J., ZHANG, Y. Q., PARK, D. J., LEE, J. C., XU, L. H. & JIANG, C. L. 2005. *Nesterenkonia sandarakina* sp. nov. and *Nesterenkonia lutea* sp. nov., novel actinobacteria, and emended description of the genus *Nesterenkonia*. *Int. J. Syst. Evol. Microbiol.,* 55**,** 463-6.

LI, W. J., CHEN, H. H., ZHANG, Y. Q., SCHUMANN, P., STACKEBRANDT, E., XU, L. H. & JIANG, C. L. 2004. *Nesterenkonia halotolerans* sp. nov. and *Nesterenkonia xinjiangensis* sp. nov., actinobacteria from saline soils in the west of China. *Int. J. Syst. Evol. Microbiol.,* 54**,** 837-41.

LUO, H.-Y., MIAO, L.-H., FANG, C., YANG, P.-L., WANG, Y.-R., SHI, P.-J., YAO, B. & FAN, Y.-L. 2008. *Nesterenkonia flava* sp. nov., isolated from paper-mill effluent. *Int. J. Syst. Evol. Microbiol.,* 58**,** 1927-1930.

LUO, H.-Y., WANG, Y.-R., MIAO, L.-H., YANG, P.-L., SHI, P.-J., FANG, C.-X., YAO, B. & FAN, Y.-L. 2009. *Nesterenkonia alba* sp. nov., an alkaliphilic actinobacterium isolated from the black liquor treatment system of a cotton pulp mill. *Int. J. Syst. Evol. Microbiol.,* 59**,** 863-868.

MARCHLER-BAUER, A., DERBYSHIRE, M. K., GONZALES, N. R., LU, S., CHITSAZ, F., GEER, L. Y., GEER, R. C., HE, J., GWADZ, M., HURWITZ, D. I., LANCZYCKI, C. J., LU, F., MARCHLER, G. H., SONG, J. S., THANKI, N., WANG, Z., YAMASHITA, R. A., ZHANG, D., ZHENG, C. & BRYANT, S. H. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res.,* 43**,** D222-D226.

MARLES-WRIGHT, J. & LEWIS, R. J. 2007. Stress responses of bacteria. *Curr. Opin. Struct. Biol.,* 17**,** 755-760.

MEDIGUE, C., KRIN, E., PASCAL, G., BARBE, V., BERNSEL, A., BERTIN, P., CHEUNG, F., CRUVEILLER, S., D'AMICO, S. & DUILIO, A. 2005. Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res.,* 15**,** 1325 - 1335.

MESBAH, N. M. & WIEGEL, J. 2008. Life at Extreme Limits. *Ann. N. Y. Acad. Sci.,* 1125**,** 44-57.

METHE, B., NELSON, K., DEMING, J., MOMEN, B., MELAMUD, E., ZHANG, X., MOULT, J., MADUPU, R., NELSON, W. & DODSON, R. 2005. The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. *Proc Natl Acad Sci U S A,* 102**,** 10913 - 10918.

MIRA, A., OCHMAN, H. & MORAN, N. A. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet.,* 17**,** 589-596.

MOLINA-HEREDIA, F. P., HOUÉE-LEVIN, C., BERTHOMIEU, C., TOUATI, D., TREMEY, E., FAVAUDON, V., ADAM, V. & NIVIÈRE, V. 2006. Detoxification of superoxide without production of H2O2: antioxidant activity of superoxide reductase complexed with ferrocyanide. *Proc Natl Acad Sci U S A,* 103**,** 14750-14755.

MONNET, C., LOUX, V., GIBRAT, J.-F., SPINNLER, E., BARBE, V., VACHERIE, B., GAVORY, F., GOURBEYRE, E., SIGUIER, P. & CHANDLER, M. 2010. The *Arthrobacter arilaitensis* Re117 genome sequence reveals its genetic adaptation to the surface of cheese. *PLoS One,* 5**,** e15489.

MORGAN-KISS, R. M., PRISCU, J. C., POCOCK, T., GUDYNAITE-SAVITCH, L. & HUNER, N. P. 2006. Adaptation and acclimation of photosynthetic microorganisms to permanently cold environments. *Microbiol. Mol. Biol. Rev.,* 70**,** 222-252.

MOUSTAFA, D. A., JAIN, N., SRIRANGANATHAN, N. & VEMULAPALLI, R. 2010. Identification of a single-nucleotide insertion in the promoter region affecting the sodc promoter activity in *Brucella neotomae*. *PloS One,* 5**,** e14112.

MYKYTCZUK, N. C., FOOTE, S. J., OMELON, C. R., SOUTHAM, G., GREER, C. W. & WHYTE, L. G. 2013. Bacterial growth at -15 degrees C; molecular insights from the permafrost bacterium *Planococcus halocryophilus* Or1. *ISME J,* 7**,** 1211-26.

NEL, A., TUFFIN, I., SEWELL, B. & COWAN, D. 2011. Unique aliphatic amidase from a psychrotrophic and haloalkaliphilic *Nesterenkonia* isolate. *Appl. Environ. Microbiol.,* 77**,** 3696-3702.

NYSTROM, T. 2006. Oxidation of bacterial proteome in response to starvation. *Methods Biochem Anal,* 49**,** 89.

OREN, A. 2007. Biodiversity in highly saline environments. *In:* GLANSDORFF, N. & GERDAY, C. (eds.) *Physiology and biochemistry of extremophiles.* Washington: ASM Press, pp. 223-231.

OREN, A. 2008. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Systems,* 4**,** 13.

OVERBEEK, R., OLSON, R., PUSCH, G. D., OLSEN, G. J., DAVIS, J. J., DISZ, T., EDWARDS, R. A., GERDES, S., PARRELLO, B. & SHUKLA, M. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.,* 42**,** D206-D214.

PARTE, A. C. 2013. LPSN—list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res.***,** gkt1111.

PHADTARE, S. 2004. Recent developments in bacterial cold-shock response. *Curr. Issues Mol. Biol.,* 6**,** 125-136.

PHADTARE, S. 2011. Unwinding activity of cold shock proteins and RNA metabolism. *RNA biology,* 8**,** 394-397.

PHADTARE, S., ALSINA, J. & INOUYE, M. 1999. Cold-shock response and cold-shock proteins. *Curr. Opin. Microbiol.,* 2**,** 175-180.

PHADTARE, S., KAZAKOV, T., BUBUNENKO, M., COURT, D. L., PESTOVA, T. & SEVERINOV, K. 2007. Transcription Antitermination by Translation Initiation Factor IF1. *J. Bacteriol.,* 189**,** 4087-4093.

POWELL, S., FORSLUND, K., SZKLARCZYK, D., TRACHANA, K., ROTH, A., HUERTA-CEPAS, J., GABALDÓN, T., RATTEI, T., CREEVEY, C. & KUHN, M. 2013. eggNOG v4. 0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.*, gkt1253.

POWELL, S., SZKLARCZYK, D., TRACHANA, K., ROTH, A., KUHN, M., MULLER, J., ARNOLD, R., RATTEI, T., LETUNIC, I. & DOERKS, T. 2012. eggNOG v3. 0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.,* 40**,** D284-D289.

RABUS, R., RUEPP, A., FRICKEY, T., RATTEI, T., FARTMANN, B., STARK, M., BAUER, M., ZIBAT, A., LOMBARDOT, T., BECKER, I., AMANN, J., GELLNER, K., TEELING, H., LEUSCHNER, W. D., GLOCKNER, F. O., LUPAS, A. N., AMANN, R. & KLENK, H. P. 2004. The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. *Environ. Microbiol.,* 6**,** 887-902.

REAL, G., FAY, A., ELDAR, A., PINTO, S. M., HENRIQUES, A. O. & DWORKIN, J. 2008. Determinants for the subcellular localization and function of a nonessential SEDS protein. *J. Bacteriol.,* 190**,** 363-376.

RILEY, M., STALEY, J. T., DANCHIN, A., WANG, T. Z., BRETTIN, T. S., HAUSER, L. J., LAND, M. L. & THOMPSON, L. S. 2008. Genomics of an extreme psychrophile, Psychromonas ingrahamii. *BMC Genomics,* 9**,** 210.

RODRIGUES, D. F. & TIEDJE, J. M. 2008. Coping with our cold planet. *Appl. Environ. Microbiol.,* 74**,** 1677-1686.

RUDNER, D. Z. & LOSICK, R. 2010. Protein subcellular localization in bacteria. *Cold Spring Harb Perspect Biol,* 2**,** a000307.

SARIKHAN, S., AZARBAIJANI, R., YEGANEH, L. P., FAZELI, A. S., AMOOZEGAR, M. A. & SALEKDEH, G. H. 2011. Draft genome sequence of *Nesterenkonia* sp. strain F, isolated from Aran-Bidgol Salt Lake in Iran. *J. Bacteriol.,* 193**,** 5580.

SAYERS, E. W., BARRETT, T., BENSON, D. A., BOLTON, E., BRYANT, S. H., CANESE, K., CHETVERNIN, V., CHURCH, D. M., DICUCCIO, M. & FEDERHEN, S. 2011. Database resources of the national center for biotechnology information. *Nucleic Acids Res.,* 39**,** D38-D51.

SCHWARTZ, E., VAN HORN, D. J., BUELOW, H. N., OKIE, J. G., GOOSEFF, M. N., BARRETT, J. E. & TAKACS-VESBACH, C. D. 2014. Characterization of growing bacterial populations in McMurdo Dry Valley soils through stable isotope probing with 18O-water. *FEMS Microbiol Ecol. ,* 89**,** 415-425.

SEIFART GOMES, C., IZAR, B., PAZAN, F., MOHAMED, W., MRAHEIL, M. A., MUKHERJEE, K., BILLION, A., AHARONOWITZ, Y., CHAKRABORTY, T. & HAIN, T. 2011. Universal Stress Proteins Are Important for Oxidative and Acid Stress Resistance and Growth of *Listeria monocytogenes* EGD-e In Vitro and In Vivo. *PLoS One,* 6**,** e24965.

SGHAIER, H. 2011. DNA repair: Lessons from the evolution of ionizing-radiation-resistant prokaryotes-fact and theory. *In:* CHEN, C. (ed.) *Selected topics in DNA repair.* Rijeka, Croatia: INTECH, pp. 145-156.

SGHAIER, H., GHEDIRA, K., BENKAHLA, A. & BARKALLAH, I. 2008. Basal DNA repair machinery is subject to positive selection in ionizing-radiation-resistant bacteria. *BMC Genomics,* 9**,** 297.

SOLOVYEV, V. & SALAMOV, A. 2011. Automatic annotation of microbial genomes and metagenomic sequences. *In:* LI, R. W. (ed.) *Metagenomics and its applications in agriculture, biomedicine and environmental studies.* New York: Nova Science, pp. 61-78.

SRINIVASAN, G., JAMES, C. M. & KRZYCKI, J. A. 2002. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science,* 296**,** 1459-1462.

STACKEBRANDT, E. 2012. Genus VI. *Nesterenkonia*. *In:* WHITMAN, W., GOODFELLOW, M., KÄMPFER, P., BUSSE, H.-J., TRUJILLO, M., LUDWIG, W., SUZUKI, K.-I. & PARTE, A. (eds.) *Bergey's Manual of Systematic Bacteriology.* New York: Springer, pp. 636-642.

STACKEBRANDT, E., KOCH, C., GVOZDIAK, O. & SCHUMANN, P. 1995. Taxonomic dissection of the genus *Micrococcus*: *Kocuria* gen. nov., *Nesterenkonia* gen. nov., *Kytococcus* gen. nov., *Dermacoccus* gen. nov., and *Micrococcus* Cohn 1872 gen. emend. *Int. J. Syst. Bacteriol.,* 45**,** 682-692.

TATUSOVA, T., DICUCCIO, M., BADRETDIN, A., CHETVERNIN, V., CIUFO, S. & LI, W. 2013. *Prokaryotic genome annotation pipeline*. *In* The NCBI handbook [Internet]. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); 2013-. Available from: http://www.ncbi.nlm.nih.gov/books/NBK174280/.

TEHEI, M. & ZACCAI, G. 2005. Adaptation to extreme environments: macromolecular dynamics in complex systems. *Biochim. Biophys. Acta,* 1724**,** 404-10.

THANNICKAL, V. J. & FANBURG, B. L. 2000. Reactive oxygen species in cell signaling. *Am J Physiol Lung Cell Mol Physiol.,* 279**,** L1005-L1028.

THIERINGER, H. A., JONES, P. G. & INOUYE, M. 1998. Cold shock and adaptation. *Bioessays,* 20**,** 49-57.

UTRATNA, M., COSGRAVE, E., BAUSTIAN, C., CEREDIG, R. H. & O'BYRNE, C. P. 2014. Effects of Growth Phase and Temperature on Activity within a *Listeria monocytogenes* Population: Evidence for RsbV-Independent Activation of at Refrigeration Temperatures. *Biomed Res Int,* 2014.

VAN DOMSELAAR, G. H., STOTHARD, P., SHRIVASTAVA, S., CRUZ, J. A., GUO, A., DONG, X., LU, P., SZAFRON, D., GREINER, R. & WISHART, D. S. 2005. BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Res.,* 33**,** W455-W459.

VAN SCHAIK, W. & ABEE, T. 2005. The role of σB in the stress response of Gram-positive bacteria – targets for food preservation and safety. *Curr. Opin. Biotechnol.,* 16**,** 218-224.

VENTURA, M., CANCHAYA, C., TAUCH, A., CHANDRA, G., FITZGERALD, G. F., CHATER, K. F. & VAN SINDEREN, D. 2007. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol. Mol. Biol. Rev.,* 71**,** 495-548.

VINCENT, W. F. 2000. Evolutionary origins of Antarctic microbiota: invasion, selection and endemism. *Antarct. Sci.,* 12**,** 374-385.

WARGO, M. J., SZWERGOLD, B. S. & HOGAN, D. A. 2008. Identification of two gene clusters and a transcriptional regulator required for *Pseudomonas aeruginosa* glycine betaine catabolism. *J. Bacteriol.,* 190**,** 2690-2699.

WEISSBACH, H., RESNICK, L. & BROT, N. 2005. Methionine sulfoxide reductases: history and cellular role in protecting against oxidative damage. *Biochim. Biophys. Acta,* 1703**,** 203-212.

WHITE, O., EISEN, J. A., HEIDELBERG, J. F., HICKEY, E. K., PETERSON, J. D., DODSON, R. J., HAFT, D. H., GWINN, M. L., NELSON, W. C. & RICHARDSON, D. L. 1999. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science,* 286**,** 1571-1577.

WOOD, J. M. 2015. Bacterial responses to osmotic challenges. *J. Gen. Physiol.,* 145**,** 381-388.

YOON, J. H., JUNG, S. Y., KIM, W., NAM, S. W. & OH, T. K. 2006. *Nesterenkonia jeotgali* sp. nov., isolated from jeotgal, a traditional Korean fermented seafood. *Int. J. Syst. Evol. Microbiol.,* 56**,** 2587-92.

YU, C.-S., CHENG, C.-W., SU, W.-C., CHANG, K.-C., HUANG, S.-W., HWANG, J.-K. & LU, C.-H. 2014. CELLO2GO: a web server for protein subcellular localization prediction with functional gene ontology annotation. *PLoS One,* 9**,** e99368.

YU, C. S., CHEN, Y. C., LU, C. H. & HWANG, J. K. 2006. Prediction of protein subcellular localization. *Proteins,* 64**,** 643-651.

ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.,* 18**,** 821-9.

ZHANG, X., PENG, C., ZHANG, G. & GAO, F. 2015. Comparative analysis of essential genes in prokaryotic genomic islands. *Sci Rep,* 5.

ZHOU, Y., LIANG, Y., LYNCH, K., DENNIS, J. & WISHART, D. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res.,* 39**,** W347 - W352.

Chapter 3

3 Comparative genomic analyses of four *Nesterenkonia* species

3.1 Introduction

Genome wide comparison of the protein encoding genes, the corresponding proteins and other genomic features in closely related organisms is a valid strategy for identifying differences in their phenotypes (MacLean *et al.*, 2009, Qin *et al.*, 2014b). The principle is that differences in traits of distinct organisms, e.g., pathogenicity, host specificity and range, metabolic capacities, niche specialisation and survival strategies are driven by differences in the gene contents of the organisms. These difference maybe in the form of the presence or absence of certain genes coding for particular biological functions or sufficient variation (mutations) in genes to affect the function of the encoded proteins (De Maayer *et al.*, 2014a, Tettelin *et al.*, 2008).

Many methods have been employed to compare the genomes of closely related organisms. These methods generally involve determination of the core and the accessory genomes (Lefébure and Stanhope, 2007, Lefébure and Stanhope, 2009, Medini *et al.*, 2005) of the compared organisms. The accessory genome constitutes those genes that are present or absent in some of the compared organism as a result of loss or gain of genetic elements associated with genome reduction or lateral gene acquisition (Mira *et al.*, 2001, Tettelin *et al.*, 2008). The accessory genome can be queried among closely related organisms for the determination of niche specialization (Medini *et al.*, 2005, Mira *et al.*, 2001, Polz *et al.*, 2013). The core genome, on the other hand, includes genes that are generally associated with essential regulatory, replication and core metabolic functions and is stably maintained through speciation events. These conserved genes are used to determine long term evolutionary relatedness among the compared organisms as well as microevolution defining their adaptations in different environments (Mira *et al.*, 2001, Tettelin *et al.*, 2008). These microevolution events are driven by selective pressures in the environmental niche to which a microorganism must adapt, where positive selection involves the accumulation of beneficial mutations that confer fitness

advantages, while deleterious mutations are reduced by negative selection and neutral mutational events do not affect fitness (Anisimova *et al.*, 2001, Yang and Nielsen, 2002). Thus, for a comprehensive evaluation of adaptation strategies in any given set of related microorganisms, both the core and the accessory genomes are assessed.

The genus *Nesterenkonia* comprised of highly versatile haloakaliphilic Gram-positive bacteria, belonging to the family Micrococcaceae and phylum Actinobacteria (Rafael *et al.*, 2011). *Nesterenkonia* strains have been isolated in samples obtained from a wide range of environmental sources, including extreme environments. For instance, *Nesterenkonia* strains have been isolated from desert and saline soils (Li *et al.*, 2005, Li *et al.*, 2004, Li *et al.*, 2008), and from hyper-saline and soda lakes (Collins *et al.*, 2002, Delgado *et al.*, 2006).

The draft genomes of four *Nesterenkonia* strains have been sequenced. *Nesterenkonia* sp. AN1 is the first cold-adapted strain of the genus, for which the genome sequence has been determined (Aliyu *et al.*, 2014). The genome sequence revealed a repertoire of protein-encoding genes that can be associated with several extremophilic adaptations (Chapter 2). The draft genome sequences of two mesophilic isolates, *Nesterenkonia* sp. F (Sarikhan *et al.*, 2011) and *Nesterenkonia* sp. NP1 (Edouard *et al.*, 2014) have also been published. The former was isolated from a hyper-saline lake in Iran (Sarikhan *et al.*, 2011), while the latter was derived from human faeces obtained from an HIV-infected individual in France (Edouard *et al.*, 2014). Furthermore, the genome sequence data for *N. alba* DSM 19423, isolated from cotton pulp mill in China (Luo *et al.*, 2009), are also publically available.

Here, we have performed holistic comparative genomic analyses of both the core and accessory genomes of *Nesterenkonia* sp. AN1 and the three mesophilic *Nesterenkonia* isolates with a view to elucidate the genetic determinants underlying the ability of *Nesterenkonia* sp. AN1 to survive in Antarctic soils.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

3.2 Materials and Methods

3.2.1 Genome sequences and structural annotation

The draft genome sequences of *Nesterenkonia* sp. AN1 and those of three mesophilic *Nesterenkonia* strains were included in this study (Table 3-1). Structural annotations were standardized by using a single gene calling application, FgenesB (Solovyev and Salamov, 2011). The predicted protein coding sequence (CDS) sets for each genome were further standardized through a pair-wise local BlastP and BlastN analysis using Bioedit v 7.2.5, thereby repairing protein coding gene sequences with frame-shift mutations resulting in premature stop-codons (Hall, 1999). The mean genomic G+C contents were also determined using Bioedit. The tRNA coding sequences in the draft genomes of the four *Nesterenkonia* strains were predicted using the ARAGORN web-server (Laslett and Canback, 2004). To assess the pattern of amino acid usage, the amino acid composition of the predicted proteins in the four strains was computed using Bioedit v 7.2.5 (Hall, 1999).

3.2.2 Phylo-genomic analyses

The draft genomes were uploaded on the CVTree web server (Qi *et al.*, 2004, Xu and Hao, 2009) to infer a whole genome phylogeny (Qi *et al.*, 2004, Xu and Hao, 2009). Average amino acid identity (AAI) values were calculated for the proteins shared by the four *Nesterenkonia* strains (core proteins) using Bioedit v 7.2.5 (Hall, 1999, Hall, 2011). The core protein sets, determined using the orthology parameters discussed in section 3.2.3, were pair-wise compared by localized BlastP and the average amino acid identity values calculated using the formula: SUM[(amino acid identity for each orthologous protein x alignment length)/100]/ number of orthologous proteins. To construct the AAI phylogeny, AAI values for each of the pair-wise compared genomes were converted to distance values using the formula: 1-(AAI/100). The distance matrix was uploaded and analysed using PHYLIP (Felsenstein, 1993, Retief, 1999).

The resultant Neighbour-joining tree was visualised using MEGA 6 (Tamura *et al.*, 2013). *In silico* digital 'DNA-DNA hybridization' (dDDH) values were computed using the online 'Genome-to-Genome Distance Calculator' GGDC 2.0 (Auch *et al.*, 2010, Meier-Kolthoff *et al.*, 2013). Average nucleotide identity was computed using JSpeciesWS (Richter et al., 2015). ANI values were determined using the BLAST method (ANIb) proposed by Goris *et al.* (2007). To validate the genus level assignment of the four *Nesterenkonia* strains, the "percentage of conserved proteins (POCP)" proposed by Qin *et al.* (2014) was computed. POCP was calculated as $[(C^1+C^2)/(T^1+T^2)] \times 100$, where C^1 and C^2 represents the number of conserved proteins (E-value < $1e^{-5}$, alignment coverage > 50 % and amino acid identity value > 40 %) in a pair of genomes and T^1 and T^2 represent the total number of proteins predicted in the two genomes (Qin *et al.*, 2014a).

3.2.3 "Pan-genome" analysis

A Reciprocal Best Blast Hit (RBBH) approach was used to determine orthologs between the standardised protein sets for each pair of strains (Moreno-Hagelsieb and Latimer, 2008). Orthologs were assigned on the basis of amino acid identity and sequence coverage cut-off values of > 30 % and > 70 %, respectively. Orthologous CDSs that are present in all four strains were assigned to the core genome. The remaining CDSs common to only two or three strains and those unique to only one strain were assigned to the appropriate accessory genome fractions. Functional annotation for each of the core and accessory proteins in the pan-genome was subsequently computed by localised BLASTP analysis against the Actinobacteria non-supervised orthologous groups (actNOG) multiple alignment obtained from the 'Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups' (eggNOG v3.0) database (Powell *et al.*, 2012). Core and accessory proteins were further functionally classified into their 'clusters of orthologous groups' (COG) using the WebMGA server (Wu *et al.*, 2011).

3.2.4 Detection of positive selection

3.2.4.1 Determination of Orthologous gene sets

Orthologous genes for all four genomes were determined using Proteinortho V5.11 (Lechner *et al.*, 2011). The program detects orthologs by implementing an improved reciprocal best hits heuristic algorithm. The identified orthologous gene families of the four genomes were extracted using an *ad hoc* perl script[1]. We implemented the T_COFFEE (v.11.00.8cbe486) (Magis *et al.*, 2014, Notredame *et al.*, 2000) alignment program to generate high quality multiple sequence alignments. The program combines different alignment algorithms such as K-align (Lassmann and Sonnhammer, 2005), MAFFT (Katoh and Standley, 2014), MUSCLE (Edgar, 2004) and T-COFFEE (Magis *et al.*, 2014, Notredame *et al.*, 2000) and then integrates the different alignment scores to generate the optimal alignment. The quality of the alignments were assessed with MaxAlign (Gouveia-Oliveira *et al.*, 2007) to identify high quality alignments for further downstream analyses (Taly *et al.*, 2011). Furthermore, to improve alignment quality and reduce false positive signals that may arise from gap rich segments of the alignments, the stringent gap removal option of Gblocks v.0.91b program was applied (Castresana, 2000, Talavera and Castresana, 2007).

3.2.4.2 Detection of recombination and positive selection

Recombination signals were detected using the PhiPack program, using default parameters (Bruen and Bruen, 2005, Su *et al.*, 2013). The program computes p-values as a measure of recombination signatures on multiple sequence alignments from ‘pairwise homoplasy index’ (Phi), ‘Maximum chi-square’ (Max $\chi 2$) and ‘neighbor similarity score’ (NSS). The p-values obtained were corrected (recombination and all subsequent analyses) for multiple testing using

[1]Perl script graciously provided by Dr Luke T Dunning of Imperial College London

BH (Benjamini and Hochberg, 1995) correction and a significance p-value of 0.5 with the statistical package R 3.0.1 (http://www.R-project.org). The orthologous genes were tested for imprints of positive selection using the branch-site and site test model of the Codeml program implemented in Paml4.8 (Yang, 2007). For the site test, the neutral model M7 (β) was compared to M8 (β and ω) alternative model with an additional $\omega > 1$. In contrast, the branch-site test involved sequential assignment of foreground branches and comparison to the other branches to which the null model was applied (Hughes and Friedman, 2008). In both models, likelihood ratio test (LRT) was computed using the formula $LRT = 2\Delta l$, where Δl is the difference between the selection and null models. A $\chi 2$ distribution test was used to compute p-values for positive selection with degrees of freedom 1 and 2 for the branch-site and site test respectively. BH correction for multiple testing was applied as described above. Protein sequences for all genes with signatures of positive selection were functionally annotated against the clusters of orthologous groups (COG) using WebMGA server (Wu *et al.*, 2011) which implements the 'Reversed Position Specific BLAST' (RPS-BLAST) program. In addition, proteins that did not return any hits against the COG database were functionally annotated using eggNOG v3.0 (Powell *et al.*, 2012). The proteins were also mapped against the GO database using the trial version of Blast2GO (Conesa *et al.*, 2005, Götz *et al.*, 2008). GO annotations for all the proteins were set as the background against which the GO annotations of the positively selected genes families were queried to test for enrichment of particular functions. Functional category enrichment analysis was computed based on the Fisher's exact test implemented via GOEAST (Zheng and Wang, 2008).

3.3 Results and Discussion

3.3.1 General genome metrics

The genomic sequence metrics for the four *Nesterenkonia* strains included in this study are summarised in Table 3-1. The genomes range in size from 2.59 to 3.04 Mb, with G+C contents of 62.2-71.5 %. *Nesterenkonia* sp. AN1 has the largest genome (~ 3.04 Megabases) with a mean G+C content of 67.4 % and 2,853 protein coding sequences (Appendix F 1) (Aliyu et al., 2014). The precise determinant of the differences in genome size between the four strains of *Nesterenkonia* is not clear, although the larger genome size of *Nesterenkonia* sp. AN1 may, in part, be explained by the horizontal acquisition of genomic islands (Chapter 2, section 2.4.3) which may confer selective advantages for survival of this strain in the cold Antarctic desert soils.

The number and variety of tRNA encoding genes as well as patterns of amino acid usage have been used to distinguish psychrophilic microorganisms from their mesophilic relatives (De Maayer *et al.*, 2014a). In general, proteins in psychrophilic bacteria are enriched in amino acids that are associated with the increased conformational flexibility required for proper enzyme functioning in cold environments (Metpally and Reddy, 2009). There is also a corresponding reduction in those amino acids which impose conformational restrictions (Siddiqui *et al.*, 2013). Empirical data from whole genome studies suggests that those amino acids which do not preferentially form helices, such as Alanine, Aspartic acid, Glycine, Serine and Tyrosine are preferred in psychrophiles. Conversely, those favouring helix formation, such as Glutamine, Phenylalanine, Lysine and Leucine, are less favoured (Metpally and Reddy, 2009). A comparison of the proteome of *Nesterenkonia* sp. AN1 and those of the three mesophilic *Nesterenkonia* strains revealed similar patterns of amino acid usage in all four strains. There are high proportions of both the preferred amino acids Alanine, Aspartic acid, Glycine, Serine and Tyrosine and the un-favoured amino acids Glutamine and Leucine (Figure 3-1a). These

results are in agreement with findings from comparative genomic analyses of psychrophilic *Arthrobacter* strains and their temperate relatives (Dsouza *et al.*, 2015). The high proportion of both the favoured and the un-favoured amino acids in the proteins of psychotolerant bacterial isolates from soils in Antarctica, is perhaps an important adaptation strategy for coping with the daily and seasonal fluctuations in temperature reported in Antarctic Dry Valley soils (Aislabie *et al.*, 2006, Dreesens *et al.*, 2014).

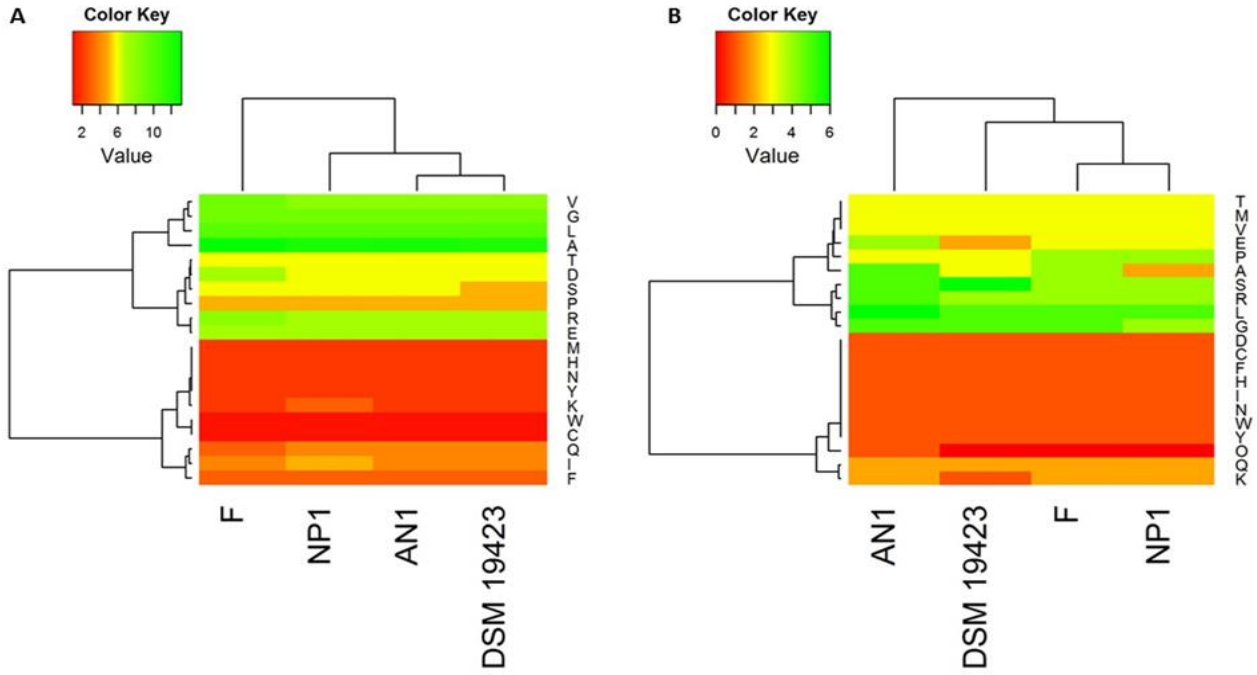


Figure 3-1: Heat maps showing (A) Comparison of amino acids proportions in proteins of four strains of *Nesterenkonia*. (B) Abundance of tRNA genes in the four compared *Nesterenkonia* strains. A –Y represent the standard amino acids single letter codes.

The genomes of psychrophilic microorganisms surviving under low temperature conditions typically encode a larger number and variety of transfer RNAs (tRNAs) (De Maayer *et al.*, 2014a). These have been reported to serve as an adaptation to counteract the cold imposed slow mobilization of tRNA species to translation sites (Math *et al.*, 2012, Satapathy *et al.*, 2010, Sharp *et al.*, 2005). Comparison of the tRNA encoding genes (Laslett and Canback, 2004), showed that the *Nesterenkonia* sp. AN1 genome encodes a larger number of tRNAs than its

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

mesophilic counterparts (Figure 3-1b, Table 3-1). Furthermore, a gene that codes for pyrrolysine tRNA ($tRNA^{Pyl}$) is found exclusively in *Nesterenkonia* sp. AN1. . Pyrrolysine is the 22nd proteinogenic amino acid, encoded by the stop codon UAG, and is essential for methanogenesis from methylamines (Ambrogelly *et al.*, 2007, Have *et al.*, 2013, Gaston *et al.*, 2011b). So far, this amino acid has only been reported in the genomes of six Archaea and six bacteria (Gaston *et al.*, 2011a, Srinivasan *et al.*, 2002).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 3-1: General genome features of the compared *Nesterenkonia* strains. The isolation source and optimum growth temperatures (Topt), number of assembled contigs and the corresponding genome size, percent G+C content, number of CDSs and transfer RNA (tRNAs) encoded on the genomes are given. Furthermore, the number of genomic islands (GIs) detected on each of the genome, total size of the GIs and number of genes contributed by the GIs are indicated.

Strain ID	GenBank ID	Source	Topt.	Nº of Contigs	Genome size (Mb)	G + C (%)	Nº of CDSs	Nº of tRNA	Nº of GIs	GI size (Kb)	Nº of genes
AN1	JEMO00000000	Antarctic dry Soil	21°C	41	3.05	67.4	2852	55	16	166	169
DSM 19423	ATXP00000000	Cotton pulp mill	42°C	36	2.59	63.8	2395	50	11	162	164
F	AFRW00000000	Hyper-saline lake	-	138	2.81	71.5	2543	47	5	53	76
NP1	CBLL000000000	Human faeces	37°C	175	2.67	62.2	2523	48	10	117	126

3.3.2 Phylo-genomics of *Nesterenkonia* spp

The whole genome phylogeny of the four *Nesterenkonia* strains and other members of the family *Micrococcaceae* (Qi *et al.*, 2004, Xu and Hao, 2009) revealed that the four *Nesterenkonia* strains form a distinct clade within the family (Figure 3-2). The average amino acid identity (AAI) of the four *Nesterenkonia* spp. ranges between 67 and 72 %. The Neighbour-joining phylogeny built from the AAI distance matrices of the four *Nesterenkonia* spp. and related organisms in the family *Micrococcaceae* also clustered the four *Nesterenkonia* as a distinct group in the family (Figure 3-3). AAI values > 95 % have been shown to correspond with the > 70 % threshold DNA-DNA hybridisation threshold required for the description of new species (Konstantinidis and Tiedje, 2005a, Konstantinidis and Tiedje, 2005b). The digital DNA-DNA hybridization (dDDH) values between *Nesterenkonia* sp. AN1 and genomes of the mesophilic *Nesterenkonia* (Figure 3-4) are low (24.4 – 26.8 %). Furthermore, the average nucleotide identity (ANI) (Goris *et al*., 2007) values for the four *Nesterenkonia* genomes ranged from 72.91 to 75.73 % (Figure 3-4). These ANI values are below the 95 - 96 % cut-off for the circumscription of bacterial species (Richter and Rosselló-Móra, 2009). The combination of the CVtree whole genome phylogeny, the AAI phylogeny, dDDH and ANI therefore suggests that the four strains included in this study belong to distinct species within the genus *Nesterenkonia.* The grouping of the four genomes in the genus *Nesterenkonia* is further supported by the percentage of conserved proteins (POCP) (Qin et al., 2014a). Pairwise comparison of the four genomes showed POCP greater than the 50 % recommended threshold (Figure 3-5) for assigning bacterial strains to the same genus (Qin *et al*., 2014a). A major constraint to the delineation of the sequenced *Nesterenkonia* sp. is the lack of genome data from the type strains. On the basis of a 16S rRNA phylogeny *Nesterenkonia* sp. AN1 clusters with the type strains of the species *N. halotolerans, N. sandarakina, N. jeotgali,* and *N. lutea* (Nel *et al.*, 2011). Additional genomic data, particular of these type strains

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

may enable the phylo-genomic delineation of the Antarctic *Nesterenkonia* sp. AN1 to the species level.

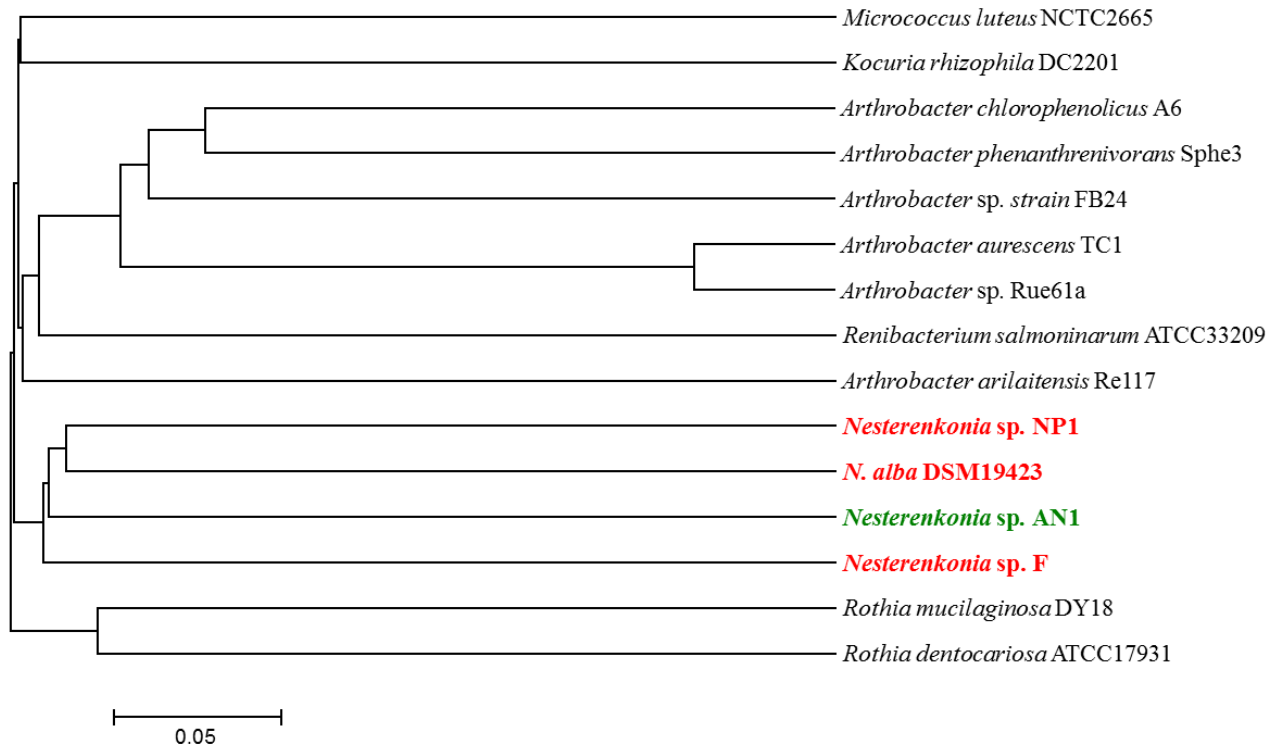


Figure 3-2: Whole genome phylogeny of the four *Nesterenkonia* strains for which genomes are available and the genomes of other members of the family Micrococcaceae. The phylogenetic relationship was inferred using the CVTree web tool, and the neighbour-joining tree was constructed using the MEGA6 program.

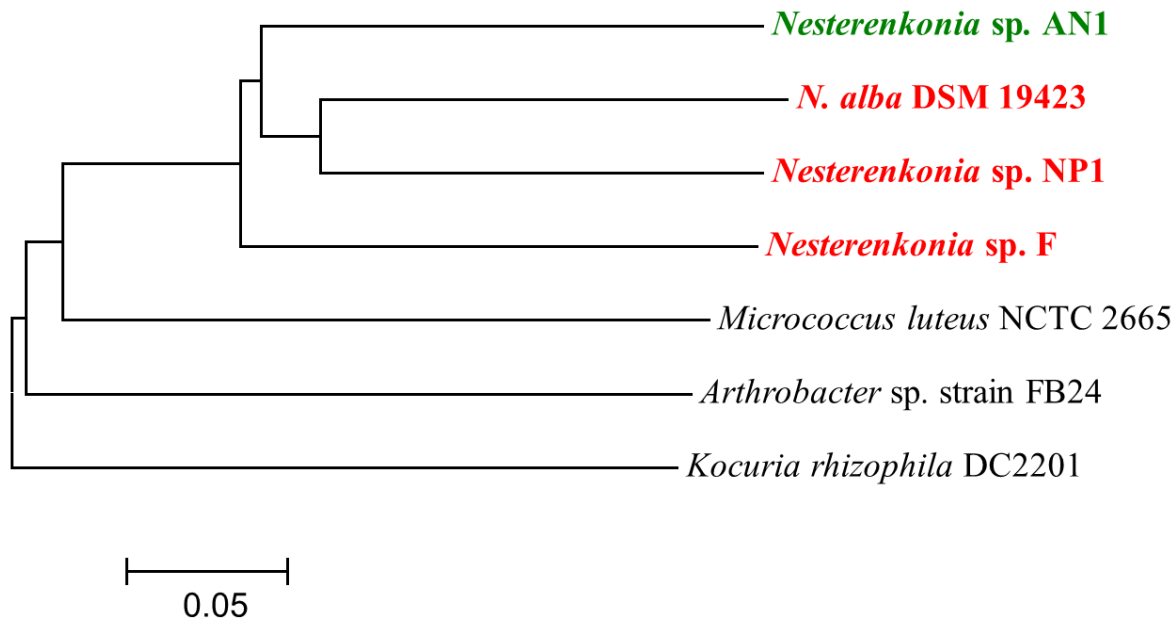


Figure 3-3: Phylogeny constructed on the basis of the average amino acid identity (AAI) values for the core proteins of the four *Nesterenkonia* strains and selected members *Micrococceae* using MEGA6 program. The genome of *Kocuria. Rhizophila* DC2201was used as an out group to root the tree.

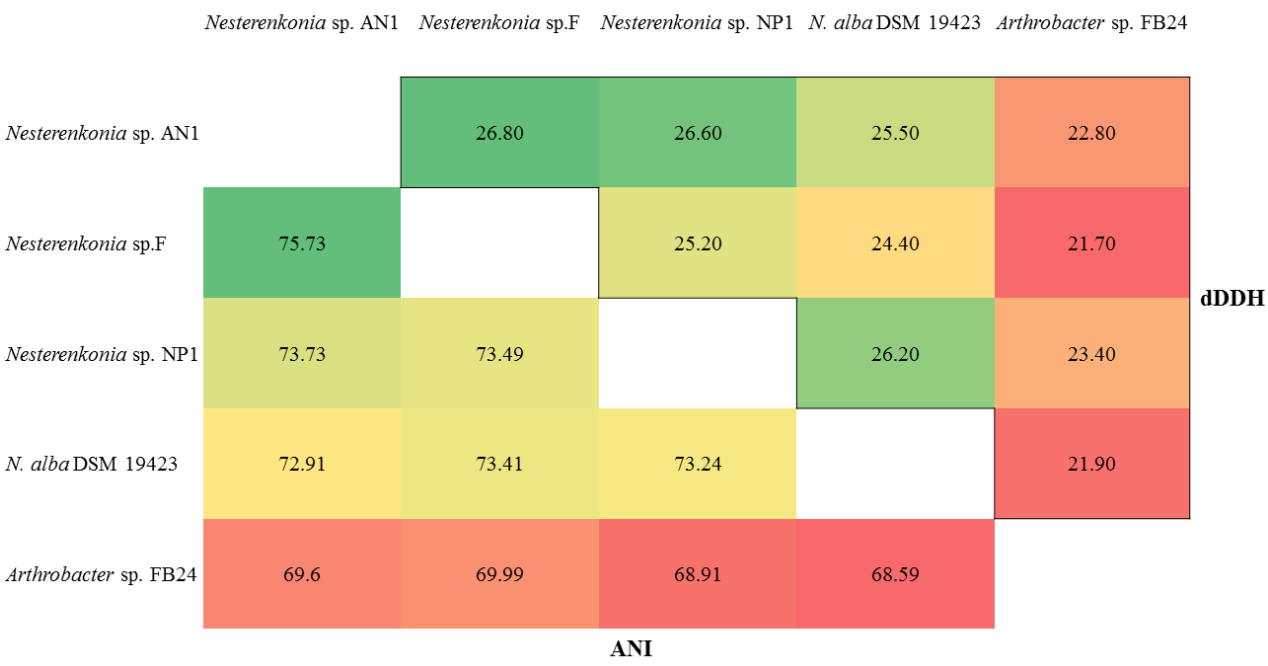


Figure 3-4: ANI and dDDH among the four *Nesterenkonia* strains and *Arthrobacter* sp. FB24. The upper triangle shows the dDDH values and the lower triangle shows the ANI values.

	***Nesterenkonia* sp. AN1**	***Nesterenkonia* sp.F**	***Nesterenkonia* sp. NP1**	***N. alba* DSM 19423**	***Arthrobacter* sp. FB24**
***Nesterenkonia* sp. AN1**	**100.000**	**61.316**	**59.926**	**55.746**	**43.382**
***Nesterenkonia* sp.F**	**61.316**	**100.000**	**58.409**	**57.918**	**41.959**
***Nesterenkonia* sp. NP1**	**59.926**	**58.409**	**100.000**	**61.183**	**38.531**
***N. alba* DSM 19423**	**55.746**	**57.918**	**61.183**	**100.000**	**37.112**
***Arthrobacter* sp. FB24**	**43.382**	**41.959**	**38.531**	**37.112**	**100.000**

Figure 3-5: POCP of four *Nesterenkonia* strains and *Arthrobacter* sp. FB24. Pair of bacterial genomes with POCP value > 50 % have been proposed to belong the same genus (Qin et al., 2014a).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

3.3.3 The 'pan-genome' of four *Nesterenkonia* strains

Comparative analysis of the proteins encoded on the genomes of the *Nesterenkonia* strains revealed that the pan-genome of these four strains is comprised of 4,818 non-orthologous proteins. These include 1,332 proteins (28 % of the total pan-genome proteins) that are common to all four strains (core genome) as well as an extensive accessory genome of 3,486 proteins (72 % of the total pan-genome proteins) (Figure 3-5). The core genome contributes ~ 47 % of the total proteins in *Nesterenkonia* sp. AN1 and 52 – 56 % in the other strains. About 70 % of proteins in the accessory genome are strain-specific (Figure 3-5) and this could be indicative of extensive acquisition of genes via horizontal gene transfer (Tettelin *et al.*, 2005) which might have played a crucial role in niche specialisation of each of the four strains. Progressive addition of the third and fourth genome sequences during the computation of the pan-genome yielded 670 and 570 new genes, respectively. The incorporation of novel genes in the pan-genome upon addition of new genome sequences suggests that the genus *Nesterenkonia* possesses an open pan-genome. An open pan-genome has been demonstrated to be characteristic of ecologically diverse taxa in which the organisms acquire a substantial portion of their genomes laterally (Abby and Daubin, 2007, De Maayer *et al.*, 2014b). However, more genome sequences need to be incorporated in order to ascertain a more accurate picture of the pan-genome of the genus *Nesterenkonia*.

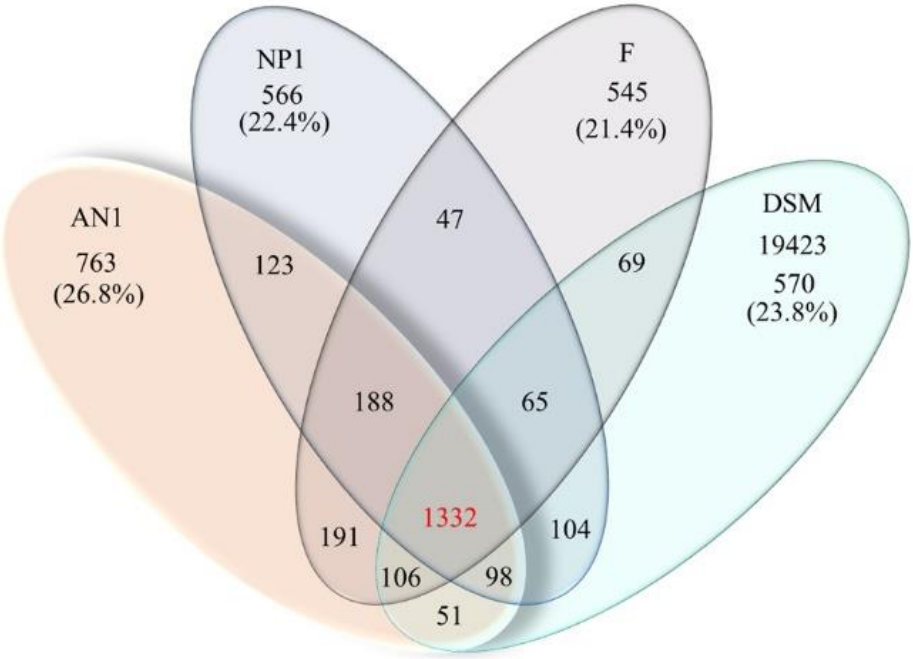


Figure 3-6: Venn diagram representation of the comparison between sequenced genomes of the genus *Nesterenkonia* based on orthologous relationships among the predicted CDSs. AN1, NP1, F and DSM 19423 represent the genome compartments of *Nesterenkonia* sp. AN1, *Nesterenkonia* sp. NP1,*Nesterenkonia* sp. F and *N. alba* DSM 19423, respectively.

To determine the functional roles of the proteins encoded within the different fractions of the pan-genome of four *Nesterenkonia* strains, the proteins were queried against cluster of orthologous (COG) and the Actinobacteria non-supervised orthologous groups (actNOG) databases. The results revealed that approximately 84 % of the proteins could be assigned to functional categories (Figures 3-6). The remainder, ~ 16 % of the proteins in the pan-genome, have no orthologs in either the COG or the actNOG databases and are therefore categorised as "proteins of unknown function". Over 90 % of the 773 proteins of unknown function are unique to a particular strain and thus contribute extensively to the unique genome fractions of the individual strains.

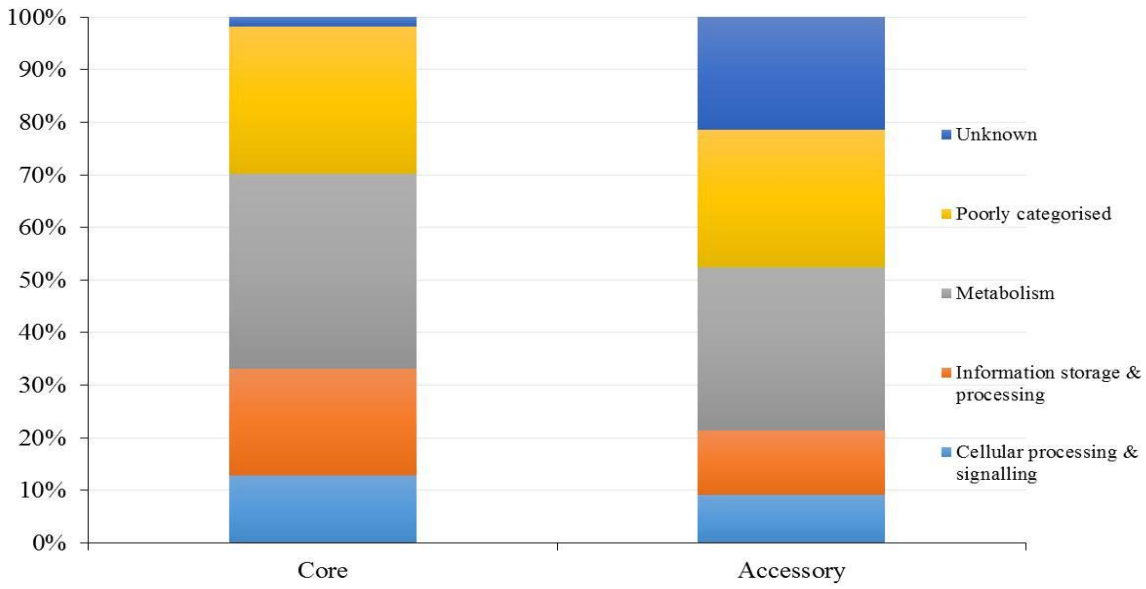


Figure 3-7: Proportion of proteins CDSs in the *Nesterenkonia* core and accessory genome fractions involved in the super functional categories in the COG and EGGNOG databases.

Analyses of the functional distribution of the 4, 045 proteins which have significant hits with characterised proteins in the COG and EGGNOG databases showed that the majority of the proteins, both in the accessory and core genome fractions, belong to the super-functional categories 'metabolism' (~ 33 %) and 'poorly categorised' (~ 27 %). Approximately 31 % of the accessory proteins and 37 % of the core proteins were clustered in the super-functional category 'metabolism' (Figure 3-6). These results represent a marked departure from results of several recent comparative genomic studies where it was reported that majority of the accessory proteins belong to the 'poorly characterised' functional group (De Maayer *et al.*, 2014b, Ozer *et al.*, 2014, Zhang and Sievert, 2014).

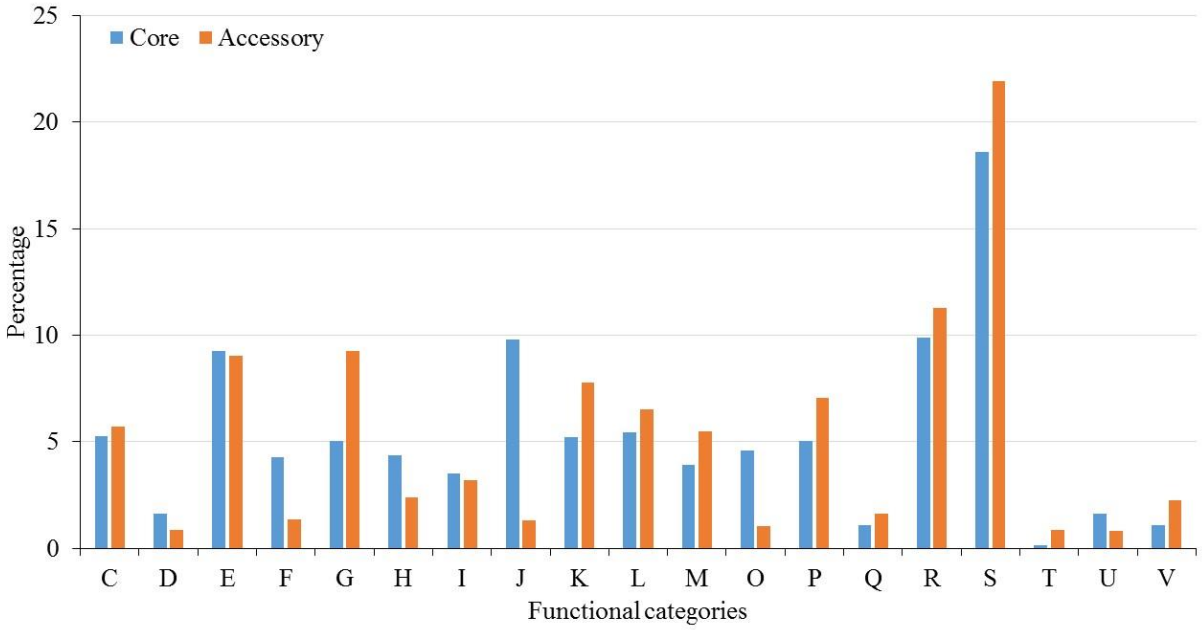


Figure 3-8: Functional categories of core and accessory proteins in the four compared *Nesterenkonia* strains. The COG categories are defined as follows: "translation, ribosomal structure and biogenesis (J), transcription (K), replication, recombination and repair (L) cell cycle control, cell division, chromosome partitioning (D), defence mechanisms (V), signal transduction mechanisms (T), cell wall/membrane/envelope biogenesis (M), intracellular trafficking, secretion, and vesicular transport (U) and posttranslational modification, protein turnover, chaperones (O). energy production and conversion (C), carbohydrate transport and metabolism (G), amino acid transport and metabolism (E), nucleotide transport and metabolism (F), coenzyme transport and metabolism (H), lipid transport and metabolism (I), inorganic ion transport and metabolism (P) and secondary metabolites biosynthesis, transport and catabolism (Q). POORLY CHARACTERIZED; general function prediction only (R) and function unknown (S)".

Furthermore, the core genome was enriched with proteins associated with 'coenzyme and nucleotide metabolism' and transport', 'translation' and 'posttranslational modification' relative to the accessory genome which contain higher proportions of proteins in the functions 'carbohydrate and inorganic ion metabolism and transport' when compared with the core proteins (Figure 3-7). Comparative genomic analyses of the psychrophilic genera *Glaciecola* (Qin *et al.*, 2014b) and *Shewanella* (Konstantinidis *et al.*, 2009) revealed similar enrichment of the core genome but the relative proportions of the accessory proteins were different from those reported in this work. Since their divergence, the four *Nesterenkonia* strains might have

acquired or lost a cohort of genes associated with specialised metabolic pathways as they colonised new environments. For instance, evaluation of the proteins (aside the poorly characterised) within accessory genomes of the individual strains revealed large proportions of the proteins corresponding to 'amino acid transport and metabolism' in all the strains, 'carbohydrate transport and metabolism' in *Nesterenkonia* sp. F and NP1 and *N. alba* DSM 19423; 'inorganic ion transport and metabolism' in *Nesterenkonia* sp. AN1; 'DNA replication in *Nesterenkonia* sp. NP and 'transcription' in *Nesterenkonia* sp. AN1,*Nesterenkonia* sp. F and *N. alba* DSM 19423 (Figure 3-8).

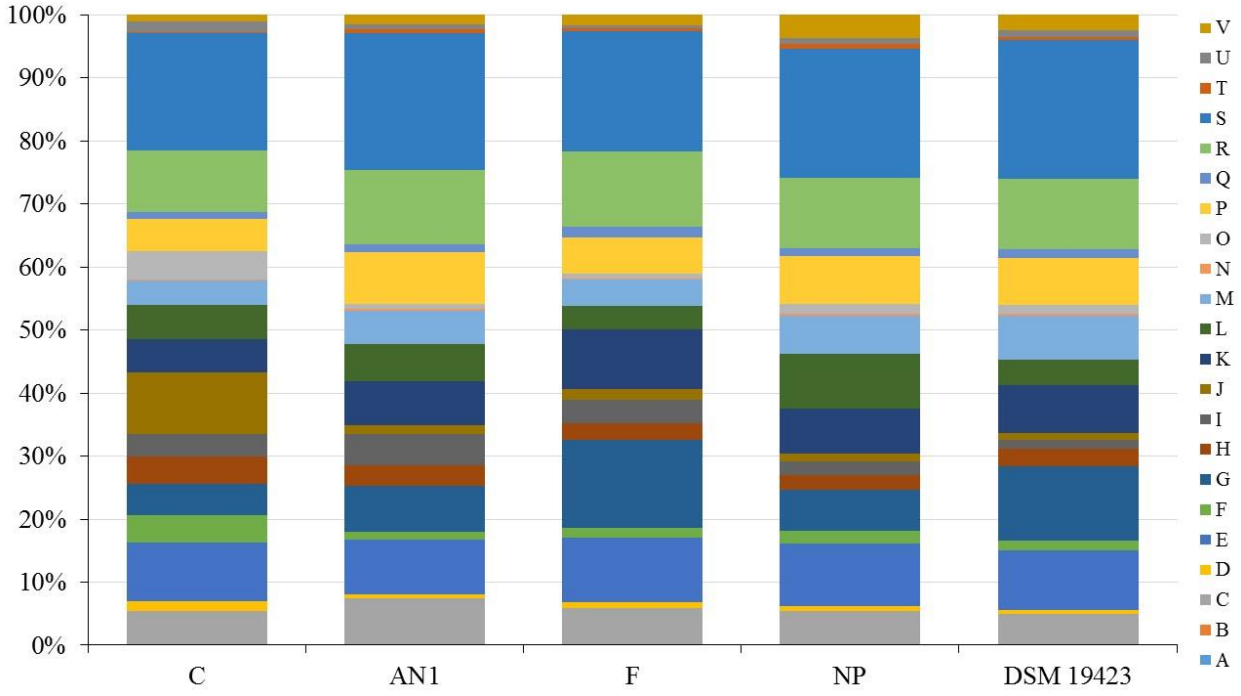


Figure 3-9: Proportion of proteins in the core and accessory genomes of four *Nesterenkonia* strains characterised based using COG and EGGNOG. C represents the orthologous genes shared (core) while AN1, F, NP1 and DSM 19423 represent the accessory genes in *Nesterenkonia* sp. AN1, *Nesterenkonia* sp. F, *Nesterenkonia* sp. NP1 and *N.* alba DSM 19423, respectively. The COG categories have been defined in Figure 3-7.

3.3.3.1 Comparative genomics identifies unique stress response proteins in Nesterenkonia sp. AN1

Comparisons of the four strains in terms of the 189 adaptive genes/proteins identified in *Nesterenkonia* sp. AN1 (Chapter 2) showed that ~ 58 % were common to all four strains (Figure 3-9). The common features include nineteen out of the twenty cold shock and acclimation proteins, eighteen out of the nineteen DNA repair proteins and the majority of the proteins involved in osmotic and oxidative stress response. These findings indicate that members of the genus *Nesterenkonia* may be naturally resilient to multiple environmental stresses and potentially capable of adapting to multiple stress conditions. These results agree with findings from studies comparing several Antarctic strains of *Arthrobacter* and *Paenibacillus* with their respective temperate relatives (Dsouza *et al.*, 2014, Dsouza *et al.*, 2015).

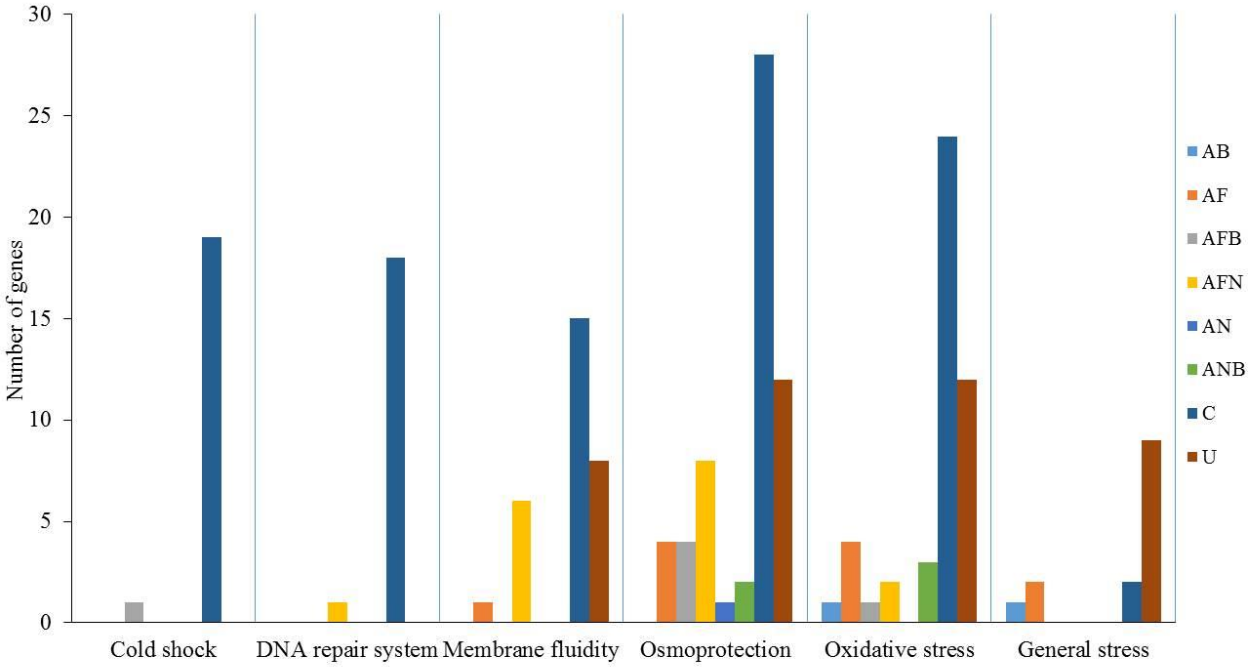


Figure 3-10: Number of adaptation genes associated with different pan-genome fractions. (U: *Nesterenkonia* sp. AN1 unique fraction, C: core, A = *Nesterenkonia* sp. AN1, B: *N. alba* DSM 19423, N: *Nesterenkonia* sp. NP1 and F: *Nesterenkonia* sp. F).

The strain-specific fraction of the *Nesterenkonia* sp. AN1 genome encodes a large number of proteins that are putatively involved in diverse stress response systems (Figure 3-9). These include proteins associated with modulation of membrane fluidity at low temperature, such as the carotenoid proteins, phytoene dehydrogenase (CrtN) and lycopene elongase (CrtEb) as well as six copies of 3-oxoacyl-[acyl-carrier protein] reductase (FabG) (Chintalapati *et al.*, 2004). The unique genome fraction in *Nesterenkonia* sp. AN1 also contains an additional locus that codes for the Ktr potassium uptake system and several genes encoding transporters of glycine/betaine/proline which are known systems for the accumulation of compatible solutes (Kappes *et al.*, 1999, Mykytczuk *et al.*, 2013). Many of these genes reside on genomic islands identified in the genome of *Nesterenkonia* sp. AN1 (Chapter 2) and, as such, may have been acquired through lateral gene transfer events (Juhas *et al.*, 2009).

The unique fraction of the *Nesterenkonia* sp. AN1 genome also encodes proteins putatively involved in oxidative stress tolerance which include the scavengers of reactive oxygen species, protocatechuate 3, 4-dioxygenase beta chain (PcaH), protocatechuate 3, 4-dioxygenase alpha chain (PcaG) and 4-hydroxyphenylpyruvate dioxygenase (HPPH) (D'Argenio *et al.*, 1999, Raspail *et al.*, 2011). The strain-specific genomic proteins encoded by *Nesterenkonia* sp. AN1 also include three sigma B regulatory proteins (RsbV, RsbW, RsbU) and four of the seven copies of the universal stress response protein (UspA) gene (Foster, 2005, Garnier *et al.*, 2010, Marles-Wright and Lewis, 2007, Utratna *et al.*, 2014).

Nesterenkonia sp. AN1 stands out among the four *Nesterenkonia* spp. in terms of unique genomic elements which potentially play a role in coping with oxidative stress and modulation of membrane fluidity, both of which are crucial for survival under extreme cold condition. Microorganisms inhabiting cold environments are especially constrained by oxidative stress as a result of an increase in the solubility of gasses at cold temperature (Chattopadhyay, 2006,

D'Amico *et al.*, 2006). Organisms living in cold environments are also subject to cold imposed rigidity of the lipid membrane (D'Amico *et al.*, 2006, Feller, 2003, Feller and Gerday, 2003).

3.3.4 Signatures of positive selection are prevalent in the core genome of *Nesterenkonia*

3.3.4.1 Orthologs sets and alignment strategy

A total of 1,235 orthologous genes were identified using Proteinortho program (Lechner *et al.*, 2011), which correlates well with the *Nesterenkonia* core genes previously predicted using the RBBH approach above (1,332 core proteins). Of these, 1,092 gene families (88.4 %) were selected based on the alignment quality filter implemented in T-COFFEE (Magis *et al.*, 2014, Notredame *et al.*, 2000). The whole genome consensus phylogeny of *Nesterenkonia* (Figure 3-10) was, however, inferred from 1,108 concatenated core genes of the four *Nesterenkonia* strains and *Arthrobacter* sp. FB24 (included as an out-group). The tree was applied in subsequent analysis to identify specific signatures of positive selection in the individual strains.

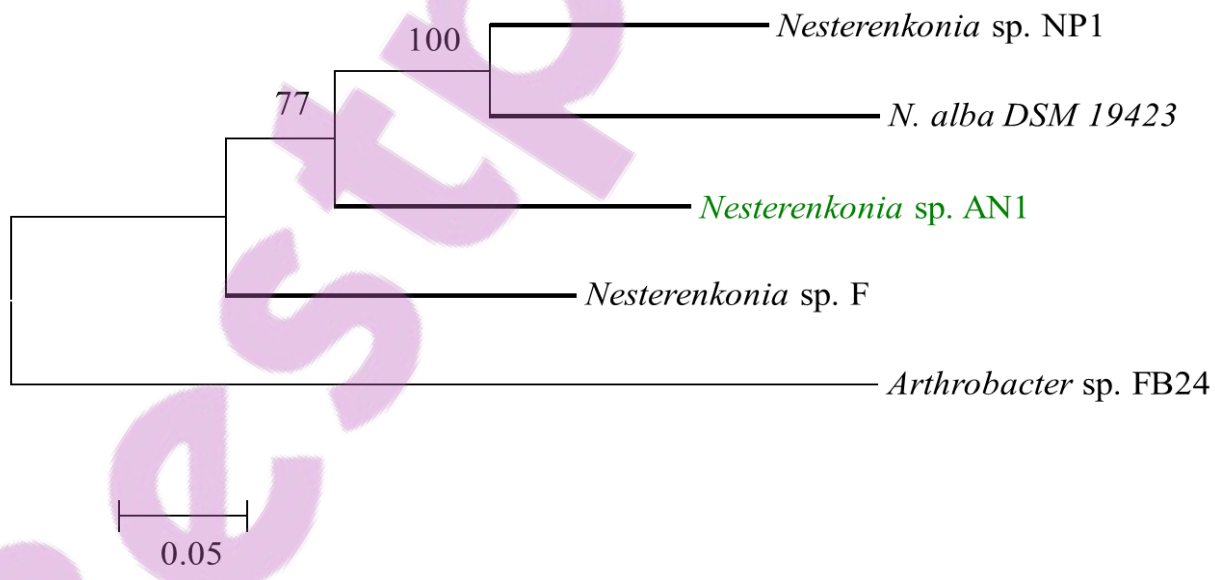


Figure 3-11: Phylogeny of four *Nesterenkonia* strains. The maximum likelihood tree was computed from 1,108 core genes of *Nesterenkonia* spp and *Arthrobacter* sp. FB24 trimmed using Gblocks0.9b.

3.3.4.2 Recombination and positive selection

Homologous recombination is an important evolutionary force. Recombination has been identified to be the major driving force in the microevolution of environmental strains of pathogenic bacteria (Keymer and Boehm, 2011). Recombination could therefore interfere with the accuracy of phylogenetic inferences, particularly in testing for positive selection in the genome sequences of related strains. This partly arises from the violation of the unique species tree topology assumed under positive selection because recombination introduces incongruent trees within and between sequences (Lefébure and Stanhope, 2007, Lefébure and Stanhope, 2009). We applied the program PhiPack to identify genes showing recombination events. A total of 160 gene sets, representing ~ 15 % of the core genes included in this analysis, were determined to contain significant (False Discovery Rate (FDR) adjusted p- value ≤ 0.05) recombination events by at least one of the programs included in PhiPack. The proportions of recombinants detected by Phi, χ2 and NSS were 0.4, 0.3 and 0.2 respectively. Of the 160 genes with signs of recombination, only 6 genes, less than 1 % of the core genes, were adjudged to be recombinant by the three substitution methods (Phi, χ2 and NSS) (Lefébure and Stanhope, 2007) at the same time (Figure 3-11). This implies that recombination was not a significant evolutionary driver of the molecular adaptations within the core genome of the *Nesterenkonia* strains included in this study.

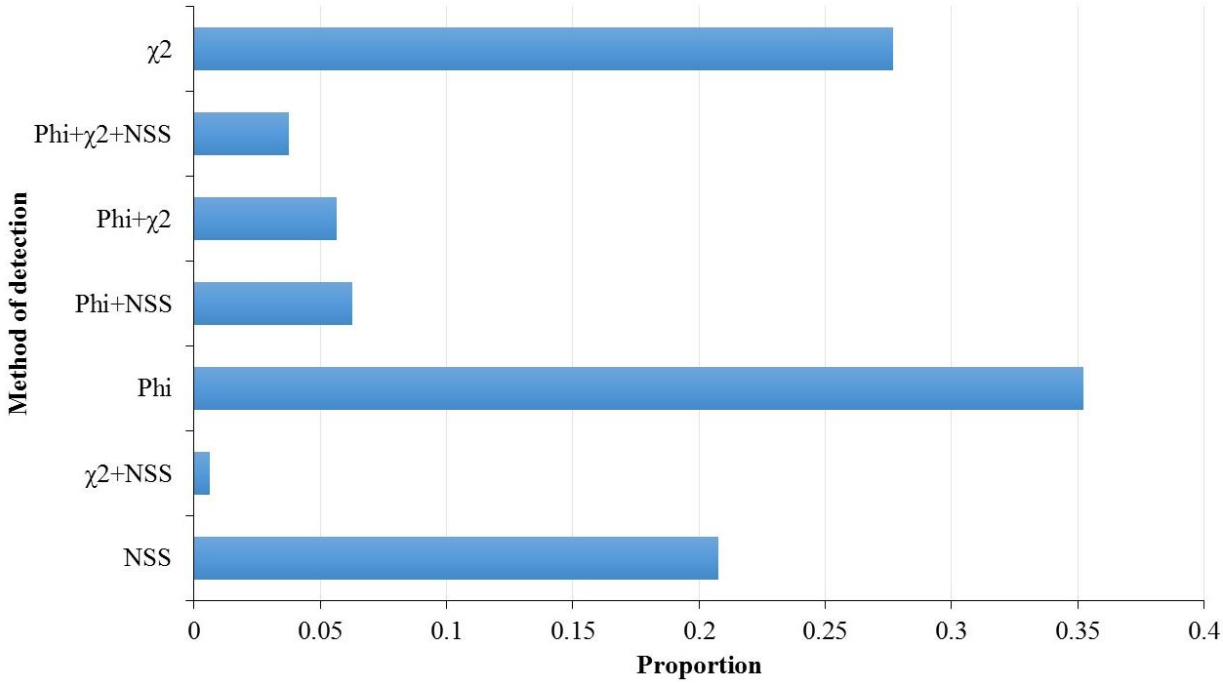


Figure 3-12: Proportions of recombination detection of 159 gene families using the three methods included in PhiPack program.

Despite the assumption of minimal impact of recombination in the core genes, the M7 (β) and M8 (β and ω) model, which has been described as a robust method of estimating positive selection in the presence of low recombination rates (Anisimova *et al.*, 2003), was applied in these studies. Recombination is not likely to bias the identification of positive selection where the strains under consideration are quite diverged (Castillo-Ramirez *et al.*, 2011). Thus, to identify gene families under positive selection in the four *Nesterenkonia* genomes, the Codeml M7 and M8 evolutionary models were implemented using 1,092 high quality alignments of the core genes. Overall, 391 (~ 36 %) of the core gene families showed significant (FDR adjusted p- value ≤ 0.05) signals for positive selection for the different lineages assigned as the foreground branch (Appendix T 3). Furthermore, 147 (~ 14 %) of the genes families in *Nesterenkonia* sp. AN1 branch showed evidence for positive selection (Appendix T 3, Figure 3-12). The branches of the mesophilic strains also showed similar proportions of positive selection signature. There was, however, no significant difference ($\chi 2$, $p \leq 0.05$) in the amount

of positive selection among branches. Of the 391 gene families that were subjected to positive selection, 150 (~ 38 %) were, however, selected across multiple branches and 241 (62 %) were uniquely selected in different branches (Figure 3-12). The identification of genes families that are positively selected in multiple lineages have been reported among members of the genera *Campylobacter* and *Streptococcus* (Lefébure and Stanhope, 2007, Lefébure and Stanhope, 2009).

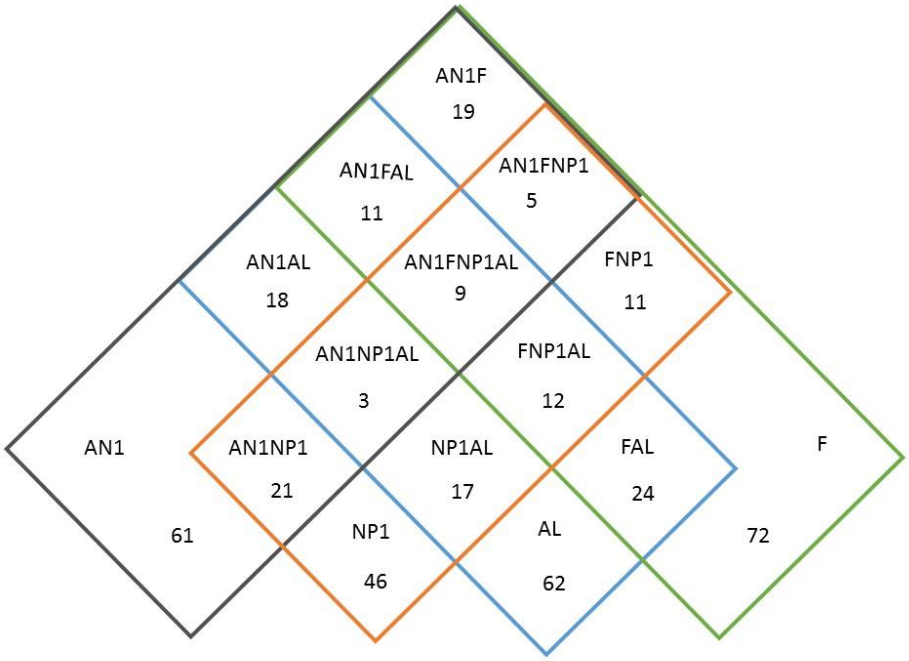


Figure 3-13: Distribution of gene families under positive selection based on branch-site test (codeml) in *Nesterenkonia* spp. AN1= *Nesterenkonia* sp. AN1; F= *Nesterenkonia* sp. F; NP1= *Nesterenkonia* sp. NP1 and AL= *Nesterenkonia alba* DSM 19423.

Functional annotation of the gene families based on COG (Tatusov *et al.*, 2000, Tatusov *et al.*, 2001) showed that there was higher proportion of significant positive selection with features involved in 'carbohydrate transport and metabolism' (G), 'coenzyme transport and metabolism' (H) and 'secondary metabolites biosynthesis, transport & catabolism' (Q) in *Nesterenkonia* sp. AN1. In *Nesterenkonia* sp. NP1, there is higher proportion of gene annotated in the categories 'energy production and conversion' (C) and 'inorganic ion transport and metabolism' (P),

whereas positive selection was higher in one category, 'amino acid transport and metabolism', in *Nesterenkonia* sp. F (Figure 3-13).

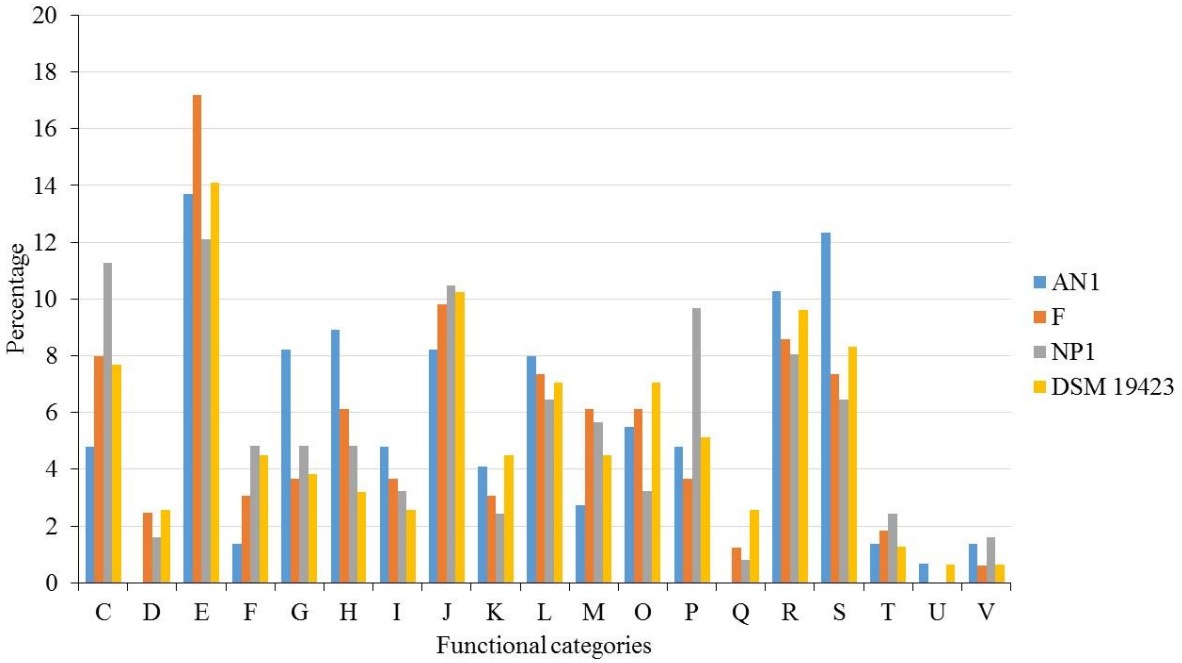


Figure 3-14: COG features of all gene families under positive selection based on branch-site test (Codeml) in *Nesterenkonia* spp. AN1= *Nesterenkonia* sp. AN1; F= *Nesterenkonia* sp. F; NP1= *Nesterenkonia* sp. NP1 and AL= *Nesterenkonia* alba DSM 19423. The COG categories have been defined in Figure 3-7.

To provide a better perspective of specific functions that are significantly associated with the positively selected genes, the corresponding proteins were annotated using GO and evaluated for function enrichment. Approximately 76 % of the 1,092 genes were mapped to specific GO terms by Blast2GO (Conesa *et al.*, 2005, Götz *et al.*, 2008). The GO term enrichment analysis for the gene families with positive selection signatures showed that sixteen functional categories were significantly (adjusted p-value ≤ 0.05) enriched for all the positively selected gene families (Table 3-2). These include two biological processes ("tricarboxylic acid cycle" and "DNA biosynthetic process") and five molecular functions ("acid-amino acid ligase

activity", "DNA-directed DNA polymerase activity", "DNA-dependent ATPase activity", "iron-sulphur cluster binding" and "3'-5' exonuclease activity") in *Nesterenkonia* sp. AN1.

Table 3-2: GO terms enriched for all positively selected genes identified using branch-site test in the four strains of *Nesterenkonia*.

Strain	GO ID	Ontology	Term	a. p-value
Nesterenkonia sp. AN1	GO:0071897	BP	DNA biosynthetic process	0.03
Nesterenkonia sp. AN1	GO:0006099	BP	tricarboxylic acid cycle	0.01
Nesterenkonia sp. AN1	GO:0003887	MF	DNA-directed DNA polymerase activity	0.03
Nesterenkonia sp. AN1	GO:0051536	MF	iron-sulfur cluster binding	0.00
Nesterenkonia sp. AN1	GO:0008408	MF	3'-5' exonuclease activity	0.03
Nesterenkonia sp. AN1	GO:0016881	MF	acid-amino acid ligase activity	0.04
Nesterenkonia sp. AN1	GO:0008094	MF	DNA-dependent ATPase activity	0.05
Nesterenkonia sp. F	GO:0009084	BP	glutamine family amino acid biosynthetic process	0.02
Nesterenkonia sp. F	GO:0000287	MF	magnesium ion binding	0.00
Nesterenkonia sp. F	GO:0005524	MF	ATP binding	0.03
Nesterenkonia sp. F	GO:0050661	MF	NADP binding	0.05
Nesterenkonia sp. F	GO:0004222	MF	metalloendopeptidase activity	0.02
Nesterenkonia sp. NP1	GO:0001510	BP	RNA methylation	0.00
Nesterenkonia sp. NP1	GO:0006400	BP	tRNA modification	0.05
Nesterenkonia sp. NP1	GO:0043190	CC	ATP-binding cassette (ABC) transporter complex	0.01
N. alba DSM 19423	GO:0006099	BP	tricarboxylic acid cycle	0.01
N. alba DSM 19423	GO:0016831	MF	carboxy-lyase activity	0.03

DNA biosynthetic processes were subject to positive selection in *Nesterenkonia* sp. AN1 only. These could play a role in protection against UV damage and desiccation in the Antarctic dry soils (Cowan *et al.*, 2014, De Maayer *et al.*, 2014a). 3'-5' exonuclease and DNA-directed DNA polymerase activities have been linked to DNA repair system of bacteria (Khare and Eckert, 2002, Reha-Krantz, 2010, Sghaier, 2011). Genes that code for DNA polymerase I (Nest_486), DNA polymerase III (alpha subunit) (Nest_542) and DNA polymerase III (gamma/tau subunit) (Nest_1057) (Table 3-2) were subject to positive selection. These were among the set of genes subject to positive selection in the ionizing-radiation-resistant bacteria (IRRB), which likely contribute to the resilience of the IRRB against ionising radiation (Sghaier *et al.*, 2008). Furthermore, DNA-dependent ATPase activity (GO:0008094) was enriched by positively selected genes coding for RecA/RadA recombinase (Nest_604), type IIA topoisomerase B subunit (Nest_638) and superfamily I DNA and RNA helicases (Nest_1185) (Table 3-2). These genes were among the DNA-associated genes in IRRB that were reported to be subject to positive selection (Sghaier *et al.*, 2008).

Five genes which encode enzymes associated with the tricarboxylic acid (TCA) cycle were subject to positive selection in *Nesterenkonia* sp. AN1 (Figure 3-14, Table 3-2).These include genes encoding isocitrate dehydrogenase (Nest_1064), succinyl-CoA synthetase subunit beta (Nest_27) succinate dehydrogenase/fumarate reductase, Fe-S protein subunit (Nest_1083) and flavoprotein subunit (Nest_1084) and malate:quinone oxidoreductase (Nest_1004), which catalyses the oxidation of malate to oxaloacetate (Appendix T 3). Nest_27, Nest_1004 and Nest_1064 were also subject to positive selection in *N. alba* DSM 19423.

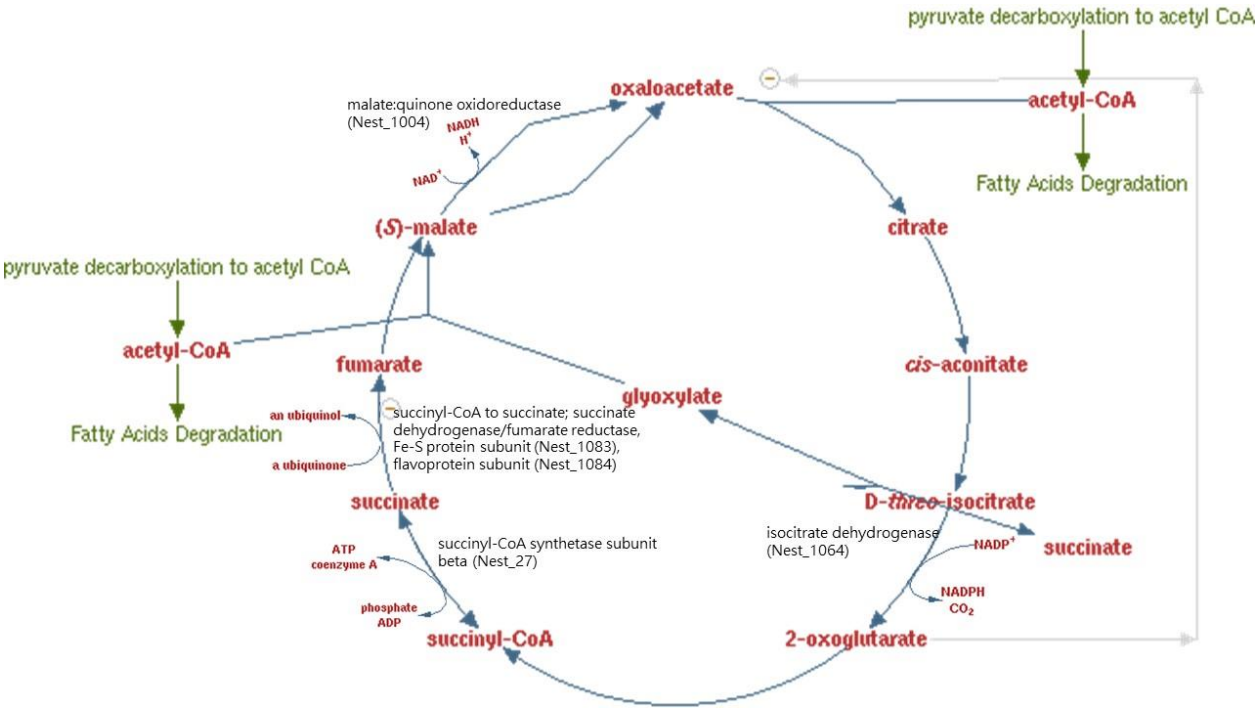


Figure 3-15: The TCA and glyoxylate pathways highlighting specific *Nesterenkonia sp.* AN1 enzymes subjected to positive selection. The pathway diagram was modified from the MetaCyc database (Caspi et al., 2014).

The NADP-dependent isocitrate dehydrogenase (Nest_1064) catalyses the conversion of isocitrate to 2-oxoglutarate (Mailloux *et al.*, 2007). Apart from its general role in TCA cycle, this enzyme has also been shown to play a role in protection against reactive oxygen species (ROS) via the production of NADPH and oxoglutarate. NADH is crucial for antioxidant regeneration while oxoglutarate act as an antioxidant (Mailloux *et al.*, 2007, Singh *et al.*, 2005). The enzyme is also known to be cold- and acetate-inducible in the psychrophiles *Colwellia maris* and *C. psychrerythraea* (Maki *et al.*, 2006). Succinyl-CoA synthetase is involved in the interconversion of succinyl-CoA to succinate, which is coupled with the production of ATP (Huynen *et al.*, 1999). Succinyl-CoA synthetase was observed to be upregulated under condition of aluminium toxicity in *Pseudomonas fluorescens* and this increase was linked to increased production of ATP required for survival under the stress condition (Singh *et al.*, 2009). Succinate dehydrogenase catalyses the reversibly conversion of succinate to fumarate

(Huynen *et al.*, 1999). The preferential expression of succinate dehydrogenase over fumarate reductase has been suggested to be an important adaptation to minimise the formation of ROS under aerobic conditions (Rustin and Rötig, 2002, Yankovskaya *et al.*, 2003, Imlay, 2003). The enzymes has also been shown to be involve in cold adaptation in *Propionibacterium freudenreichii* CIRM-BIA1^T (Dalmasso *et al.*, 2012). Malate:quinone oxidoreductase catalyses the irreversible oxidation of malate to oxaloacetate in citric acid cycle and glyoxylate cycle (Kretzschmar *et al.*, 2002, van der Rest *et al.*, 2000). It been demonstrated that the enzyme is essential for the utilization of C2-compounds through the glyoxylate cycle in *Pseudomonas citronellolis* and *Pseudomonas aeruginosa*

3.4 Conclusion

Nesterenkonia sp. AN1 is the first member of the genus *Nesterenkonia* to be isolated from a low temperature environment. Genome-wide comparisons of *Nesterenkonia* sp. AN1 and three of its mesophilic relatives were performed to provide insights into the evolutionary imprints of adaptive strategies that might determine the survival of *Nesterenkonia* sp. AN1 in the Antarctic soil habitat. Analysis of the *Nesterenkonia* pan-genome revealed pervasive adaptation strategies, consistent with the general characteristics of the genus as a taxon of highly resilient haloalkaliphilic bacteria. As seen with other ecologically diverse bacterial taxa; e.g., the pan-genome of the genus *Streptococcus* (Lefébure and Stanhope, 2007, Lefébure and Stanhope, 2009), *Nesterenkonia* strains exhibited an open pan-genome. In the course of their evolution, the strains of *Nesterenkonia* may have undergone intensive gene gain and loss which might have enabled them colonise different environments. Analysis of the adaptation proteins identified in *Nesterenkonia* sp. AN1 showed that the majority of the proteins were encoded on the shared (core) genome fraction, suggesting an inherent resilience to environmental stresses among *Nesterenkonia* species. Evaluation of the core genome, however, revealed the presence

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

of major adaptive differences which resulted from positive selection events in a large number of the core genes of *Nesterenkonia* sp. AN1. Furthermore, the strain-specific genome fraction of *Nesterenkonia* sp. AN1 encodes proteins that could be crucial for survival under multiple stress conditions. The evolutionary histories of all the strain-specific genes could not be ascertained because of limitation of genomes available for this study. However, evidence from genomic island analysis indicates that some of the genes, including those linked to responses to oxidative and salinity stresses, were acquired horizontally by *Nesterenkonia* sp. AN1.

Overall, comparison of the *Nesterenkonia* genomes revealed that *Nesterenkonia* sp. AN1 had retained a repertoire of ancestral resilience gene stock and, in the course of evolution, also gained a number of adaptation genes that probably facilitate the survival of the organism in the edaphic environment of Antarctica.

3.5 References

ABBY, S. & DAUBIN, V. 2007. Comparative genomics and the evolution of prokaryotes. *Trends Microbiol.,* 15**,** 135-141.

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 Biol. Biochem.,* 38**,** 3041-3056.

ALIYU, H., DE MAAYER, P., REES, J., TUFFIN, M. & COWAN, D. A. 2014. Draft Genome Sequence of the Antarctic Polyextremophile *Nesterenkonia* sp. Strain AN1. *Genome Announc,* 2.

AMBROGELLY, A., PALIOURA, S. & SÖLL, D. 2007. Natural expansion of the genetic code. *Nat. Chem. Biol.,* 3**,** 29-35.

ANISIMOVA, M., BIELAWSKI, J. P. & YANG, Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.,* 18**,** 1585-1592.

ANISIMOVA, M., NIELSEN, R. & YANG, Z. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics,* 164**,** 1229-1236.

AUCH, A. F., VON JAN, M., KLENK, H.-P. & GÖKER, M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci,* 2**,** 117.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 289-300.

BRUEN, T. & BRUEN, T. 2005. PhiPack: PHI test and other tests of recombination. *McGill University, Montréal, Québec, Canada*.

CASPI, R., ALTMAN, T., BILLINGTON, R., DREHER, K., FOERSTER, H., FULCHER, C. A., HOLLAND, T. A., KESELER, I. M., KOTHARI, A., KUBO, A., KRUMMENACKER, M., LATENDRESSE, M., MUELLER, L. A., ONG, Q., PALEY, S., SUBHRAVETI, P., WEAVER, D. S., WEERASINGHE, D., ZHANG, P. & KARP, P. D. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.,* 42**,** D459-D471.

CASTILLO-RAMIREZ, S., HARRIS, S. R., HOLDEN, M. T., HE, M., PARKHILL, J., BENTLEY, S. D. & FEIL, E. J. 2011. The impact of recombination on dN/dS within recently emerged bacterial clones. *PLoS Path.,* 7**,** e1002129.

CASTRESANA, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.,* 17**,** 540-552.

CHATTOPADHYAY, M. 2006. Mechanism of bacterial adaptation to low temperature. *J. Biosci. (Bangalore),* 31**,** 157-165.

CHINTALAPATI, S., KIRAN, M. & SHIVAJI, S. 2004. Role of membrane lipid fatty acids in cold adaptation. *Cell Mol Biol (Noisy-le-grand),* 50**,** 631-642.

COLLINS, M. D., LAWSON, P. A., LABRENZ, M., TINDALL, B. J., WEISS, N. & HIRSCH, P. 2002. *Nesterenkonia lacusekhoensis* sp. nov., isolated from hypersaline Ekho Lake, East Antarctica, and emended description of the genus *Nesterenkonia*. *Int. J. Syst. Evol. Microbiol.,* 52**,** 1145-1150.

CONESA, A., GÖTZ, S., GARCÍA-GÓMEZ, J. M., TEROL, J., TALÓN, M. & ROBLES, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics,* 21**,** 3674-3676.

COWAN, D. A., MAKHALANYANE, T. P., DENNIS, P. G. & HOPKINS, D. W. 2014. Microbial ecology and biogeochemistry of continental Antarctic soils. *Front Microbiol,* 5**,** 154.

D'AMICO, S., COLLINS, T., MARX, J. C., FELLER, G. & GERDAY, C. 2006. Psychrophilic microorganisms: challenges for life. *EMBO Rep,* 7**,** 385-9.

D'ARGENIO, D. A., VETTING, M. W., OHLENDORF, D. H. & ORNSTON, L. N. 1999. Substitution, insertion, deletion, suppression, and altered substrate specificity in functional protocatechuate 3, 4-dioxygenases. *J. Bacteriol.,* 181**,** 6478-6487.

DALMASSO, M., AUBERT, J., BRIARD-BION, V., CHUAT, V., DEUTSCH, S.-M., EVEN, S., FALENTIN, H., JAN, G., JARDIN, J., MAILLARD, M.-B., PARAYRE, S., PIOT, M., TANSKANEN, J. & THIERRY, A. 2012. A Temporal -omic Study of *Propionibacterium freudenreichii* CIRM-BIA1^{T} Adaptation Strategies in Conditions Mimicking Cheese Ripening in the Cold. *PLoS One,* 7**,** e29083.

DE MAAYER, P., ANDERSON, D., CARY, C. & COWAN, D. A. 2014a. Some like it cold: understanding the survival strategies of psychrophiles. *EMBO Rep,* 15**,** 508-517.

DE MAAYER, P., CHAN, W., RUBAGOTTI, E., VENTER, S., TOTH, I., BIRCH, P. R. & COUTINHO, T. 2014b. Analysis of the *Pantoea ananatis* pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts. *BMC Genomics,* 15**,** 404.

DELGADO, O., QUILLAGUAMÁN, J., BAKHTIAR, S., MATTIASSON, B., GESSESSE, A. & HATTI-KAUL, R. 2006. *Nesterenkonia aethiopica* sp. nov., an alkaliphilic, moderate halophile isolated from an Ethiopian soda lake. *Int. J. Syst. Evol. Microbiol.,* 56**,** 1229-1232.

DREESENS, L. L., LEE, C. K. & CARY, S. C. 2014. The Distribution and Identity of Edaphic Fungi in the McMurdo Dry Valleys. *Biology,* 3**,** 466-483.

DSOUZA, M., TAYLOR, M. W., TURNER, S. J. & AISLABIE, J. 2014. Genome-based comparative analyses of Antarctic and temperate species of *Paenibacillus*. *PLoS One,* 9**,** e108009.

DSOUZA, M., TAYLOR, M. W., TURNER, S. J. & AISLABIE, J. 2015. Genomic and phenotypic insights into the ecology of *Arthrobacter* from Antarctic soils. *BMC Genomics,* 16**,** 36.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.,* 32**,** 1792-1797.

EDOUARD, S., SANKAR, S., DANGUI, N. P., LAGIER, J. C., MICHELLE, C., RAOULT, D. & FOURNIER, P. E. 2014. Genome sequence and description of *Nesterenkonia massiliensis* sp. nov. strain NP1(T.). *Stand Genomic Sci,* 9**,** 866-82.

FELLER, G. 2003. Molecular adaptations to cold in psychrophilic enzymes. *Cell. Mol. Life Sci.,* 60**,** 648-62.

FELLER, G. & GERDAY, C. 2003. Psychrophilic enzymes: hot topics in cold adaptation. *Nat. Rev. Microbiol.,* 1**,** 200 - 208.

FELSENSTEIN, J. 1993. {PHYLIP}: phylogenetic inference package, version 3.5 c. *Cladistics*, 5, 163-166

FOSTER, P. L. 2005. Stress responses and genetic variation in bacteria. *Mutat. Res.,* 569**,** 3-11.

GARNIER, M., MATAMOROS, S., CHEVRET, D., PILET, M.-F., LEROI, F. & TRESSE, O. 2010. Adaptation to cold and proteomic responses of the psychrotrophic biopreservative *Lactococcus piscium* strain CNCM I-4031. *Appl. Environ. Microbiol.,* 76**,** 8011-8018.

GASTON, M. A., JIANG, R. & KRZYCKI, J. A. 2011a. Functional context, biosynthesis, and genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.,* 14**,** 342-349.

GASTON, M. A., ZHANG, L., GREEN-CHURCH, K. B. & KRZYCKI, J. A. 2011b. The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. *Nature,* 471**,** 647-650.

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. *Int. J. Syst. Evol. Microbiol*., 57, 81-91.

GÖTZ, S., GARCÍA-GÓMEZ, J. M., TEROL, J., WILLIAMS, T. D., NAGARAJ, S. H., NUEDA, M. J., ROBLES, M., TALÓN, M., DOPAZO, J. & CONESA, A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.,* 36**,** 3420-3435.

GOUVEIA-OLIVEIRA, R., SACKETT, P. W. & PEDERSEN, A. G. 2007. MaxAlign: maximizing usable data in an alignment. *BMC Bioinformatics,* 8**,** 312.

HALL, T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser,* 41**,** 95 - 98.

HALL, T. 2011. BioEdit: An important software for molecular biology. *GERF Bulletin of Bioscience,* 2**,** 60-61.

HAVE, C. T., ZAMBACH, S. & CHRISTIANSEN, H. 2013. Effects of using coding potential, sequence conservation and mRNA structure conservation for predicting pyrrolysine containing genes. *BMC Bioinformatics,* 14**,** 118.

HUGHES, A. L. & FRIEDMAN, R. 2008. Codon-based tests of positive selection, branch lengths, and the evolution of mammalian immune system genes. *Immunogenetics,* 60**,** 495-506.

HUYNEN, M. A., DANDEKAR, T. & BORK, P. 1999. Variation and evolution of the citric-acid cycle: a genomic perspective. *Trends Microbiol.,* 7**,** 281-291.

IMLAY, J. A. 2003. Pathways of Oxidative Damage. *Annu. Rev. Microbiol.,* 57**,** 395-418.

JUHAS, M., VAN DER MEER, J. R., GAILLARD, M., HARDING, R. M., HOOD, D. W. & CROOK, D. W. 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.,* 33**,** 376-393.

KAPPES, R. M., KEMPF, B., KNEIP, S., BOCH, J., GADE, J., MEIER-WAGNER, J. & BREMER, E. 1999. Two evolutionarily closely related ABC transporters mediate the uptake of choline for synthesis of the osmoprotectant glycine betaine in *Bacillus subtilis*. *Mol. Microbiol.,* 32**,** 203-216.

KATOH, K. & STANDLEY, D. M. 2014. MAFFT: iterative refinement and additional methods. *Multiple Sequence Alignment Methods.* Springer, pp. 131-146.

KHARE, V. & ECKERT, K. A. 2002. The proofreading 3′→5′ exonuclease activity of DNA polymerases: a kinetic barrier to translesion DNA synthesis. *Mutat Res,* 510**,** 45-54.

KONSTANTINIDIS, K. T., SERRES, M. H., ROMINE, M. F., RODRIGUES, J. L., AUCHTUNG, J., MCCUE, L.-A., LIPTON, M. S., OBRAZTSOVA, A., GIOMETTI, C. S. & NEALSON, K. H. 2009. Comparative systems biology across an evolutionary gradient within the Shewanella genus. *Proc Natl Acad Sci U S A,* 106**,** 15909-15914.

KONSTANTINIDIS, K. T. & TIEDJE, J. M. 2005a. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A,* 102**,** 2567-2572.

KONSTANTINIDIS, K. T. & TIEDJE, J. M. 2005b. Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.,* 187**,** 6258-6264.

KRETZSCHMAR, U., RÜCKERT, A., JEOUNG, J.-H. & GÖRISCH, H. 2002. Malate:quinone oxidoreductase is essential for growth on ethanol or acetate in *Pseudomonas aeruginosa*. *Microbiology,* 148**,** 3839-3847.

LASLETT, D. & CANBACK, B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.,* 32**,** 11-16.

LASSMANN, T. & SONNHAMMER, E. L. 2005. Kalign–an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics,* 6**,** 298.

LECHNER, M., FINDEIß, S., STEINER, L., MARZ, M., STADLER, P. F. & PROHASKA, S. J. 2011. Proteinortho: Detection of (Co-) orthologs in large-scale analysis. *BMC Bioinformatics,* 12**,** 124.

LEFÉBURE, T. & STANHOPE, M. J. 2007. Evolution of the core and pan-genome of Streptococcus: positive selection, recombination, and genome composition. *Genome Biol,* 8**,** R71.

LEFÉBURE, T. & STANHOPE, M. J. 2009. Pervasive, genome-wide positive selection leading to functional divergence in the bacterial genus Campylobacter. *Genome Res.,* 19**,** 1224-1232.

LI, W.-J., ZHANG, Y.-Q., SCHUMANN, P., LIU, H.-Y., YU, L.-Y., ZHANG, Y.-Q., STACKEBRANDT, E., XU, L.-H. & JIANG, C.-L. 2008. *Nesterenkonia halophila* sp. nov., a moderately halophilic, alkalitolerant actinobacterium isolated from a saline soil. *Int. J. Syst. Evol. Microbiol.,* 58**,** 1359-1363.

LI, W. J., CHEN, H. H., KIM, C. J., ZHANG, Y. Q., PARK, D. J., LEE, J. C., XU, L. H. & JIANG, C. L. 2005. *Nesterenkonia sandarakina* sp. nov. and *Nesterenkonia lutea* sp. nov., novel actinobacteria, and emended description of the genus *Nesterenkonia*. *Int. J. Syst. Evol. Microbiol.,* 55**,** 463-6.

LI, W. J., CHEN, H. H., ZHANG, Y. Q., SCHUMANN, P., STACKEBRANDT, E., XU, L. H. & JIANG, C. L. 2004. *Nesterenkonia halotolerans* sp. nov. and *Nesterenkonia xinjiangensis* sp. nov., actinobacteria from saline soils in the west of China. *Int. J. Syst. Evol. Microbiol.,* 54**,** 837-41.

LUO, H.-Y., WANG, Y.-R., MIAO, L.-H., YANG, P.-L., SHI, P.-J., FANG, C.-X., YAO, B. & FAN, Y.-L. 2009. *Nesterenkonia alba* sp. nov., an alkaliphilic actinobacterium isolated from the black liquor treatment system of a cotton pulp mill. *Int. J. Syst. Evol. Microbiol.,* 59**,** 863-868.

MACLEAN, D., JONES, J. D. & STUDHOLME, D. J. 2009. Application of'next-generation'sequencing technologies to microbial genetics. *Nat. Rev. Microbiol.,* 7**,** 287-296.

MAGIS, C., TALY, J.-F., BUSSOTTI, G., CHANG, J.-M., DI TOMMASO, P., ERB, I., ESPINOSA-CARRASCO, J. & NOTREDAME, C. 2014. T-coffee: tree-based consistency objective function for alignment evaluation. *Multiple Sequence Alignment Methods.* Springer, pp. 117-129.

MAILLOUX, R. J., BÉRIAULT, R., LEMIRE, J., SINGH, R., CHÉNIER, D. R., HAMEL, R. D. & APPANNA, V. D. 2007. The tricarboxylic acid cycle, an ancient metabolic network with a novel twist. *PLoS One,* 2**,** e690-e690.

MAKI, S., YONETA, M. & TAKADA, Y. 2006. Two isocitrate dehydrogenases from a psychrophilic bacterium, *Colwellia psychrerythraea*. *Extremophiles,* 10**,** 237-249.

MARLES-WRIGHT, J. & LEWIS, R. J. 2007. Stress responses of bacteria. *Curr. Opin. Struct. Biol.,* 17**,** 755-760.

MATH, R. K., JIN, H. M., KIM, J. M., HAHN, Y., PARK, W., MADSEN, E. L. & JEON, C. O. 2012. Comparative genomics reveals adaptation by *Alteromonas* sp. SN2 to marine tidal-flat conditions: cold tolerance and aromatic hydrocarbon metabolism. *PLoS One,* 7**,** e35784.

MEDINI, D., DONATI, C., TETTELIN, H., MASIGNANI, V. & RAPPUOLI, R. 2005. The microbial pan-genome. *Curr. Opin. Genet. Dev.,* 15**,** 589 - 594.

MEIER-KOLTHOFF, J. P., AUCH, A. F., KLENK, H.-P. & GÖKER, M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics,* 14**,** 60.

METPALLY, R. & REDDY, B. 2009. Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins. *BMC Genomics,* 10**,** 11.

MIRA, A., OCHMAN, H. & MORAN, N. A. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet.,* 17**,** 589-596.

MORENO-HAGELSIEB, G. & LATIMER, K. 2008. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics,* 24**,** 319-324.

MYKYTCZUK, N. C., FOOTE, S. J., OMELON, C. R., SOUTHAM, G., GREER, C. W. & WHYTE, L. G. 2013. Bacterial growth at -15 degrees C; molecular insights from the permafrost bacterium *Planococcus halocryophilus* Or1. *ISME J,* 7**,** 1211-26.

NEL, A., TUFFIN, I., SEWELL, B. & COWAN, D. 2011. Unique aliphatic amidase from a psychrotrophic and haloalkaliphilic *Nesterenkonia* isolate. *Appl. Environ. Microbiol.,* 77**,** 3696-3702.

NOTREDAME, C., HIGGINS, D. G. & HERINGA, J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.,* 302**,** 205-217.

OZER, E. A., ALLEN, J. P. & HAUSER, A. R. 2014. Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGEnt. *BMC Genomics,* 15**,** 737.

POLZ, M. F., ALM, E. J. & HANAGE, W. P. 2013. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet.,* 29**,** 170-175.

POWELL, S., SZKLARCZYK, D., TRACHANA, K., ROTH, A., KUHN, M., MULLER, J., ARNOLD, R., RATTEI, T., LETUNIC, I. & DOERKS, T. 2012. eggNOG v3. 0:

orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.,* 40**,** D284-D289.

QI, J., LUO, H. & HAO, B. 2004. CVTree: a phylogenetic tree reconstruction tool based on whole genomes. *Nucleic Acids Res.,* 32**,** W45-W47.

QIN, Q.-L., XIE, B.-B., ZHANG, X.-Y., CHEN, X.-L., ZHOU, B.-C., ZHOU, J., OREN, A. & ZHANG, Y.-Z. 2014a. A proposed genus boundary for the prokaryotes based on genomic insights. *J. Bacteriol*., 196, 2210-2215.

QIN, Q. L., XIE, B. B., YU, Y., SHU, Y. L., RONG, J. C., ZHANG, Y. J., ZHAO, D. L., CHEN, X. L., ZHANG, X. Y. & CHEN, B. 2014b. Comparative genomics of the marine bacterial genus *Glaciecola* reveals the high degree of genomic diversity and genomic characteristic for cold adaptation. *Environ. Microbiol.,* 16**,** 1642-1653.

RAFAEL, R., SÁNCHEZ-PORRO, C., MÁRQUEZ, M. & VENTOSA, A. 2011. Taxonomy of halophiles. *Extremophiles handbook.* Springer, pp. 255-308.

RASPAIL, C., GRAINDORGE, M., MOREAU, Y., CROUZY, S., LEFÈBVRE, B., ROBIN, A. Y., DUMAS, R. & MATRINGE, M. 2011. 4-Hydroxyphenylpyruvate dioxygenase catalysis identification of catalytic residues and production of a hydroxylated intermediate shared with a structurally unrelated enzyme. *J. Biol. Chem.,* 286**,** 26061-26070.

REHA-KRANTZ, L. J. 2010. DNA polymerase proofreading: Multiple roles maintain genome stability. *Biochim. Biophys. Acta,* 1804**,** 1049-1063.

RETIEF, J. D. 1999. Phylogenetic Analysis Using PHYLIP. In: MISENER, S. & KRAWETZ, S. (eds.) *Bioinformatics methods and rotocols*. Humana Press, pp. 243-258.

RICHTER, M., ROSSELLÓ-MÓRA, R., GLÖCKNER, F. O. & PEPLIES, J. 2015. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics,* 32**,** 929-931.

RUSTIN, P. & RÖTIG, A. 2002. Inborn errors of complex II – Unusual human mitochondrial diseases. *Biochim. Biophys. Acta,* 1553**,** 117-122.

SARIKHAN, S., AZARBAIJANI, R., YEGANEH, L. P., FAZELI, A. S., AMOOZEGAR, M. A. & SALEKDEH, G. H. 2011. Draft genome sequence of *Nesterenkonia* sp. strain F, isolated from Aran-Bidgol Salt Lake in Iran. *J. Bacteriol.,* 193**,** 5580.

SATAPATHY, S. S., DUTTA, M. & RAY, S. K. 2010. Higher tRNA diversity in thermophilic bacteria: A possible adaptation to growth at high temperature. *Microbiol. Res.,* 165**,** 609-616.

SGHAIER, H. 2011. DNA repair: Lessons from the evolution of ionizing-radiation-resistant prokaryotes-fact and theory. *In:* CHEN, C. (ed.) *Selected topics in DNA repair.* Rijeka, Croatia: INTECH, pp. 145-156.

SGHAIER, H., GHEDIRA, K., BENKAHLA, A. & BARKALLAH, I. 2008. Basal DNA repair machinery is subject to positive selection in ionizing-radiation-resistant bacteria. *BMC Genomics,* 9**,** 297.

SHARP, P. M., BAILES, E., GROCOCK, R. J., PEDEN, J. F. & SOCKETT, R. E. 2005. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res.,* 33**,** 1141-1153.

SIDDIQUI, K. S., WILLIAMS, T. J., WILKINS, D., YAU, S., ALLEN, M. A., BROWN, M. V., LAURO, F. M. & CAVICCHIOLI, R. 2013. Psychrophiles. *Annu Rev Earth Planet Sci,* 41**,** 87-115.

SINGH, R., BERIAULT, R., MIDDAUGH, J., HAMEL, R., CHENIER, D., APPANNA, V. D. & KALYUZHNYI, S. 2005. Aluminum-tolerant Pseudomonas fluorescens: ROS toxicity and enhanced NADPH production. *Extremophiles,* 9**,** 367-373.

SINGH, R., LEMIRE, J., MAILLOUX, R. J., CHÉNIER, D., HAMEL, R. & APPANNA, V. D. 2009. An ATP and oxalate generating variant tricarboxylic acid cycle counters aluminum toxicity in *Pseudomonas fluorescens*. *PLoS One,* 4**,** e7344.

SOLOVYEV, V. & SALAMOV, A. 2011. Automatic annotation of microbial genomes and metagenomic sequences. *In:* LI, R. W. (ed.) *Metagenomics and its applications in agriculture, biomedicine and environmental studies.* New York: Nova Science, pp. 61-78.

SRINIVASAN, G., JAMES, C. M. & KRZYCKI, J. A. 2002. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science,* 296**,** 1459-1462.

SU, F., OU, H.-Y., TAO, F., TANG, H. & XU, P. 2013. PSP: rapid identification of orthologous coding genes under positive selection across multiple closely related prokaryotic genomes. *BMC Genomics,* 14**,** 924.

TALAVERA, G. & CASTRESANA, J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.,* 56**,** 564-577.

TALY, J.-F., MAGIS, C., BUSSOTTI, G., CHANG, J.-M., DI TOMMASO, P., ERB, I., ESPINOSA-CARRASCO, J., KEMENA, C. & NOTREDAME, C. 2011. Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. *Nat Protoc,* 6**,** 1669-1682.

TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. & KUMAR, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.,* 30**,** 2725-9.

TATUSOV, R., GALPERIN, M., NATALE, D. & KOONIN, E. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucl Acid Res,* 28**,** 33 - 36.

TATUSOV, R., NATALE, D., GARKAVTSEV, I., TATUSOVA, T., SHANKAVARAM, U., RAO, B., KIRYUTIN, B., GALPERIN, M., FEDOROVA, N. & KOONIN, E. 2001. The COG

database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.,* 29**,** 22 - 28.

TETTELIN, H., MASIGNANI, V., CIESLEWICZ, M. J., DONATI, C., MEDINI, D., WARD, N. L., ANGIUOLI, S. V., CRABTREE, J., JONES, A. L., DURKIN, A. S., DEBOY, R. T., DAVIDSEN, T. M., MORA, M., SCARSELLI, M., MARGARIT Y ROS, I., PETERSON, J. D., HAUSER, C. R., SUNDARAM, J. P., NELSON, W. C., MADUPU, R., BRINKAC, L. M., DODSON, R. J., ROSOVITZ, M. J., SULLIVAN, S. A., DAUGHERTY, S. C., HAFT, D. H., SELENGUT, J., GWINN, M. L., ZHOU, L., ZAFAR, N., KHOURI, H., RADUNE, D., DIMITROV, G., WATKINS, K., O'CONNOR, K. J., SMITH, S., UTTERBACK, T. R., WHITE, O., RUBENS, C. E., GRANDI, G., MADOFF, L. C., KASPER, D. L., TELFORD, J. L., WESSELS, M. R., RAPPUOLI, R. & FRASER, C. M. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A,* 102**,** 13950-5.

TETTELIN, H., RILEY, D., CATTUTO, C. & MEDINI, D. 2008. Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.,* 11**,** 472-477.

UTRATNA, M., COSGRAVE, E., BAUSTIAN, C., CEREDIG, R. H. & O'BYRNE, C. P. 2014. Effects of Growth Phase and Temperature on Activity within a *Listeria monocytogenes* Population: Evidence for RsbV-Independent Activation of at Refrigeration Temperatures. *Biomed Res Int,* 2014.

VAN DER REST, M. E., FRANK, C. & MOLENAAR, D. 2000. Functions of the Membrane-Associated and Cytoplasmic Malate Dehydrogenases in the Citric Acid Cycle of *Escherichia coli*. *J. Bacteriol.,* 182**,** 6892-6899.

WU, S., ZHU, Z., FU, L., NIU, B. & LI, W. 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics,* 12**,** 444.

XU, Z. & HAO, B. 2009. CVTree update: a newly designed phylogenetic study platform using composition vectors and whole genomes. *Nucleic Acids Res.,* 37**,** W174-W178.

YANG, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.,* 24**,** 1586-1591.

YANG, Z. & NIELSEN, R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.,* 19**,** 908-917.

YANKOVSKAYA, V., HORSEFIELD, R., TÖRNROTH, S., LUNA-CHAVEZ, C., MIYOSHI, H., LÉGER, C., BYRNE, B., CECCHINI, G. & IWATA, S. 2003. Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science,* 299**,** 700-704.

ZHANG, Y. & SIEVERT, S. M. 2014. Pan-genome analyses identify lineage- and niche-specific markers of evolution and adaptation in Epsilonproteobacteria. *Front Microbio,* 5**,** 110.

ZHENG, Q. & WANG, X.-J. 2008. GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic Acids Res.,* 36**,** W358-W363.

Chapter 4

4 Transcriptome analysis reveals several key adaptive features for the survival of *Nesterenkonia* sp. AN1 under cold conditions

4.1 Introduction

Although temperatures in the Dry Valleys of Antarctica are generally in the sub-zero range (mean annual range ~ -20 to -25°C) for the greater part of the year (Cowan and Ah Tow, 2004), bacteria isolated from the Antarctica dry soils are generally psychrotrophic (psychrotolerant), and have the requisite genetic makeup to survive the frequent daily and seasonal temperature fluctuations. For example, soil surface temperature during the austral summer usually rise to +15°C when the ground is exposed to direct solar irradiation but drop to zero when the ground is shadowed (Cowan and Ah Tow, 2004, Kirby *et al.*, 2011). Life processes in the Dry Valleys are further constrained by drought, high alkalinity and osmotic stress, fluxes of UV irradiation and scarcity of nutrients (Aislabie *et al.*, 2006, De Maayer *et al.*, 2014).

Genomic analyses of bacterial strains obtained from extreme environments generally reveal the presence of genetic features which are linked to tolerance to the environmental stresses to which they are exposed. For instance, recent assessments of the genome sequences of *Arthrobacter* spp. (Dsouza *et al.*, 2015) and *Paenibacillus* spp. (Dsouza *et al.*, 2014) isolated from Antarctic soils revealed a range of adaptation mechanisms which may potentially underlie their survival in the cold, arid soils. However, the majority of these features could also be observed in the genomes of their temperate relatives (Dsouza *et al.*, 2014, Dsouza *et al.*, 2015). Furthermore, our comparative analysis of the genome of the Antarctic strain and three mesophilic *Nesterenkonia* strains (Chapter 3) concurs with these findings. It should, however, be stressed that the presence of these genetics features does not necessarily correlate to the phenotype to which they are linked (Lockhart and Winzeler, 2000), especially as it relates to gene function under extreme conditions. While the gene may be present, it may not necessarily be expressed under cold conditions. For example, although, orthologs of genes encoding the cold shock protein CspC, which has been shown to function in cold acclimation in psychrophiles

(De Maayer *et al.*, 2014, Siddiqui *et al.*, 2013), are present in both *Bacillus subtilis* and *Escherichia coli*, the gene was only demonstrated to be cold stress inducible in *B. subtilis* and not expressed under cold conditions in *E. coli* (Graumann and Marahiel, 1999). In order to study the role of a genetic element in an adaptive strategy, such as cold tolerance, it is therefore imperative to determine its expression under the conditions at which it should be active.

The psychrotolerant Gram-positive bacterium *Nesterenkonia* sp. AN1 was isolated from soil samples collected from Miers Valley, Antarctica (Nel *et al.*, 2011). Analysis of the growth characteristics of this strain revealed that it is a polyextremophile, being tolerant to low temperatures, high pH and high salt concentrations (Nel *et al.*, 2011). Analyses of the genome sequence (Aliyu *et al.*, 2014) of the Antarctic *Nesterenkonia* strain revealed a plethora of adaptive features potentially crucial to the responses to cold and other stresses which likely define the survival of the strain in the Dry Valleys (Chapter 2). Here, RNA-seq based transcriptome analysis was employed, in order to provide additional insights into the roles played by the putative stress response elements identified from the genome, with particular emphasis on how this strain can cope with the cold conditions and temperature fluctuations it may be exposed to in Antarctic soils.

4.2 Material and Methods

4.2.1 Growth conditions

Nesterenkonia sp. AN1 was isolated from soil samples obtained from Antarctica as described (Nel *et al.*, 2011). The isolate was grown in triplicate in 20 ml modified Luria-Bertani (mLB) medium containig of 10 mg/mL tryptone, 5 mg/mL yeast extract and 50 mg/mL NaCl adjusted to pH 8.5. The culture was incubated in a rotary shaker (180 rpm) at 21 ºC until an optical density (OD_{600}) of 0.5 was obtained. Cells were pelleted by centrifugation (8000 rpm, 5 minutes) at 21ºC. The pellets were re-suspended in 1.5 ml of mLB broth. 0.25 µl of the cells was transferred in triplicate to freshly prepared 50 ml of mLB broth. These were incubated in a rotary shaker (180 rpm) at the optimum growth

temperature of 21 ºC and the test temperature of 5 ºC to an exponential growth phase; absorbance value (OD_{600}) of 0.5 (approximately similar cell numbers for the two temperature regimes). The cultures attained the OD600 value of 0.5 in ~ 24 hours and ~ 106 hours for the 21 ºC and 5 ºC, respectively (Figure 4-1).

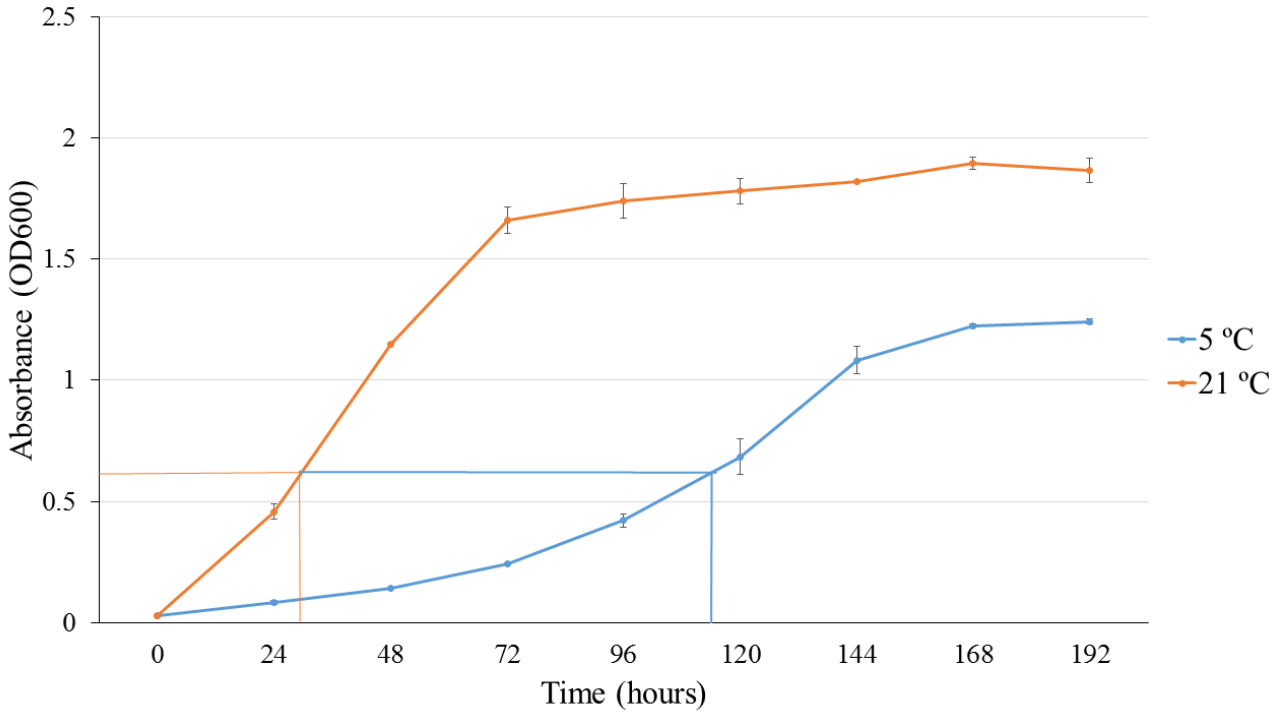


Figure 4-1: Growth profiles of cultures of *Nesterenkonia* sp. AN1 grown at 5 ºC and 21 ºC. Absorbance value (OD_{600}) of 0.5 and the corresponding time at which total RNA was extracted for the two treatment conditions are indicated.

4.2.2 RNA preparation

Total RNA was extracted from two replicate samples from each treatment condition. The bacterial cultures (20 ml from each replicate) were pelleted at 8,000 rpm for 5 minutes at 5 ºC or 21 ºC based on the treatment conditions. The cell pellets were re-suspended in 1 ml TRI Reagent® (Zymo Research Corp., USA). RNA extraction was performed using the Direct-zol™ RNA Mini Prep kits (Zymo Research Corp., USA) according to the manufacturer's instructions with some modifications. Briefly, the cells were re-suspended in 1ml TRI Reagent® and then transferred to ZR Bashing Bead™Lysis

Tubes. The cells were lysed using a PowerLyser™ (MO Bio Lab. Inc., USA) at 2,000 rpm for one minute. The homogenate was centrifuged at 8,000 rpm for 1 minute and the supernatant transferred into RNAse-free 1.5ml tubes. DNase I treatment and subsequent RNA purification were performed as per the protocol provided by the manufacturer. Preliminary quality assessment of the extracted RNA was performed using a NanoDrop spectrophotometer. The RNA was preserved at - 80 ºC.

4.2.3 RNA sequencing

RNA sequencing was performed using the Illumina MiSeq platform at Inqaba Biotech, South Africa. The basic workflow involved selective enrichment of mRNA by depleting the ribosomal RNA using the Ribo-Zero™ rRNA removal kit (Epicentre, USA) and RNA quality assessment using the Agilent Bioanalyzer™. An Illumina paired-end library was prepared and the samples were multiplexed and sequenced in a single lane of the Illumina Miseq (600 cycles with an expected data yield of 1.2GB).

4.2.4 RNA-Seq analysis

The paired end sequencing read datasets were analysed using the CLC Genomics Workbench 7.5, using the draft genome of *Nesterenkonia* sp. AN1 as a reference. The reads were mapped against the reference genome using the default mapping options. Differentially expressed genes were assessed using the empirical analysis of differentially expressed genes (EDGE) tool with a stringent filter (absolute fold change of > 1.5, false discovery rate (FDR) < 0.05 and Bonferreni correction for multiple testing $p < 0.05$). The RNA-Seq data was deposited in the NCBI GEO repository (Barrett *et al.*, 2013) under the accession number GSE71132. Functional annotation of the overexpressed genes was done using clusters of orthologous genes (COG) (Tatusov *et al.*, 2003), accessed via WebMGA (Wu *et al.*, 2011), in combination with an in-house adapted multiple alignment of actNOG obtained from the Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups (eggNOG v3.0) database (Powell *et al.*, 2012). The differentially expressed genes were assessed for gene ontology (GO) term

enrichment using the web-based GO enrichment toolkit GOEAST (Zheng and Wang, 2008). The up-regulated and down-regulated transcripts associated with GO terms were tested separately against the *Nesterenkonia* sp. AN1 GO annotated gene set (background) using the default setting in GOEAST (Zheng and Wang, 2008).

4.3 Results and Discussion

4.3.1 RNA-Seq read mappings

RNA sequencing yielded between 1,089,506 and 1,321,644 paired-end reads per sample, comprising a total of 651 Mb of sequence data (Table 4-1). Approximately 77.36 % of the paired-end reads could be mapped to the *Nesterenkonia* sp. AN1 draft genome sequence (Aliyu *et al.*, 2014), out of which 63.34 % and 14.01 % were mapped to the genic and intergenic regions, respectively. The read mapping of the transcripts against the *Nesterenkonia* sp. AN1 draft genome was visualized using the track list function in CLC Genomics Workbench 7.5. Read coverage for a section of contig1 of the draft genome is presented in Figure 4-2.

Table 4-1: Mapping statistics for *Nesterenkonia* sp. AN1 RNA-Seq paired-end (PE) reads. The control treatment represents *Nesterenkonia* sp. AN1 grown at the optimal growth temperature (21°C), while the cold treatment involved growth at 5°C.

Sample	Number of reads (PE)	% reads mapped (genic)	% reads mapped (intergenic)	% reads not mapped
Control 1	1089506	63.84	15.71	20.45
Control 2	1321644	60.51	12.30	27.19
Cold 1	1287334	63.47	14.20	22.33
Cold 2	1122330	66.05	14.18	19.76
Total	4820814	63.34	14.01	22.64

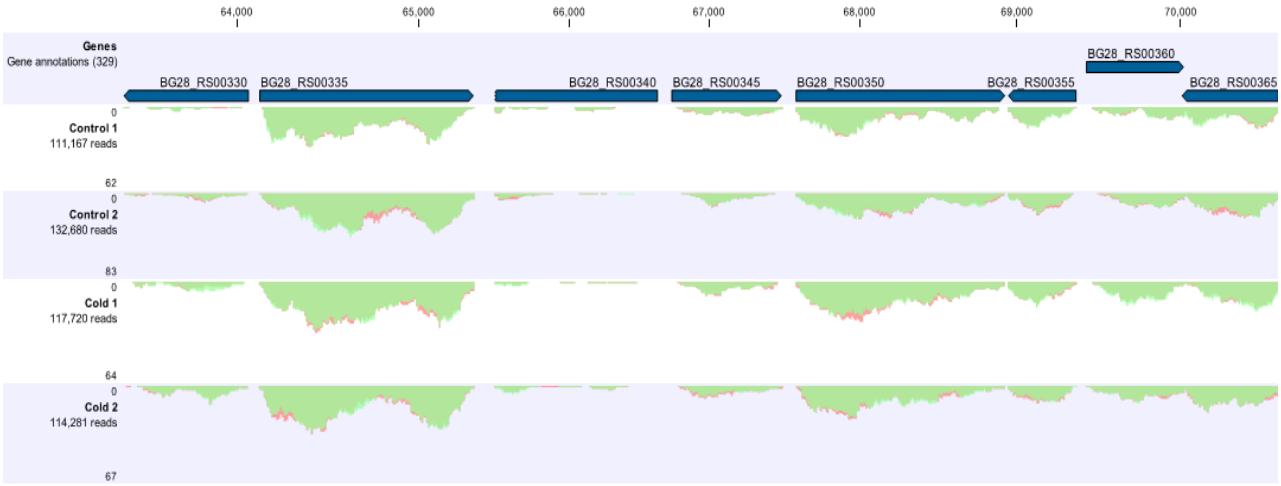


Figure 4-2: Track lists showing RNA-Seq read coverage for a section of contig 1 of the draft genome of *Nesterenkonia* sp. AN1 visualised using the track list function of the CLC Genomics Workbench 7.5.

4.3.2 Analysis of gene expression

Analyses of the transcriptome profile in *Nesterenkonia* sp. AN1 grown at 5°C (cold) and 21°C (control) revealed that a total of 2,796 (~ 97 %) of the 2,852 predicted genes were expressed, signifying high sequencing depth (Kogenaru *et al.*, 2012). The KEGG pathway mapping of the entire transcriptome profile computed using iPATH2.0 (Letunic *et al.*, 2008, Yamada *et al.*, 2011) is presented in Appendix F 2.

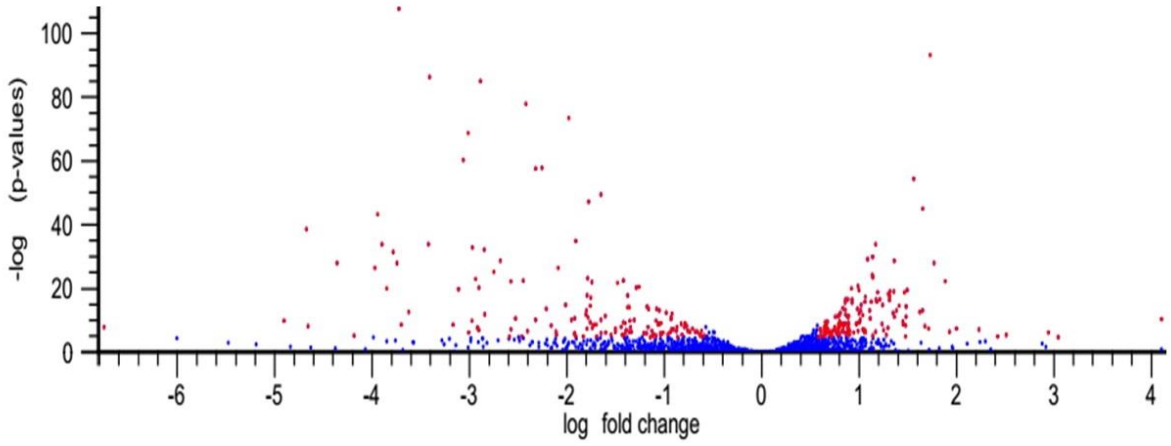


Figure 4-3: Volcano plot showing transcripts expression profile in *Nesterenkonia* sp. AN1 grown at 5 ºC and 21 ºC. Differentially expressed genes at 5 ºC are represented by the red dots, while blue dots are those genes which are constitutively expressed under both conditions. The negative and positive fold change values indicate the downregulated and upregulated transcripts, respectively.

A total of 316 transcripts (~ 11 %) showed significant (FDR < 0.05, Bonferreni adjusted P-value < 0.05, absolute fold change > 1.5) differential expression when *Nesterenkonia* sp. AN1 was grown at 5 ºC relative to 21 ºC (Figure 4-3), with 164 and 152 genes upregulated and downregulated, respectively at 5°C (Appendix T 4). Hierarchical clustering of the overexpressed genes using relative proportion of the normalised expression values showed distinct separation of the differential expressed genes at 5°C. The downregulated transcript showed an overall higher transcript abundance compared to the upregulated transcripts (Figure 4-4).

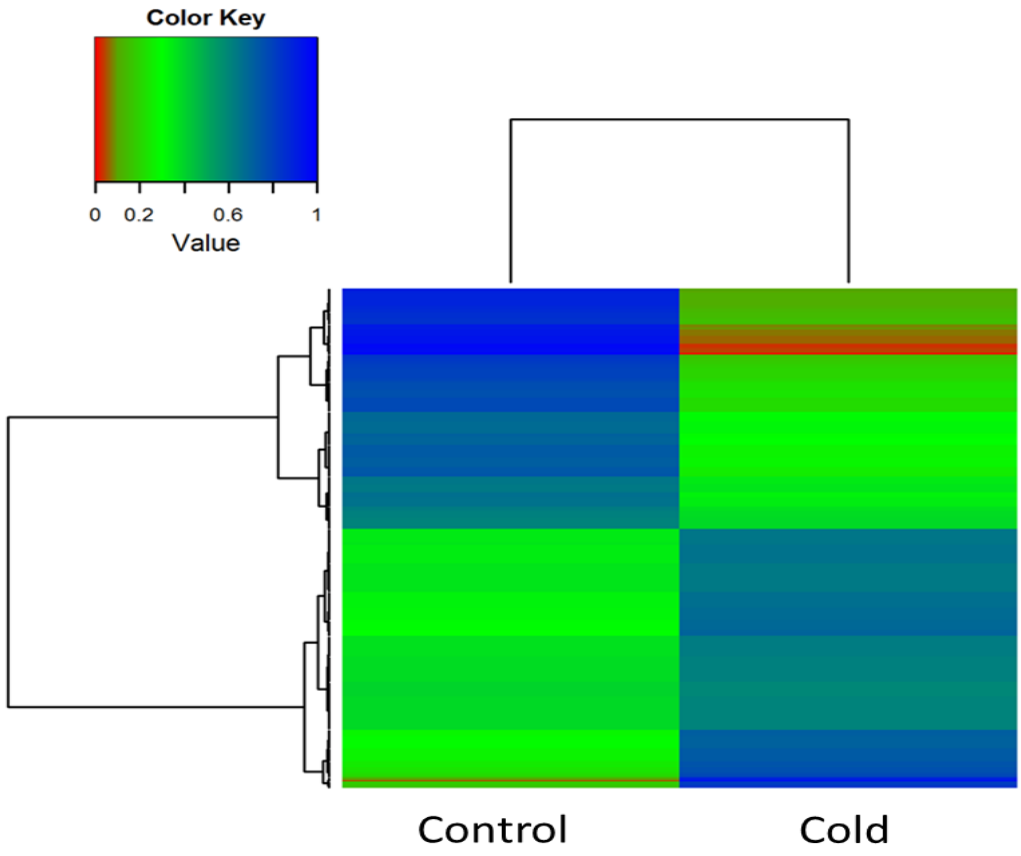


Figure 4-4: Hierarchical clustering of the differentially expressed transcripts in *Nesterenkonia* sp. AN1 at at 5°C (cold) relative to 21°C.(control). The colour key shows the relative abundance of the differentially expressed genes.

Functional annotation of the differentially expressed transcripts, using a combination of comparison against the COG and eggNOG databases (Powell *et al.*, 2012, Tatusov *et al.*, 2003), showed an overall greater number of upregulated transcripts coding for proteins categorised in the super-functional classes "cellular processes and signalling" and the "poorly characterised proteins" (Figure 4-5) at 5°C. The majority of the downregulated transcripts at 5°C belonged to the super-functional category "metabolism". Further evaluation of the functional classes revealed that among the upregulated genes, there was an over-representation of genes involved in "cell cycle control", "cell division", "chromosome partitioning", "cell membrane biogenesis", "posttranslational modification", "protein turnover", "chaperones and signal transduction" (Figure 4-6). In contrast, the downregulated genes were dominated by those involved in "information storage and processing", particularly those of the functional class "DNA replication, recombination and repair" (Figure 4-6). Similarly, genes involved in amino acid, lipid and inorganic ion transport and metabolism were down-regulated at 5 ºC. Fifty-five genes, representing ~ 31 % of the genes upregulated at 5 ºC, were classified in the 'poorly characterized' super-functional category (Figure 4-6).

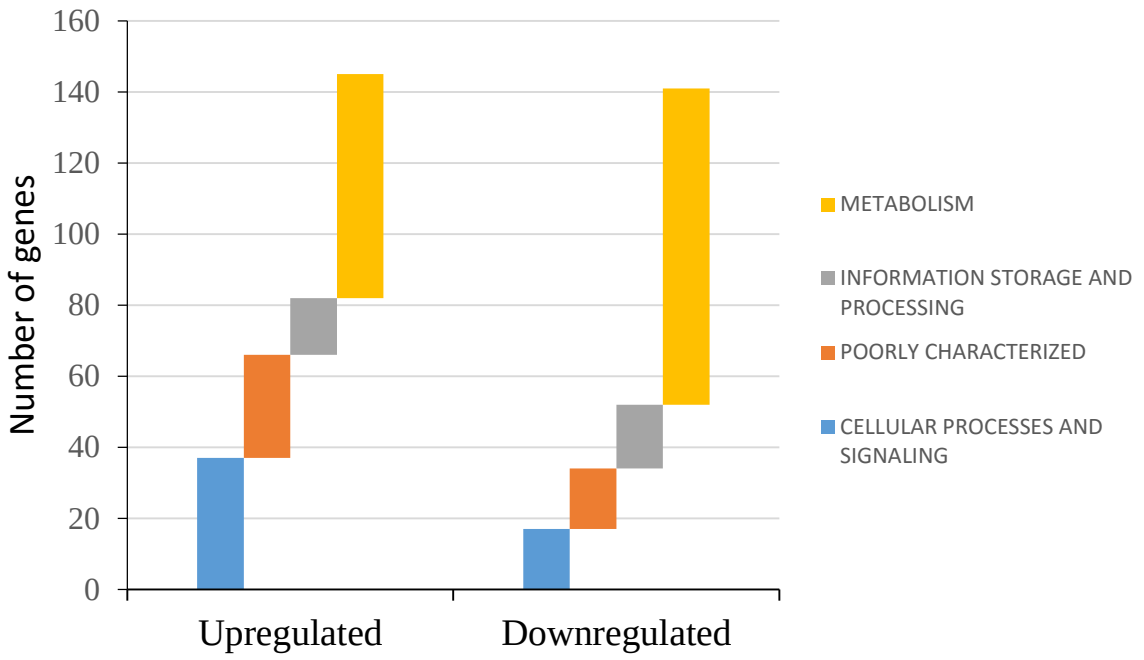


Figure 4-5*:* EggNOG functional categorization of the differentially expressed genes of *Nesterenkonia* sp. AN1 at 5°C relative to 21°C based on EGGNOG.

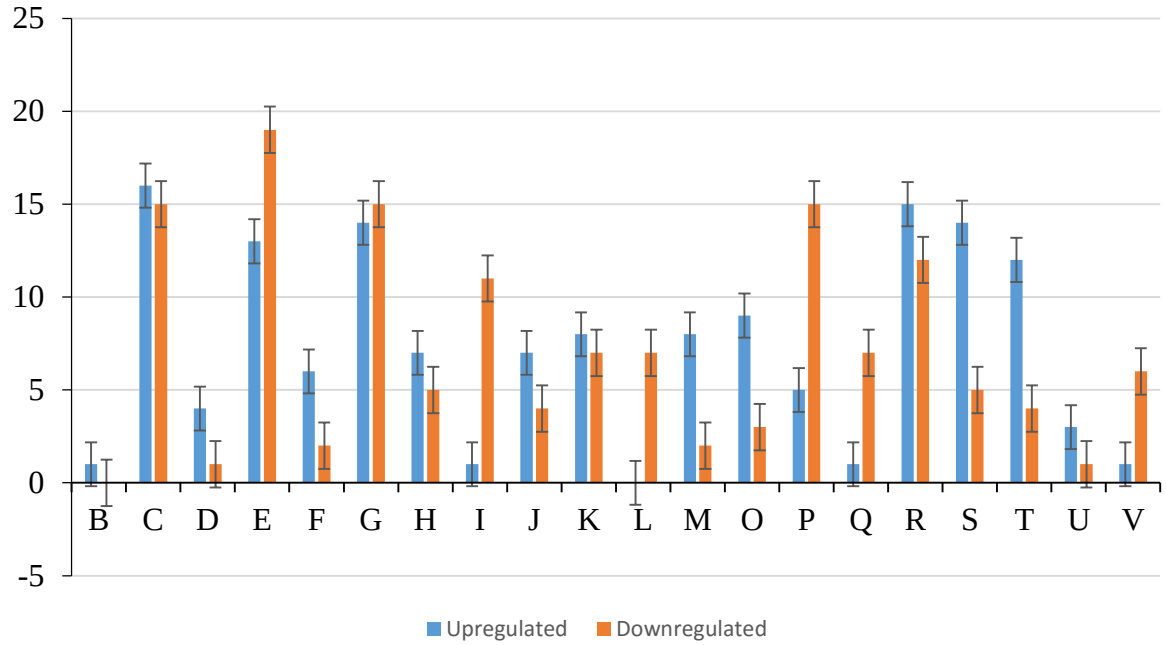


Figure 4-6: EGGNOG classification of differentially expressed genes in *Nesterenkonia* sp. AN1 grown at 5°C. The bars represent number of overexpressed transcripts. The COG categories in EGGNOG are defined as follows: "INFORMATION STORAGE AND PROCESSING; translation, ribosomal structure and biogenesis (J), transcription (K), replication, recombination and repair (L) and chromatin structure and dynamics (B). CELLULAR PROCESSES AND SIGNALING; cell cycle control, cell division, chromosome partitioning (D), defence mechanisms (V), signal transduction mechanisms (T), cell wall/membrane/envelope biogenesis (M), intracellular trafficking, secretion, and vesicular transport (U) and posttranslational modification, protein turnover, chaperones (O). METABOLISM; energy production and conversion (C), carbohydrate (G), amino acid (E), nucleotide (F), coenzyme (H), lipid (I), inorganic ion (P) transport and metabolism and secondary metabolites biosynthesis, transport and catabolism (Q). POORLY CHARACTERIZED; general function prediction only (R) and function unknown (S)". The error bars indicate the standard error.

4.3.2.1 Most stress response genes were expressed under both cold and optimal growth conditions

The transcriptome profile was also assessed to determine the expression of the putative stress response genes identified from the genomic data in Chapter 2. The results revealed that most of the genes coding for stress adaptation features (183 genes; 97 % of the 189 stress features) were expressed under both experimental conditions (5 ºC and 21 ºC). Of these, 155 were constitutively expressed (i.e., expressed at the same level under both conditions). Twenty-eight of the predicted adaptive genes were differentially expressed, with twelve down-regulated and sixteen up-regulated at 5 ºC. Surprisingly, the majority of the cold shock and acclimation features, whose function has been experimentally

elucidated in other bacteria exposed to cold temperatures, were constitutively expressed in *Nesterenkonia* sp. AN1, with the *cspA* gene being downregulated at 5 ºC, suggesting that *cspA* is not cold-induced in *Nesterenkonia* sp. AN1. By contrast, *cspA* has been shown to be upregulated transiently upon temperature downshift in *E. coli* (Polissi *et al.*, 2003). Similarly, the antioxidant thioredoxin gene *trxA* was downregulated at 5 ºC. This may, however, be compensated for by the significant upregulation of *trxC*, which produces thioredoxin Trx2, with an equivalent antioxidant function (Zeller and Klug, 2006). Additionally, there was significant upregulation at 5 ºC of *sodA* and *bcp* which encode the reactive oxygen species scavengers superoxide dismutase SodA and thiol peroxidase Bcp, respectively (Moustafa *et al.*, 2010). The induction of these antioxidants could be a strategy to deal with a possible spike in endogenous reactive oxygen species, resulting from the induction of molybdenum cofactor biosynthesis genes *moaA* and *moaB* (Mehta *et al.*, 2014, Regulski *et al.*, 2008), both of which were significantly induced at 5 ºC. Molybdo-enzymes are commonly associated with catalytic activities that lead to the production of reactive oxygen species (Mendel and Bittner, 2006, Neumann and Leimkühler, 2010).

Four of the seven copies of *uspA*, three of which are unique to *Nesterenkonia* sp. AN1 among the four *Nesterenkonia* strains compared in Chapter 3, were upregulated at 5 ºC. The *uspA* genes of the psychrotrophic bacterium *Lactococcus piscium* CNCM I-4031 have also been reported to be cold-inducible (Garnier *et al.*, 2010). UspA and UspA-like proteins have also been reported to be associated with responses to several other stresses including carbon starvation, and osmotic and oxidative stresses (Phadtare and Inouye, 2001, Seifart Gomes *et al.*, 2011). Although these genes were induced in response to cold, other stress-related roles for the *Nesterenkonia* sp. AN1 *uspA* genes cannot be excluded.

4.3.2.2 Induction of genes encoded on the core genome

Further analyses of the differentially expressed genes, with respect to the comparative genomics data for the *Nesterenkonia* spp. (Chapter 3), showed that 147, (~ 49 %) of the differentially expressed genes were encoded on the core genome and 61 genes, which accounted for ~ 20 % of the overexpressed genes at 5°C, were unique to *Nesterenkonia* sp. AN1 (Figure 4-7). It was observed that thirteen genes among the core genome repertoire subjected to positive selection, were upregulated at 5°C. These include transcripts that code for 2-heptaprenyl-1, 4-naphthoquinone methyltransferase (UbiE), 6-phosphofructokinase (Pfk), fructose-1, 6-bisphosphatase (GlpX), iron-sulfur cluster assembly protein (SufB), Lipoate synthase (LipA), multi-modular transpeptidase-transglycosylase (PbpG) and succinate dehydrogenase; flavoprotein and iron-sulfur protein subunits (SdhA and SdhB) (Chapter 3). UbiE is associated with coenzyme transport and metabolism and mutation of the gene encoding UbiE resulted in the impairment of ubiquinone and menaquinone production and disruption of growth in *E. coli* (Lee *et al.*, 1997). PbpG is one of the enzymes that catalyses the transglycosylation and transpeptidation steps during peptidoglycan biosynthesis (Haenni *et al.*, 2006). Mutational studies have shown that at least one copy of the enzyme PbpG is essential for the completion of the process (Denome *et al.*, 1999). Induction of PbpG may indicate active cell wall synthesis at low temperature since the enzyme catalyses the final stages of peptidoglycan synthesis.

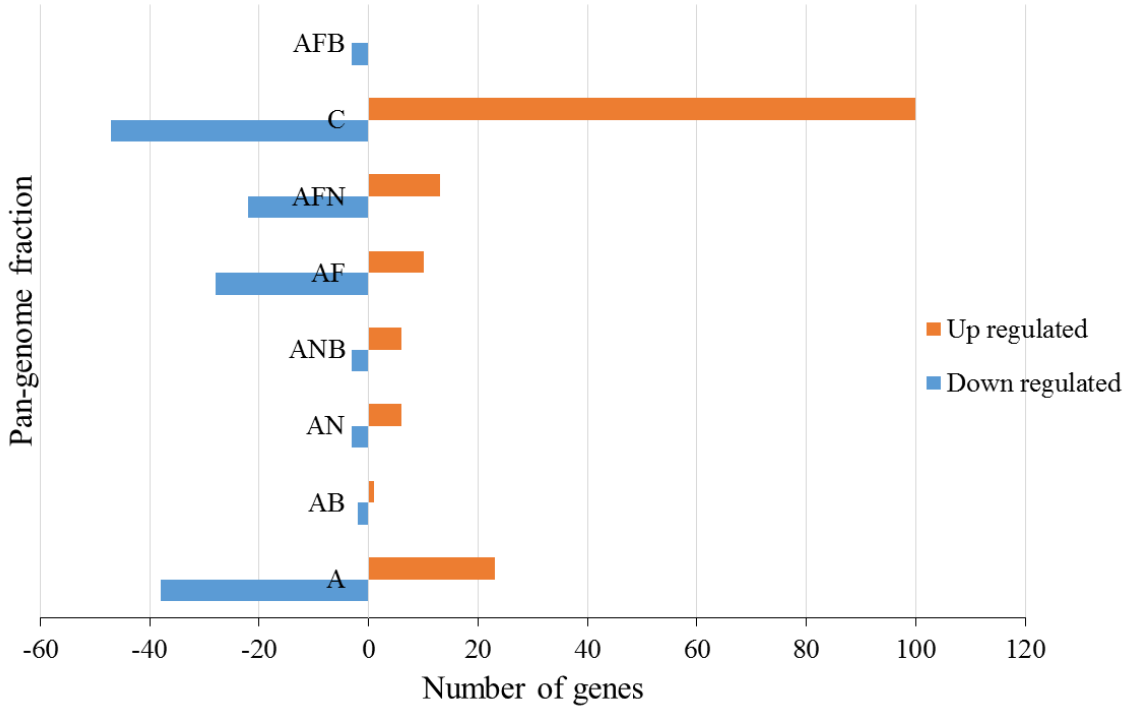


Figure 4-7: Distribution of differentially expressed genes in the pan genome of *Nesterenkonia*. Number of environmental adaptation genes associated with different pan-genome fractions (C: core, A=*Nesterenkonia* sp. AN1, B: *N. alba* DSM 19423, N: *Nesterenkonia* sp. NP1 and F: *Nesterenkonia* sp. F). Negative gene numbers represent the down-regulated transcripts at 5°C and the positive values represent the up-regulated transcripts at 5°C.

The protein was, however, not detected when *Planococcus halocryophilus* was grown under cold conditions or at high salt concentrations (Ronholm *et al.*, 2015). GplX and Pfk catalyse different stages in carbohydrate metabolism (Enstrom *et al.*, 2012). GplX has also been reported to play a role in metabolism of glycerol under cold condition in *Arthrobacter chlorophenolicus* (Unell *et al.*, 2009). SdhA and SdhB are two subunits of succinate dehydrogenase involved in energy metabolism. Succinate dehydrogenase catalyses the conversion of succinate to fumarate in the TCA cycle (Huynen *et al.*, 1999) and its potential role in resistance to reactive oxygen species (ROS) (Yankovskaya *et al.*, 2003) is discussed in Chapter 3. The genes encoding SdhA and SdhB have been shown to be upregulated under cold stress conditions in *Propionibacterium freudenreichii* CIRM-BIA1^T (Dalmasso *et al.*, 2012) and in response to high pH stress in *Corynebacterium glutamicum* ATCC 13032 (Barriuso-Iglesias *et al.*, 2008).

4.3.2.3 Enrichment of biological processes associated with growth at 5 ºC

To assess whether specific molecular mechanisms were associated with the differentially expressed genes, gene ontology (GO-term) enrichment analysis was performed using GOEAST (Zheng and Wang, 2008) and DAVID 7.5 (Huang et al., 2008). Results from the DAVID 7.5 (Huang *et al.*, 2008) annotations indicated the gene compliment was significantly enriched in 69 functions and repressed in 79 functions (Appendix F 3 and 4). Furthermore, the differentially expressed genes (at 5 ºC) were assigned to 16 biological processes annotated in the KEGG database (Appendix F 5).

Further analyses of these pathways revealed significant enrichment of upregulated genes in the pentose phosphate pathway, fructose and mannose metabolism as well as ascorbate and aldarate metabolism at 5 ºC. (Appendix F 6). However, proteins associated with general metabolism were reported to be generally suppressed when the psychrophilic bacterium *Pseudoalteromonas haloplanktis* was grown at 4°C relative to 18°C (Piette *et al.*, 2011). An elaborate evaluation of transcriptome or proteome of cold adapted organism that put into cognisance the complex association of low temperature and other conditions, e.g., nutrient availability will likely provide a clearer picture of the metabolic activities of these organisms at low temperature.

Similarly, there was greater enrichment of upregulated genes at 5°C in the alanine, aspartate, glutamate, histidine metabolism and valine, leucine and isoleucine biosynthesis pathways. Conversely, the valine, leucine, isoleucine and lysine degradation pathways, and cysteine, methionine, tyrosine and tryptophan metabolism pathways were enriched by the downregulated genes at 5°C. The cold-induction of genes involved in ‘valine, leucine and isoleucine biosynthesis pathway’ and the corresponding suppression of the ‘valine, leucine and isoleucine degradation pathway’ at 5°C probably signifies an important psychrophilic adaptation in *Nesterenkonia* sp. AN1 (Klein *et al.*, 1999). In *Bacillus subtilis*, two mechanisms for the modulation of membranes fluidity at low temperature have been characterized. One of these involves the Isoleucine mediated biosynthesis of anteiso-branched C15:0 and C17:0 fatty

acids. Additionally, valine and leucine could also serve as precursors in the branched chain fatty acids biosynthetic pathways (Klein *et al.*, 1999).

It has been observed that the DAVID 7.5 (Huang *et al.*, 2008) annotation overestimates the enrichment of specific molecular functions, probably because of the large and non-specific background gene sets used by the program. The enrichment analysis based on GOEAST (Zheng and Wang, 2008) showed that three and five molecular functions were significantly ($p \leq 0.05$) upregulated and suppressed at 5 ºC, respectively (Appendix T 5, Figure 4-8). There was significant ($p \leq 0.05$) induction of "antioxidant activity" (GO: 0016209) at 5 ºC. This was enriched by the superoxide dismutase (*sodA*) and perioxiredoxin (*bcp*) genes which are prominent scavengers of reactive oxygen species (ROS) (Imlay, 2008, Moustafa *et al.*, 2010). Furthermore, two genes, coding for citrate lyase beta subunit (Mcl2) and isocitrate lyase (ICL), associated with oxo-acid-lyase activity (GO: 0016833) were significantly induced at 5 ºC. Isocitrate lyase has been shown to be upregulated in the psychrophilic bacteria, *Colwellia maris* (Watanabe *et al.*, 2002b, Watanabe *et al.*, 2002a) and *Psychrobacter cryohalolentis* K5 (Bakermans *et al.*, 2007) in response to cold. The malate synthase gene (*aceB*) was similarly upregulated. ICL and AceB are key enzymes of the glyoxylate cycle, present only in those microorganisms that are capable of utilising C_2 compounds (e.g. acetate and ethanol) or fatty acids as sole carbon sources for energy production (Berg *et al.*, 2002, Kunze *et al.*, 2006). Low temperature induction of these genes suggests a potential mechanism for energy acquisition by *Nesterenkonia* sp. AN1 in the nutrient-deficient soils of the Antarctic Dry Valleys.

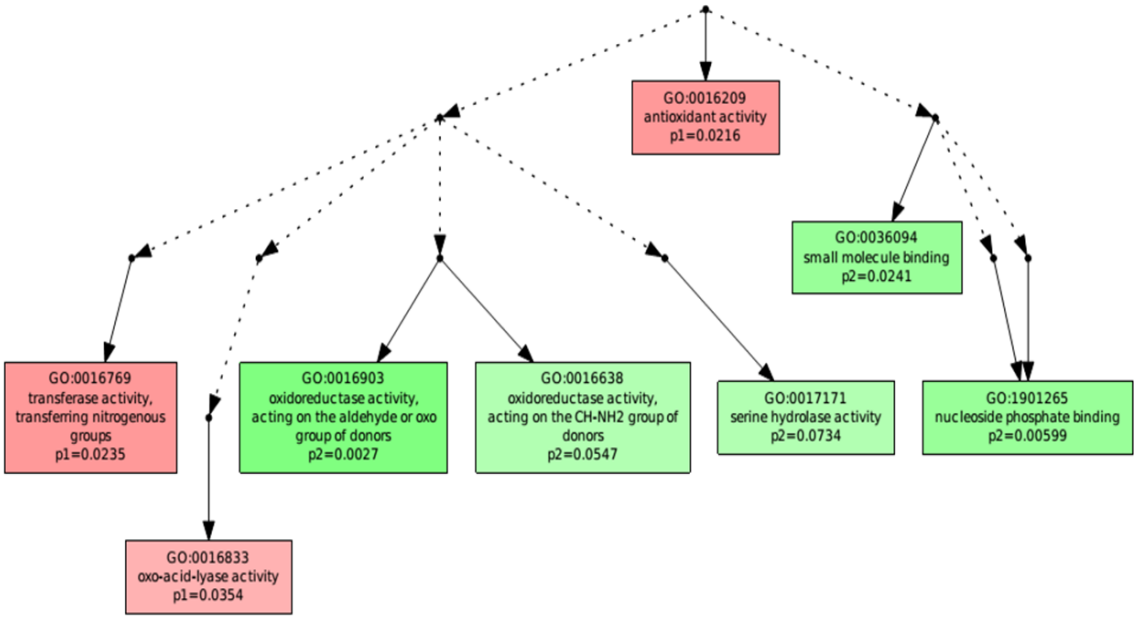


Figure 4-8: GEOEAST graph of significantly enriched molecular functions for differentially expressed genes in *Nesterenkonia* sp. AN1 grown at 5°C and 21°C. Red and green boxes represent the GO-terms for the upregulated and downregulated genes, respectively.

4.4 Conclusions

To provide additional insight into the adaptations of *Nesterenkonia* sp. AN1 to cold stress, the global RNA in *Nesterenkonia* sp. AN1 grown at 5 ºC and 21 ºC was sampled, sequenced and compared. The transcriptome profile of *Nesterenkonia* sp. AN1 grown at 5 ºC and 21 ºC showed that transcriptional signals were detected for ~ 97 % of the annotated genes, suggesting high sequencing coverage and accuracy of the gene prediction methods.

Evaluation of the differentially expressed transcripts showed that 164 and 152 genes were upregulated and downregulated, respectively, at 5 ºC. Strikingly, the majority of the transcripts coding for cold shock and cold acclimation were constitutively expressed. However, there was significant induction of transcripts that code for antioxidants at 5 ºC, demonstrated by the upregulation of *sodA*, *bcp* and *bpoA2*. The predominant strategy employed by the organism, as suggested by the data, centred on the induction of genes associated with scavenging of reactive oxygen species (ROS). The concentration of ROS have been shown to increase under low temperature as a result of cold induced increased

solubility of gasses (De Maayer *et al.*, 2014). *Nesterenkonia* sp. AN1 responded to possible ROS toxicity at 5 ºC by switching-on several of these genes with known antioxidant activities (De Maayer *et al.*, 2014, Margesin and Miteva, 2011).

There was also overexpression of universal stress protein genes related to *uspA*, along with genes encoding other characterized cold stress features. Genes encoding the two key enzymes of the glyoxylate cycle, isocitrate lyase (ICL) and malate synthase (AceB) were induced at 5 ºC, suggesting possible adaptation strategies for energy metabolism in cold habitats.

In addition to the induction of the genes reported to play a role in cold temperature adaptation in other psychrophilic and psychrotolerant microorganisms, there were also a substantial number of the transcripts which have not been previously characterized, that are cold-inducible in *Nesterenkonia* sp. AN1. These uncharacterized cold-induced transcripts may encode novel candidate cold adaptation features. In combination with the genome and comparative genomic data, the transcriptome analysis thus provides preliminary insights into the molecular determinants of cold survival and adaptation of *Nesterenkonia* sp. AN1 in Antarctic soils.

One major limitation of the current study was that, due to the prohibitive costs of transcriptome sequencing, only two transcriptome replicates were sequenced for each of the test conditions. A third replicate should be performed to provide additional statistical support. It could also be argued that the growth of *Nesterenkonia* sp. AN1 in nutrient-rich LB medium may affect the transcriptome analyses. However, as the same medium was employed for growth of this strain at both 5 ºC and 21 ºC, this should not impact the global differentiation in terms of genes induced and repressed at the different growth temperatures.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

4.5 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 Biol. Biochem.,* 38**,** 3041-3056.

ALIYU, H., DE MAAYER, P., REES, J., TUFFIN, M. & COWAN, D. A. 2014. Draft Genome Sequence of the Antarctic Polyextremophile *Nesterenkonia* sp. Strain AN1. *Genome Announc,* 2.

BAKERMANS, C., TOLLAKSEN, S. L., GIOMETTI, C. S., WILKERSON, C., TIEDJE, J. M. & THOMASHOW, M. F. 2007. Proteomic analysis of *Psychrobacter cryohalolentis* K5 during growth at subzero temperatures. *Extremophiles,* 11**,** 343-354.

BARRETT, T., WILHITE, S. E., LEDOUX, P., EVANGELISTA, C., KIM, I. F., TOMASHEVSKY, M., MARSHALL, K. A., PHILLIPPY, K. H., SHERMAN, P. M. & HOLKO, M. 2013. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.,* 41**,** D991-D995.

BARRIUSO-IGLESIAS, M., SCHLUESENER, D., BARREIRO, C., POETSCH, A. & MARTÍN, J. F. 2008. Response of the cytoplasmic and membrane proteome of *Corynebacterium glutamicum* ATCC 13032 to pH changes. *BMC Microbiol.,* 8**,** 225.

BERG, J. M., TYMOCZKO, J. L. & STRYER, L. 2002. The Glyoxylate Cycle Enables Plants and Bacteria to Grow on Acetate. *In:* FREEMAN, W. H. (ed.) *Biochemistry.* 5th edition ed. New York.

COWAN, D. A. & AH TOW, L. 2004. Endangered antarctic environments. *Annu. Rev. Microbiol.,* 58**,** 649-690.

DALMASSO, M., AUBERT, J., BRIARD-BION, V., CHUAT, V., DEUTSCH, S.-M., EVEN, S., FALENTIN, H., JAN, G., JARDIN, J., MAILLARD, M.-B., PARAYRE, S., PIOT, M., TANSKANEN, J. & THIERRY, A. 2012. A Temporal -omic Study of *Propionibacterium freudenreichii* CIRM-BIA1^{T} Adaptation Strategies in Conditions Mimicking Cheese Ripening in the Cold. *PLoS One,* 7**,** e29083.

DE MAAYER, P., ANDERSON, D., CARY, C. & COWAN, D. A. 2014. Some like it cold: understanding the survival strategies of psychrophiles. *EMBO Rep,* 15**,** 508-517.

DENOME, S. A., ELF, P. K., HENDERSON, T. A., NELSON, D. E. & YOUNG, K. D. 1999. *Escherichia coli* Mutants Lacking All Possible Combinations of Eight Penicillin Binding Proteins: Viability, Characteristics, and Implications for Peptidoglycan Synthesis. *J. Bacteriol.,* 181**,** 3981-3993.

DSOUZA, M., TAYLOR, M. W., TURNER, S. J. & AISLABIE, J. 2014. Genome-based comparative analyses of Antarctic and temperate species of *Paenibacillus*. *PLoS One,* 9**,** e108009.

DSOUZA, M., TAYLOR, M. W., TURNER, S. J. & AISLABIE, J. 2015. Genomic and phenotypic insights into the ecology of *Arthrobacter* from Antarctic soils. *BMC Genomics,* 16**,** 36.

ENSTROM, M., HELD, K., RAMAGE, B., BRITTNACHER, M., GALLAGHER, L. & MANOIL, C. 2012. Genotype-phenotype associations in a nonmodel prokaryote. *MBio,* 3**,** e00001-12.

GARNIER, M., MATAMOROS, S., CHEVRET, D., PILET, M.-F., LEROI, F. & TRESSE, O. 2010. Adaptation to cold and proteomic responses of the psychrotrophic biopreservative *Lactococcus piscium* strain CNCM I-4031. *Appl. Environ. Microbiol.,* 76**,** 8011-8018.

GRAUMANN, P. L. & MARAHIEL, M. A. 1999. Cold shock response in Bacillus subtilis. *J. Mol. Microbiol. Biotechnol.,* 1**,** 203-209.

HAENNI, M., MAJCHERCZYK, P. A., BARBLAN, J.-L. & MOREILLON, P. 2006. Mutational Analysis of Class A and Class B Penicillin-Binding Proteins in *Streptococcus gordonii*. *Antimicrob Agents Chemother,* 50**,** 4062-4069.

HUANG, D. W., SHERMAN, B. T. & LEMPICKI, R. A. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Proto,* 4**,** 44-57.

HUYNEN, M. A., DANDEKAR, T. & BORK, P. 1999. Variation and evolution of the citric-acid cycle: a genomic perspective. *Trends Microbiol.,* 7**,** 281-291.

IMLAY, J. A. 2008. Cellular defenses against superoxide and hydrogen peroxide. *Annu. Rev. Biochem.,* 77**,** 755-776.

KIRBY, B., BARNARD, D., MARLA TUFFIN, I. & COWAN, D. 2011. Ecological Distribution of Microorganisms in Terrestrial, Psychrophilic Habitats. *In:* HORIKOSHI, K. (ed.) *Extremophiles Handbook.* Springer Japan, pp. 839-863.

KLEIN, W., WEBER, M. H. & MARAHIEL, M. A. 1999. Cold shock response of Bacillus subtilis: isoleucine-dependent switch in the fatty acid branching pattern for membrane adaptation to low temperatures. *J. Bacteriol.,* 181**,** 5341-5349.

KOGENARU, S., YAN, Q., GUO, Y. & WANG, N. 2012. RNA-seq and microarray complement each other in transcriptome profiling. *BMC Genomics,* 13**,** 629.

KUNZE, M., PRACHAROENWATTANA, I., SMITH, S. M. & HARTIG, A. 2006. A central role for the peroxisomal membrane in glyoxylate cycle function. *Biochim. Biophys. Acta,* 1763**,** 1441-1452.

LEE, P. T., HSU, A. Y., HA, H. T. & CLARKE, C. F. 1997. A C-methyltransferase involved in both ubiquinone and menaquinone biosynthesis: isolation and identification of the *Escherichia coli* ubiE gene. *J. Bacteriol.,* 179**,** 1748-1754.

LETUNIC, I., YAMADA, T., KANEHISA, M. & BORK, P. 2008. iPath: interactive exploration of biochemical pathways and networks. *Trends Biochem. Sci.,* 33**,** 101-103.

LOCKHART, D. J. & WINZELER, E. A. 2000. Genomics, gene expression and DNA arrays. *Nature,* 405**,** 827-836.

MARGESIN, R. & MITEVA, V. 2011. Diversity and ecology of psychrophilic microorganisms. *Res. Microbiol.,* 162**,** 346-61.

MEHTA, A. P., ABDELWAHED, S. H., XU, H. & BEGLEY, T. P. 2014. Molybdopterin Biosynthesis: Trapping of Intermediates for the MoaA-Catalyzed Reaction Using 2′-DeoxyGTP and 2′-ChloroGTP as Substrate Analogues. *J. Am. Chem. Soc.,* 136**,** 10609-10614.

MENDEL, R. R. & BITTNER, F. 2006. Cell biology of molybdenum. *Biochim. Biophys. Acta,* 1763**,** 621-635.

MOUSTAFA, D. A., JAIN, N., SRIRANGANATHAN, N. & VEMULAPALLI, R. 2010. Identification of a single-nucleotide insertion in the promoter region affecting the sodc promoter activity in *Brucella neotomae*. *PloS One,* 5**,** e14112.

NEL, A., TUFFIN, I., SEWELL, B. & COWAN, D. 2011. Unique aliphatic amidase from a psychrotrophic and haloalkaliphilic *Nesterenkonia* isolate. *Appl. Environ. Microbiol.,* 77**,** 3696-3702.

NEUMANN, M. & LEIMKÜHLER, S. 2010. The role of system-specific molecular chaperones in the maturation of molybdoenzymes in bacteria. *Biochem Res Int,* 2011.

PHADTARE, S. & INOUYE, M. 2001. Role of CspC and CspE in Regulation of Expression of RpoS and UspA, the Stress Response Proteins in *Escherichia coli*. *J. Bacteriol.,* 183**,** 1205-1214.

PIETTE, F., D'AMICO, S., MAZZUCCHELLI, G., DANCHIN, A., LEPRINCE, P. & FELLER, G. 2011. Life in the cold: a proteomic study of cold-repressed proteins in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Appl. Environ. Microbiol.,* 77**,** 3881-3883.

POLISSI, A., DE LAURENTIS, W., ZANGROSSI, S., BRIANI, F., LONGHI, V., PESOLE, G. & DEHÒ, G. 2003. Changes in *Escherichia coli* transcriptome during acclimatization at low temperature. *Res. Microbiol.,* 154**,** 573-580.

POWELL, S., SZKLARCZYK, D., TRACHANA, K., ROTH, A., KUHN, M., MULLER, J., ARNOLD, R., RATTEI, T., LETUNIC, I. & DOERKS, T. 2012. eggNOG v3. 0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.,* 40**,** D284-D289.

REGULSKI, E. E., MOY, R. H., WEINBERG, Z., BARRICK, J. E., YAO, Z., RUZZO, W. L. & BREAKER, R. R. 2008. A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Mol. Microbiol.,* 68**,** 918-932.

RONHOLM, J., RAYMOND-BOUCHARD, I., CRESKEY, M., CYR, T., CLOUTIS, E. A. & WHYTE, L. G. 2015. Characterizing the surface-exposed proteome of *Planococcus halocryophilus* during cryophilic growth. *Extremophiles,* 19**,** 619-629.

SEIFART GOMES, C., IZAR, B., PAZAN, F., MOHAMED, W., MRAHEIL, M. A., MUKHERJEE, K., BILLION, A., AHARONOWITZ, Y., CHAKRABORTY, T. & HAIN, T. 2011. Universal Stress Proteins Are Important for Oxidative and Acid Stress Resistance and Growth of *Listeria monocytogenes* EGD-e In Vitro and In Vivo. *PLoS One,* 6**,** e24965.

SIDDIQUI, K. S., WILLIAMS, T. J., WILKINS, D., YAU, S., ALLEN, M. A., BROWN, M. V., LAURO, F. M. & CAVICCHIOLI, R. 2013. Psychrophiles. *Annu Rev Earth Planet Sci,* 41**,** 87-115.

TATUSOV, R. L., FEDOROVA, N. D., JACKSON, J. D., JACOBS, A. R., KIRYUTIN, B., KOONIN, E. V., KRYLOV, D. M., MAZUMDER, R., MEKHEDOV, S. L., NIKOLSKAYA, A. N., RAO, B. S., SMIRNOV, S., SVERDLOV, A. V., VASUDEVAN, S., WOLF, Y. I., YIN, J. J. & NATALE, D. A. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics,* 4**,** 41.

UNELL, M., ABRAHAM, P. E., SHAH, M., ZHANG, B., RÜCKERT, C., VERBERKMOES, N. C. & JANSSON, J. K. 2009. Impact of Phenolic Substrate and Growth Temperature on the *Arthrobacter chlorophenolicus* Proteome. *J Proteome Res,* 8**,** 1953-1964.

WATANABE, S., YAMAOKA, N., FUKUNAGA, N. & TAKADA, Y. 2002a. Purification and characterization of a cold-adapted isocitrate lyase and expression analysis of the cold-inducible isocitrate lyase gene from the psychrophilic bacterium *Colwellia psychrerythraea*. *Extremophiles,* 6**,** 397-405.

WATANABE, S., YAMAOKA, N., TAKADA, Y. & FUKUNAGA, N. 2002b. The cold-inducible icl gene encoding thermolabile isocitrate lyase of a psychrophilic bacterium, *Colwellia maris*. *Microbiology,* 148**,** 2579-2589.

WU, S., ZHU, Z., FU, L., NIU, B. & LI, W. 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics,* 12**,** 444.

YAMADA, T., LETUNIC, I., OKUDA, S., KANEHISA, M. & BORK, P. 2011. iPath2. 0: interactive pathway explorer. *Nucleic Acids Res.,* 39**,** W412-W415.

YANKOVSKAYA, V., HORSEFIELD, R., TÖRNROTH, S., LUNA-CHAVEZ, C., MIYOSHI, H., LÉGER, C., BYRNE, B., CECCHINI, G. & IWATA, S. 2003. Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science,* 299**,** 700-704.

ZELLER, T. & KLUG, G. 2006. Thioredoxins in bacteria: functions in oxidative stress response and regulation of thioredoxin genes. *Naturwissenschaften,* 93**,** 259-266.

ZHENG, Q. & WANG, X.-J. 2008. GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic Acids Res.,* 36**,** W358-W363.

Chapter 5

5 General conclusions

5.1 Introduction

This thesis reports on investigations into the adaptation strategies that potentially contribute to the survival of *Nesterenkonia* sp. AN1 in the coldest terrestrial environment on Earth (Cowan *et al.*, 2014, Nel *et al.*, 2011). The dry coastal soils of Antarctica are characterised by a wide range of complex 'stress' factors such as extreme cold and aridity, high alkalinity and salinity, fluxes of UV irradiation and scarcity of nutrients (Cary *et al.*, 2010, Cowan *et al.*, 2014). Although there is a large body of data on microbial abundance in Antarctic soils (Cowan *et al.*, 2014), very little is known about the strategies that are used by the 'resident' psychrotolerant bacteria (De Maayer *et al.*, 2014). As the first member of the haloalkaliphilic genus *Nesterenkonia* to be isolated from Antarctic soils, it was hypothesised that the genome sequence of *Nesterenkonia* sp. AN1 would add to the growing body of knowledge on the genetic mechanisms that are employed by the cold-active bacteria. To our knowledge, this is the first comprehensive study combining genomic, comparative genomic and transcriptomics approaches to elucidate the survival strategies of a psychrotolerant organism inhabiting the arid soils of Antarctica.

5.2 Adaptations of the Antarctic *Nesterenkonia*

The draft genome of *Nesterenkonia* sp. AN1 was assembled into 41 contigs, with a total genome size of ~ 3.04 Mb. The genome had a mean G+C content of 67.4 % and encodes 2,853 proteins, (Aliyu *et al.*, 2014). The *Nesterenkonia* sp. AN1 genome is the largest known genome for this genus. The large genome size could be attributed to the acquisition possibly by lateral gene transfer of genetic elements, evident by the presence of several genomic islands.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

The genome of *Nesterenkonia* sp. AN1 encodes a repertoire of genomic features linked to the stresses typical of its Antarctic edaphic origin (Cary *et al.*, 2010, Cowan *et al.*, 2014). These include several proteins involved in osmotic and oxidative stress responses, modulation of membrane fluidity as well as cold stress and cold acclimation proteins.

Comparison of the *Nesterenkonia* sp. AN1 genome with those of three mesophilic members of the genus showed that over 50 % of the stress response features are shared with the mesophilic *Nesterenkonia* strains, suggesting an inherent resilience to environmental stresses for members of the genus. Analysis of the shared (core) genome fraction, however, revealed evolutionary innovations in the form of positive selection in the different *Nesterenkonia* strains. The *Nesterenkonia* sp. AN1 branch showed notable signatures of positive selection in genes encoding proteins linked to defence against and repair of damages from UV irradiation (Sghaier *et al.*, 2008) and genes coding for proteins associated with tolerance to oxidative stress (Imlay, 2003, Outten *et al.*, 2004, Rustin and Rötig, 2002, Santos *et al.*, 2014, Yankovskaya *et al.*, 2003). High fluxes of UV and oxidative stress are among the major stress factors in the edaphic habitat of Antarctica (Cary *et al.*, 2010, Cowan *et al.*, 2014). The presence of pervasive positive selection suggests that *Nesterenkonia* sp. AN1 may have incorporated mutations into its genome which could enhance its adaptation capacity in the Antarctic soils. In addition to the adaptations of the core genome, a sizeable portion of the stress adaptation proteins were also encoded on the strain-specific fraction of *Nesterenkonia* sp. AN1 genome. These include proteins linked oxidative, osmotic and general stress responses. It was not immediately apparent what evolutionary events are associated with all the singletons in *Nesterenkonia* sp. AN1 genome (whether acquired via horizontal gene transfer or events of gene decay in the mesophilic genome) (Mira *et al.*, 2001, Ventura *et al.*, 2007).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Analysis of the *Nesterenkonia* sp. AN1 transcriptome when the organism was grown at 5 ºC and 21 ºC showed that ~ 97 % of the predicted genes were expressed. Evaluation of the differentially expressed transcript showed that a substantial number of the transcripts were cold-inducible. These include thirteen transcript subject to positive selection and several other transcripts which have not been previously characterised. The uncharacterised cold-induced transcripts may encode novel candidate cold adaptation proteins. Several transcript that code for proteins involved in defence against reactive oxygen species and transcript associated with TCA and glyoxylate (Berg *et al.*, 2002, Kunze *et al.*, 2006) cycles were also cold-induced and may be linked to scavenging of reactive oxygen species and survival in nutrient-scarce environments. The cold-inducible enzymes of the TCA have been variously linked to the production of intermediary molecules involved in scavenging ROS (Amato, 2013, Imlay, 2003, Rustin and Rötig, 2002, Yankovskaya *et al.*, 2003). While the induction of TCA and glyoxylate enzymes may enable an organisms utilize different carbon sources (Han *et al.*, 2008), cold induction of the glyoxylate cycle is especially crucial for survival in nutrient-scarce environments (Berg *et al.*, 2002, Kunze *et al.*, 2006).

Taken together, the combined genomic and transcriptomic data suggest that polyextremophily in the Antarctic *Nesterenkonia* sp. AN1 is likely driven by the presence of pervasive adaptation strategies to cope with cold and associated stresses (De Maayer *et al.*, 2014, Mesbah and Wiegel, 2008).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

5.3 Concluding remarks and future perspectives

The draft genome of *Nesterenkonia* sp. AN1 is, to date, the only genome sequence of a cold adapted member of the genus *Nesterenkonia*. This is also the first comprehensive study, which combines genomic, comparative genomic and transcriptomics approaches, to elucidate the survival strategies of a psychrotolerant organism inhabiting the arid soils of Antarctica.

Future research on *Nesterenkonia* sp. AN1 would benefit from the resolution of taxonomic status of the organism via physiochemical characterization, DNA-DNA hybridization and the sequencing and comparison of the genomes of the type strains of species within the genus *Nesterenkonia.* Furthermore, a substantial number of the identified genes as well as sizeable proportion of the cold-induced transcripts were either annotated as hypothetical or of unknown function. These unknown gene are likely to encode new adaptation strategies. To harness this potential, however, there is the need for the characterization of the genes to unravel their specific roles. The “pan-genome” of the genus *Nesterenkonia* presented in this thesis was based on the genomes of only four strains. The inclusion of more genomes, especially the type strains of *Nesterenkonia*, will enable a more robust description of the pan-genome and consequently, a better understanding of the evolution of members of the genus. Finally, the transcriptome analysis presented in the thesis was limited to cold adaptations. However, as highlighted in the preceding chapters, cold stress in the natural environment of *Nesterenkonia* sp. AN1 (the Antarctic soils) is compounded by many other stresses. Transcriptomic or proteomics using a combination of multiple stress conditions that reflects the soil condition of the Antarctic soils will enable a more comprehensive picture of the lifestyle of the polyextremophile, *Nesterenkonia* sp. AN1.

5.4 References

ALIYU, H., DE MAAYER, P., REES, J., TUFFIN, M. & COWAN, D. A. 2014. Draft Genome Sequence of the Antarctic Polyextremophile *Nesterenkonia* sp. Strain AN1. *Genome Announc,* 2.

AMATO, P. 2013. Energy metabolism in low-temperature and frozen conditions in cold-adapted microorganism. *In:* YUMOTO, I. (ed.) *Cold-adapted microorganisms.* Norfolk, UK: Caister Academic Press, pp. 71-95.

BERG, J. M., TYMOCZKO, J. L. & STRYER, L. 2002. The Glyoxylate Cycle Enables Plants and Bacteria to Grow on Acetate. *In:* FREEMAN, W. H. (ed.) *Biochemistry.* 5th edition ed. New York.

CARY, S. C., MCDONALD, I. R., BARRETT, J. E. & COWAN, D. A. 2010. On the rocks: the microbiology of Antarctic Dry Valley soils. *Nat. Rev. Microbiol.,* 8**,** 129-138.

COWAN, D. A., MAKHALANYANE, T. P., DENNIS, P. G. & HOPKINS, D. W. 2014. Microbial ecology and biogeochemistry of continental Antarctic soils. *Front Microbiol,* 5**,** 154.

DE MAAYER, P., ANDERSON, D., CARY, C. & COWAN, D. A. 2014. Some like it cold: understanding the survival strategies of psychrophiles. *EMBO Rep,* 15**,** 508-517.

HAN, S. O., INUI, M. & YUKAWA, H. 2008. Effect of carbon source availability and growth phase on expression of Corynebacterium glutamicum genes involved in the tricarboxylic acid cycle and glyoxylate bypass. *Microbiology,* 154**,** 3073-3083.

IMLAY, J. A. 2003. Pathways of Oxidative Damage. *Annu. Rev. Microbiol.,* 57**,** 395-418.

KUNZE, M., PRACHAROENWATTANA, I., SMITH, S. M. & HARTIG, A. 2006. A central role for the peroxisomal membrane in glyoxylate cycle function. *Biochim. Biophys. Acta,* 1763**,** 1441-1452.

MESBAH, N. M. & WIEGEL, J. 2008. Life at Extreme Limits. *Ann. N. Y. Acad. Sci.,* 1125**,** 44-57.

MIRA, A., OCHMAN, H. & MORAN, N. A. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet.,* 17**,** 589-596.

NEL, A., TUFFIN, I., SEWELL, B. & COWAN, D. 2011. Unique aliphatic amidase from a psychrotrophic and haloalkaliphilic *Nesterenkonia* isolate. *Appl. Environ. Microbiol.,* 77**,** 3696-3702.

OUTTEN, F. W., DJAMAN, O. & STORZ, G. 2004. A suf operon requirement for Fe–S cluster assembly during iron starvation in *Escherichia coli*. *Mol. Microbiol.,* 52**,** 861-872.

RUSTIN, P. & RÖTIG, A. 2002. Inborn errors of complex II – Unusual human mitochondrial diseases. *Biochim. Biophys. Acta,* 1553**,** 117-122.

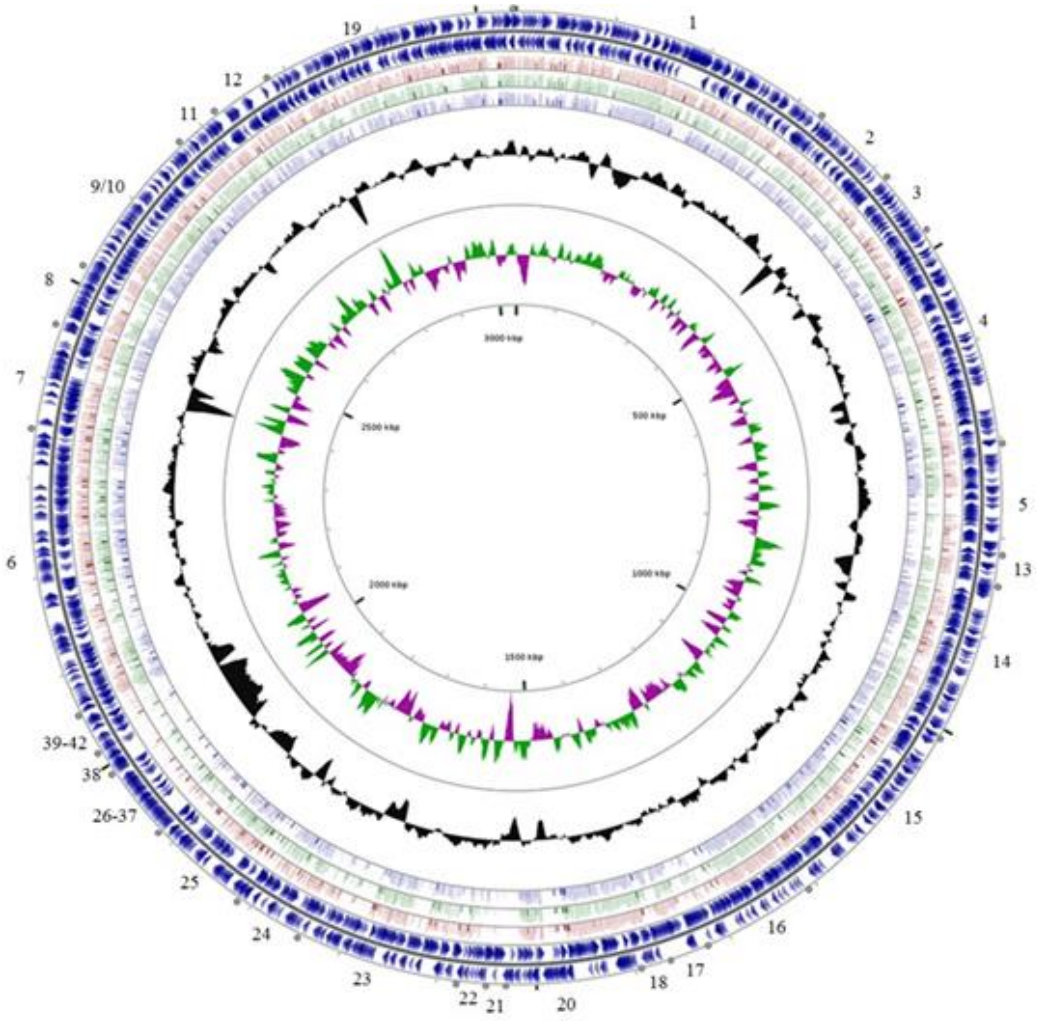
SANTOS, J. A., ALONSO-GARCÍA, N., MACEDO-RIBEIRO, S. & PEREIRA, P. J. B. 2014. The unique regulation of iron-sulfur cluster biogenesis in a Gram-positive bacterium. *Proc Natl Acad Sci U S A,* 111**,** E2251-E2260.

SGHAIER, H., GHEDIRA, K., BENKAHLA, A. & BARKALLAH, I. 2008. Basal DNA repair machinery is subject to positive selection in ionizing-radiation-resistant bacteria. *BMC Genomics,* 9**,** 297.

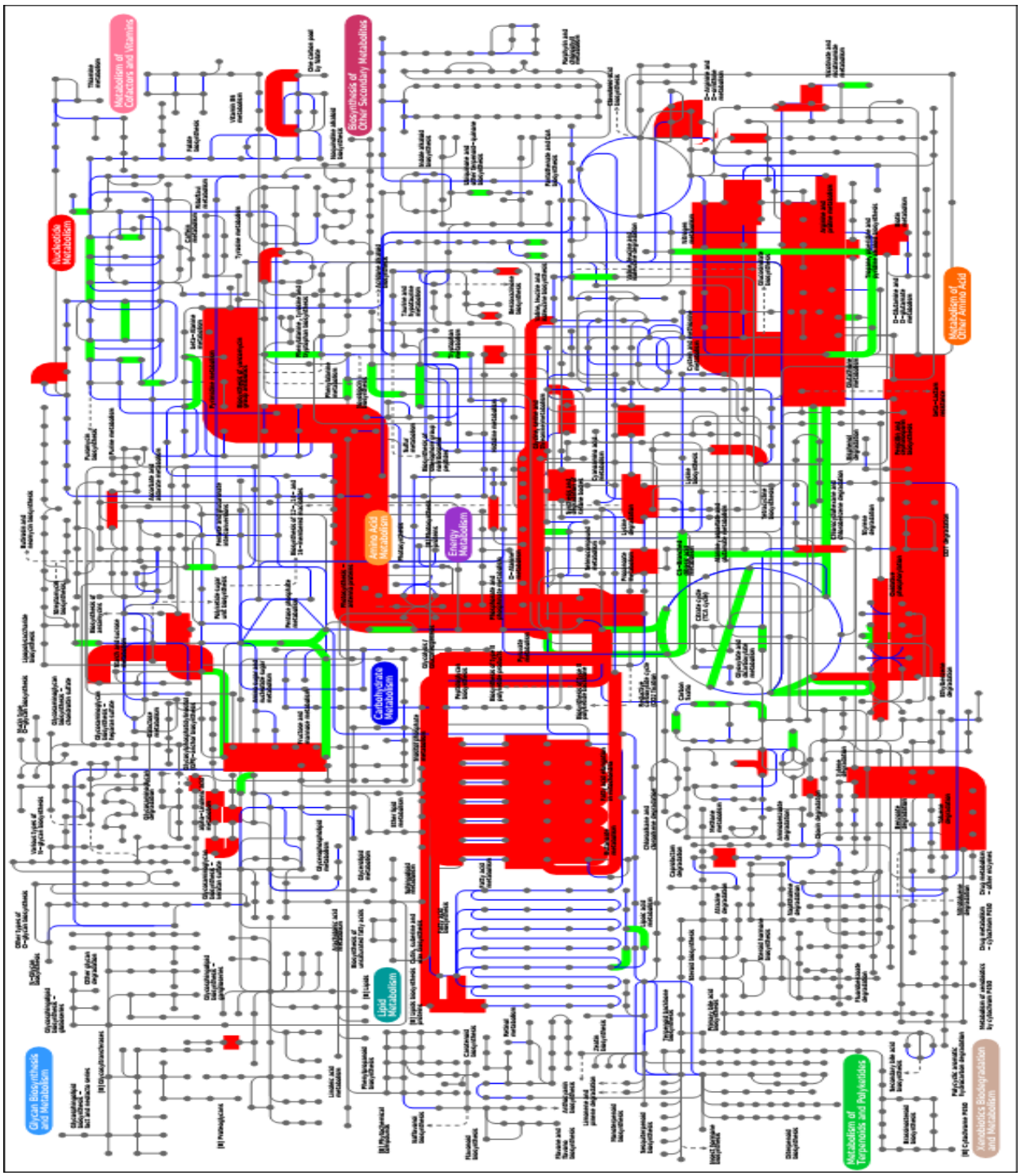
VENTURA, M., CANCHAYA, C., TAUCH, A., CHANDRA, G., FITZGERALD, G. F., CHATER, K. F. & VAN SINDEREN, D. 2007. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol. Mol. Biol. Rev.,* 71**,** 495-548.

YANKOVSKAYA, V., HORSEFIELD, R., TÖRNROTH, S., LUNA-CHAVEZ, C., MIYOSHI, H., LÉGER, C., BYRNE, B., CECCHINI, G. & IWATA, S. 2003. Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science,* 299**,** 700-704.

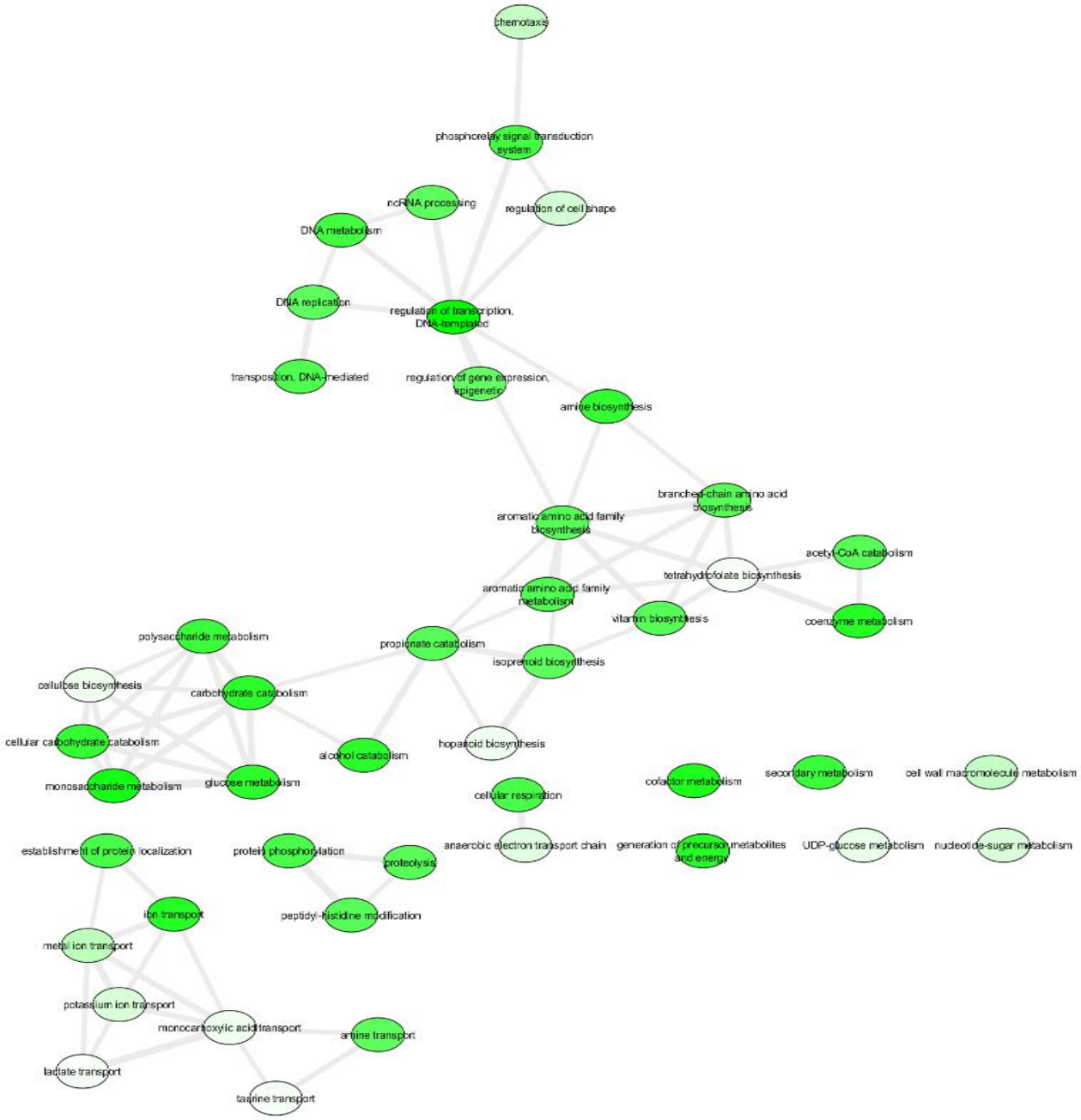
Appendices



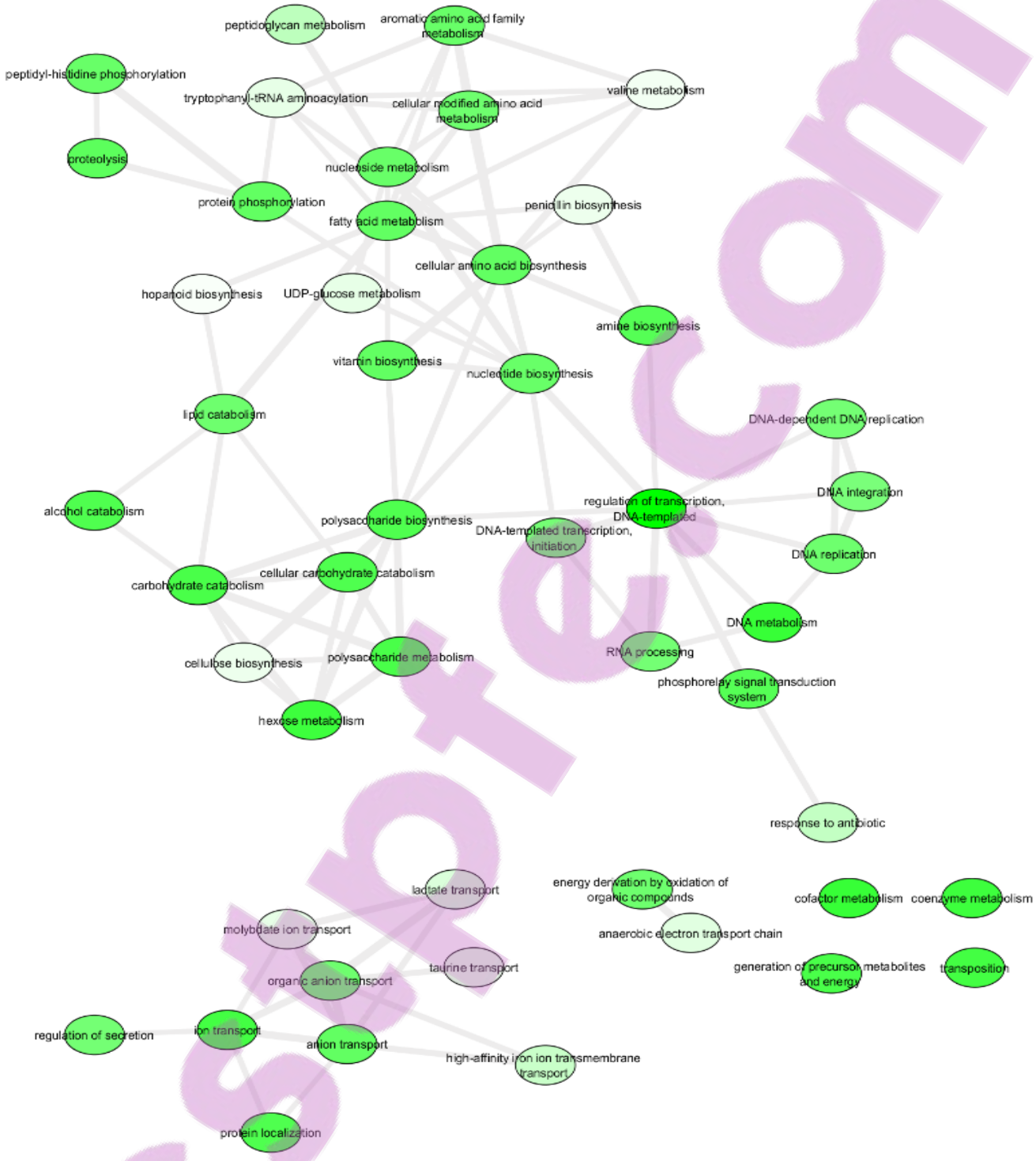
Appendix F 1: Circular representation of the genome of *Nesterenkonia* sp. AN1 and three mesophilic *Nesterenkonia* strains plotted using the CG View Server (Grant and Stothard, 2008). The first two rings from the outside show CDSs encoded on the positive (first ring) and negative (second ring) strands of the *Nesterenkonia* sp. AN1 genome. The 3rd, 4th and 5th ring represent the CDSs encoded on the genomes of *Nesterenkonia* sp. F, NP1 and *N. alba* DSM 19423, respectively. The sixth and seventh rings show the G+C content plot and GC skew (green indicating GC skew + and purple GC skew -), respectively, for the *Nesterenkonia* sp. AN1.



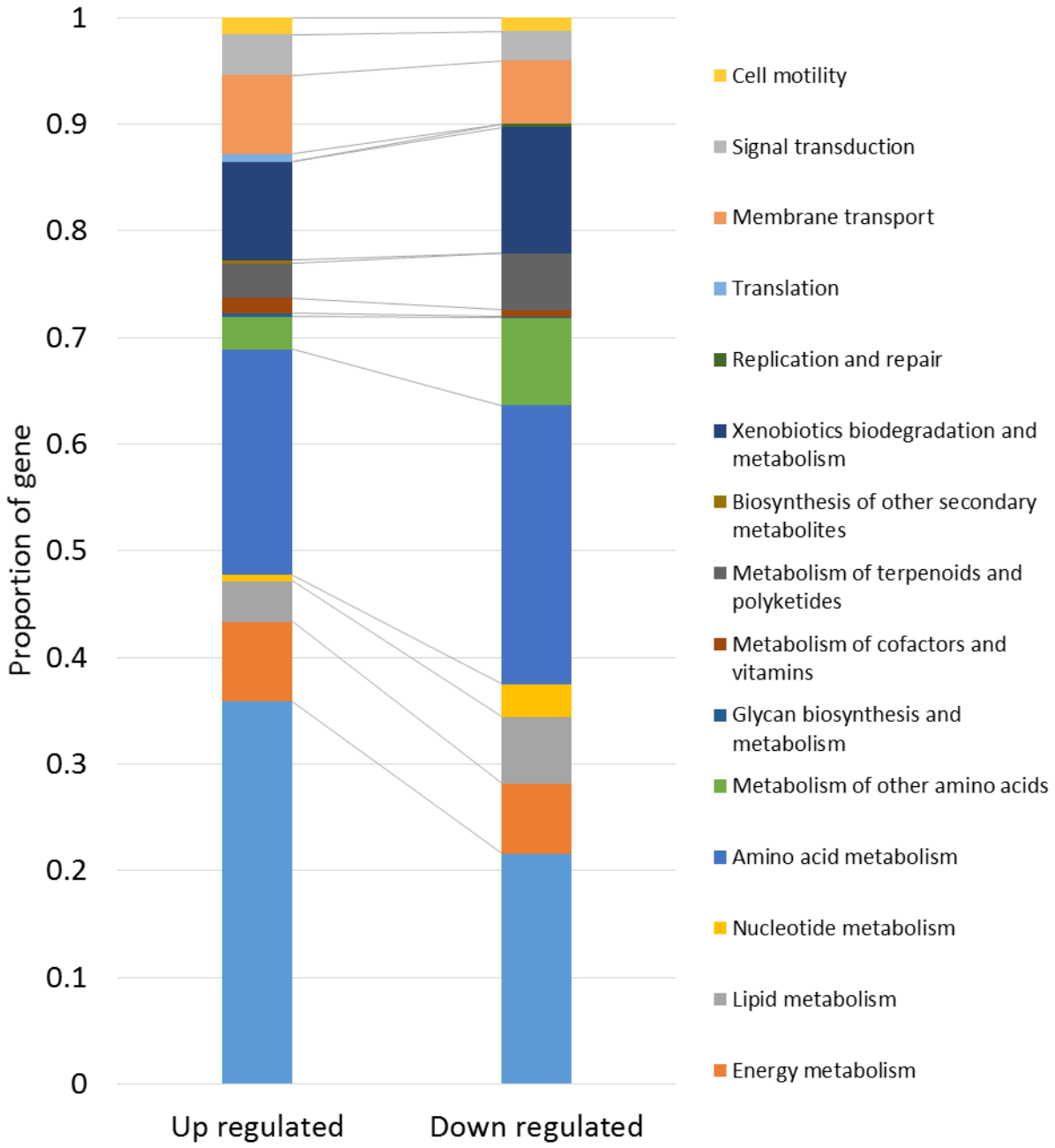
Appendix F 2: Overview of the transcriptome of *Nesterenkonia* sp. AN1 grown at 5°C and 21°C mapped against KEGG metabolic pathways using Ipath2.0 (Letunic *et al.*, 2008, Yamada *et al.*, 2011). .The up-regulated, down-regulated and the constitutively expressed genes at 5°C are shown in green, red and blue, respectively. The line width indicates expression levels for the up-regulated and down-regulated genes.



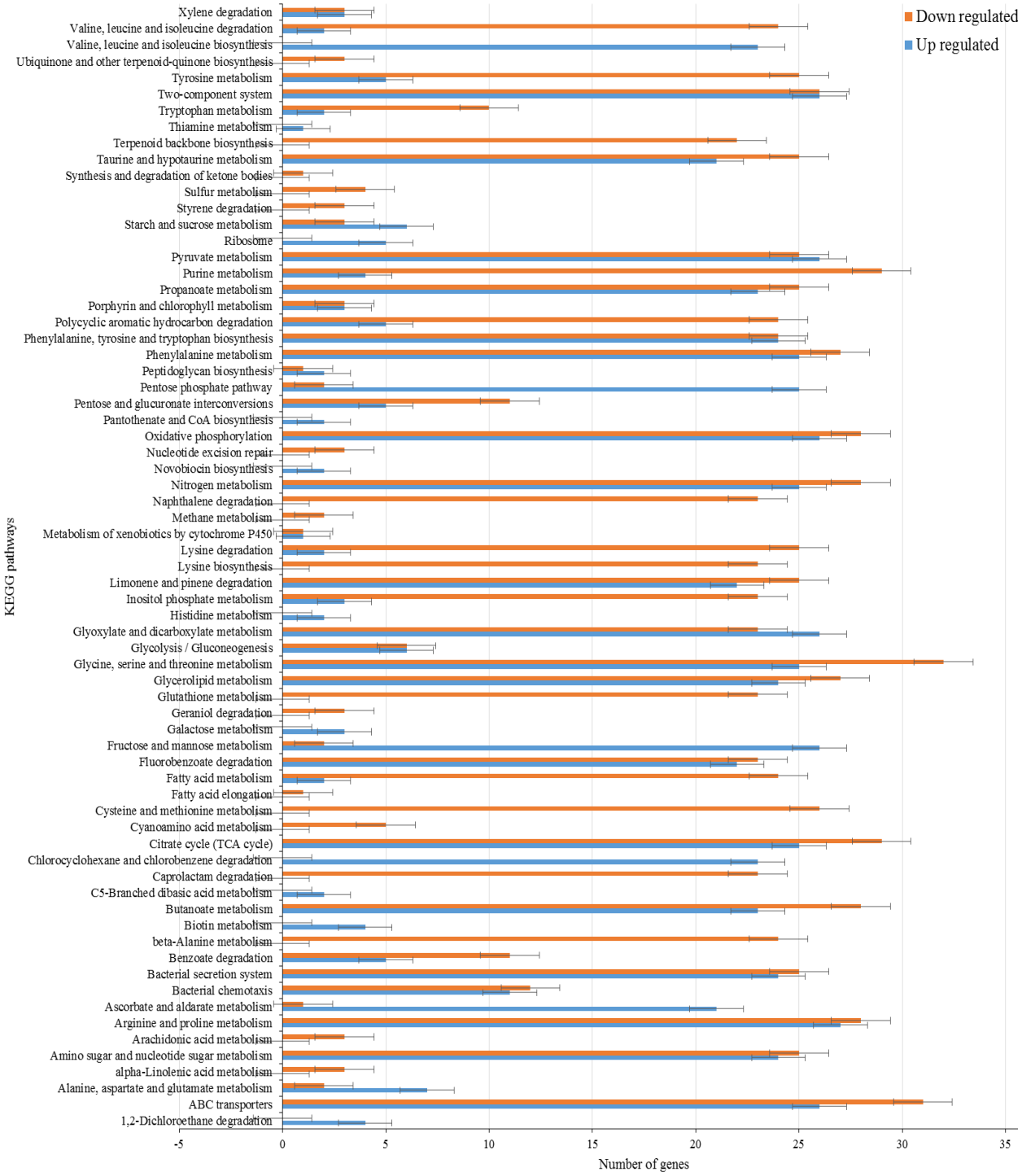
Appendix F 3: REVIGO (Supek *et al.*, 2011) interactive graph showing gene ontology (GO) terms enriched by the overexpressed genes at 5°C relative to 21°C in *Nesterenkonia* sp. AN1. The intensity of the bobble colour indicates significant levels of the enriched GO terms and the width of the edges indicates level of similarity between the GO terms.



Appendix F 4: REVIGO (Supek *et al.*, 2011) interactive graph showing gene ontology (GO) terms enriched by the repressed genes at 5°C relative to 21°C in *Nesterenkonia* sp. AN1. The intensity of the bobble colour indicates significant levels of the enriched GO terms and the width of the edges indicates level of similarity between the GO terms.



Appendix F 5: Proportion of differentially expressed genes associated with different KEGG pathways for *Nesterenkonia* sp. AN1grown at 5°C and 21°C.



Appendix F 6: Number of differentially expressed genes at 5°C relative to 21°C enriched in specific KEGG pathways. The error bars indicate the standard error for the compared gene numbers.

Appendix T 1: The genome islands (GIs) identified in *Nesterenkonia* sp. AN1.

Genomic Island	Size (kb)	G+C content	G+C deviation from chromosome	Number of CDSs	Potential features encoded on GI
GI1	21.31	67.88	-0.48	25	ABC-transporter ATP binding protein, ABC-transporter membrane protein, ABC transporter permease, alkyl hydro peroxidase;
GI2	5.92	67.01	0.39	4	DNA mismatch repair enzyme, hypothetical protein (3) and an unknown protein
GI3	16.71	68.61	-1.21	19	ABC transporter substrate-binding protein, hypothetical proteins (4), integrase (2), mobilisation proteins (2), reticulocyte binding protein 2 homolog A, transcriptional regulator, transposase, transposase-like protein and unknown proteins (6)
GI4	4.24	70.38	-2.98	6	AsnC family transcriptional regulator, hypothetical proteins (3), permease and transposase
GI5	14.57	64.6	2.80	13	Cobalt transporter, dTDP-4-dehydrorhamnose reductase, dTDP-glucose 4:6-dehydratase, glucose-1-phosphate thymidylyl transferase, glycosyl transferase family 28, histidine kinase, hypothetical protein (4), integrase and transposase
GI6	4.64	53.59	13.81	8	Cell division protein FtsW, hypothetical protein (3) and lipoate-protein ligase A
GI7	7.79	66.91	0.49	9	Oligoribonuclease, membrane proteins (3), hypothetical protein (2) and unknown (3)
GI8	6.70	69.55	-2.15	8	Heat shock protein Hsp20 (2), ornithine aminotransferase (2), transposase (2) and universal stress protein UspA
GI9	4.03	65.01	2.39	6	Flavodoxin (2), hypothetical protein (2) and integrase
GI10	4.21	67.38	0.02	8	Cytochrome C biogenesis protein; ResB, ResC (3), hypothetical protein (2), thiol-disulfide isomerase and unknown (2)

GI11	6.60	64.44	2.96	12	Hypothetical protein (5), N-acetyltransferase GCN5, phosphotransferase, potassium transporter, Potassium uptake protein KtrB and KtrC, PTS lactose transporter subunit IIC and unknown (2)
GI12	8.62	67.7	-0.30	14	Hydrolase, hypothetical protein (5), integrase (2), membrane protein, prevent-host-death family protein and unknown (4)
GI13	4.09	69.92	-2.52	5	DNA polymerase III subunit alpha, Na+/proline symporter, transcriptional regulator NrdR (2) and an unknown protein
GI14	12.80	69.63	-2.23	10	1-acyl-sn-glycerol-3-phosphate acyltransferase, excinuclease ABC subunits A and C, glycoside hydrolase families 2 & 16 protein, hypothetical protein, PBS lyase and response regulator receiver protein
GI15	11.38	67.48	-0.08	11	Chromosome partitioning protein, exopolysaccharide biosynthesis protein, glycosyl transferase, hypothetical protein, low molecular weight protein-tyrosine-phosphatase, polyprenyl glycosyl phosphotransferase, sugar/glycosyl transferase, transposase (2) and UDP-N-acetyl glucosamine 1-carboxyvinyltransferase
GI16	11.24	67.19	0.21	9	2Fe-2S ferredoxin, GntR family transcriptional regulator, HAD family hydrolase, hypothetical protein, RNA-binding protein S4, RND transporter, transcriptional regulator TetR and transglycosylase

Appendix T 2. Genome features of *Nesterenkonia* sp. AN1 involved in adaptation to different stresses. The *Nesterenkonia* sp. AN1 proteins (CDSs) are indicated as are the nearest BLAST hits (locus tag,% identity, e-value and bitscore) as well as the COG classification for each protein.

Adaptation	Locus tag	Product description	COG	Subject id	% AAI	E-value	score
Cold stress	EXF25635B.1	putative ATP-dependent RNA helicase	COG0513	WP_022871899.1	80	0.00E+00	683
	EXF26063.1	Cold shock protein CspC	COG1278	WP_022873609.1	78	1.50E-66	208
	EXF26073.1	Cold shock protein CspA	COG1278	WP_022873622.1	93	1.90E-38	132
	EXF26183.1	Translation initiation factor 1	COG0361	YP_002957735.1	92	2.20E-42	144
	EXF26122.1	DEAD-box ATP-dependent RNA helicase CshA (EC 3.6.4.13)	COG0513	WP_022873322.1	73	0.00E+00	883
	EXF25810.1	Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8)	COG1185	WP_022872001.1	86	0.00E+00	1277
	EXF25817.1	Ribosome-binding factor A	COG0858	WP_022872007.1	77	9.40E-68	214
	EXF25818.1	Translation initiation factor 2	COG0532	WP_022872008.1	72	0.00E+00	1185
	EXF25820.1	Transcription termination protein NusA	COG0195	YP_002487504.1	79	0.00E+00	520
	EXF25456.1	Chaperone protein DnaJ	COG0484	WP_022872175.1	72	7.90E-147	428
	EXF25457.1	Heat shock protein GrpE	COG0576	WP_022872176.1	50	4.00E-70	228
	EXF25458.1	Chaperone protein DnaK	COG0443	WP_022872177.1	88	0.00E+00	1073
	EXF25246.1	Capsule biosynthesis protein CapA	COG2843	WP_022871770.1	59	5.90E-157	465
	EXF24979.1	ATP-dependent RNA helicase	COG0513	WP_009773177.1	75	0.00E+00	595
	EXF25027.1	RNA methyltransferase, TrmA family	COG2265	WP_022872524.1	54	4.50E-170	501

	EXF24737.1	Carbon starvation protein A	COG1966	WP_022873111.1	80	0.00E+00	872
	EXF24631.1	Chaperone protein DnaJ	COG0484	WP_022872818.1	79	0.00E+00	567
	EXF24052.1	RNA methyltransferase *rlmB* (EC 2.1.1.-)	COG0566	WP_022872616.1	67	5.70E-156	452
	EXF25204C.1	ATP-dependent helicase HrpA	COG1643	YP_002956868.1	54	0.00E+00	1383
	EXF25884B.1	Uncharacterized tRNA/rRNA methyltransferase YsgA	COG0566	WP_022871879.1	51	6.10E-95	297
DNA repair	EXF24550B.1	DNA repair protein RecN	COG0497	WP_022872894.1	66	0.00E+00	590
	EXF26040.1	Uracil-DNA glycosylase, family 1	COG0692	WP_022873601.1	69	3.20E-106	319
	EXF25888.1	RecA/Rad A recombinase	actNOG09053	WP_022871877.1	81	1.00E-60	193
	EXF25385.1	Recombination protein RecR	COG0353	YP_946469.1	80	2.70E-114	335
	EXF25436.1	Excinuclease ABC subunit A paralog of unknown function	COG0178	YP_005331432.1	72	0.00E+00	1194
	EXF25537.1	Exodeoxyribonuclease III (EC 3.1.11.2)	COG0708	WP_007272932.1	78	1.10E-150	433
	EXF25581.1	A/G-specific adenine glycosylase	COG1194	WP_022873346.1	65	1.10E-135	401
	EXF25334.1	Excinuclease ABC subunit C	COG0322	WP_022872239.1	78	0.00E+00	1058
	EXF25348.1	Excinuclease ABC subunit B	COG0556	WP_022872248.1	86	0.00E+00	1273
	EXF25336.1	Excinuclease ABC subunit A	COG0178	WP_022872241.1	82	0.00E+00	1673
	EXF25289.1	Formamido pyrimidine-DNA glycosylase (EC 3.2.2.23)	COG0266	WP_022872658.1	60	2.40E-127	379
	EXF25253.1	RecA protein	COG0468	WP_022872390.1	91	0.00E+00	641

	EXF25252.1	Regulatory protein RecX	COG2137	WP_022872389.1	55	6.20E-58	198
	EXF25077.1	Holliday junction DNA helicase RuvA	COG0632	WP_022872562.1	63	1.00E-80	250
	EXF25076.1	Holliday junction DNA helicase RuvB	COG2255	WP_022872563.1	85	0.00E+00	534
	EXF25078.1	Crossover junction endodeoxyribonuclease RuvC (EC 3.1.22.4)	COG0817	WP_022872561.1	75	3.50E-104	312
	EXF24638.1	DNA recombination and repair protein RecO	COG1381	WP_022872825.1	70	2.20E-112	334
	EXF24705.1	Methylated-DNA-protein-cysteine methyltransferase (EC 2.1.1.63)	COG2169	WP_022872688.1	63	0.00E+00	656
	EXF23958.1	Endonuclease III (EC 4.2.99.18)	COG0177	WP_022873689.1	75	1.80E-143	415
Membrane fluidity	EXF24605B.1	Phytoene dehydrogenase and related proteins	COG1233	WP_022872919.1	50	5.50E-135	416
	EXF26105B.1	Fatty acid desaturase (EC 1.14.19.3)	COG3239	WP_022873098.1	68	0.00E+00	520
	EXF26111.1	Octaprenyl diphosphate synthase (EC 2.5.1.90)	COG0142	WP_022873338.1	71	5.60E-155	451
	EXF25790.1	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	COG1028	YP_001221884.1	68	8.80E-113	336
	EXF25789.1	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	COG1028	WP_006369012.1	62	7.10E-98	298
	EXF25423.1	Lycopene elongase (EC 2.5.1.-)	COG0382	YP_946134.1	65	3.30E-125	371
	EXF25424.1	C50 carotenoid epsilon cyclase	actNOG12951	WP_020014083.1	55	6.70E-17	80

EXF25425.1	C50 carotenoid epsilon cyclase	actNOG13176	WP_017201376.1	57	1.80E-33	123
EXF25426.1	Phytoene dehydrogenase (EC 1.14.99.-)	COG1233	WP_021200000.1	62	0.00E+00	656
EXF25427.1	Phytoene synthase (EC 2.5.1.32)	COG1562	WP_017201374.1	62	7.10E-111	334
EXF25428.1	Isopentenyl-diphosphate delta-isomerase (EC 5.3.3.2)	COG1443	WP_017201372.1	64	3.20E-68	216
EXF25510.1	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	COG1028	WP_020674122.1	74	8.90E-134	391
EXF25465.1	short-chain dehydrogenase/reductase SDR	COG0300	WP_022872188.1	66	2.90E-127	378
EXF25335.1	1-acyl-sn-glycerol-3-phosphate acyltransferase (EC 2.3.1.51)	COG0204	WP_022872240.1	65	9.40E-106	317
EXF25301.1	1-acyl-sn-glycerol-3-phosphate acyltransferase (EC 2.3.1.51)	COG0204	WP_022873211.1	58	2.00E-103	316
EXF25197.1	Geranylgeranyl diphosphate synthase (EC 2.5.1.29)	COG0142	WP_022872329.1	61	9.20E-140	414
EXF25128B.1	3-Ketoacyl-(Acyl-Carrier-Protein) Reductase	actNOG00499	WP_022872460.1	59	8.70E-51	178
EXF24852.1	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	COG1028	YP_004898815.1	57	8.80E-109	330
EXF24624.1	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	COG1028	YP_002487956.1	76	1.50E-119	360

	EXF24397.1	3-oxoacyl-[ACP] synthase III in alkane synthesis cluster	COG0332	WP_022872515.1	82	0.00E+00	586
	EXF24403.1	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	COG1028	WP_022872506.1	86	4.40E-142	409
	EXF24426.1	3-oxoacyl-[acyl-carrier-protein] synthase, KASIII (EC 2.3.1.180)	COG0332	WP_022872278.1	75	0.00E+00	547
	EXF24428.1	3-oxoacyl-[acyl-carrier-protein] synthase, KASII (EC 2.3.1.179)	COG0304	WP_022872276.1	77	0.00E+00	678
	EXF24451.1	Short-chain dehydrogenase/reductase in hypothetical Actinobacterial gene cluster	COG1028	YP_001853950.1	88	3.00E-164	467
	EXF24309.1	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	COG1028	WP_007269559.1	76	1.40E-166	477
	EXF24133.1	Lycopene elongase (EC 2.5.1.-)	COG0382	WP_021809160.1	67	3.00E-118	351
	EXF24134.1	Phytoene dehydrogenase (EC 1.14.99.-)	COG1233	YP_007684695.1	59	5.50E-171	502
	EXF24005.1	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	COG4221	WP_022886657.1	74	5.90E-124	366
	EXF25202B.1	1-acyl-sn-glycerol-3-phosphate acyltransferase (EC 2.3.1.51)	COG0204	WP_022872335.1	71	5.30E-114	338
Oxidative stress	EXF24169C.1	4-hydroxyphenylpyruvate	COG3185	WP_003800060.1	67	1.20E-19	91

	dioxygenase (EC 1.13.11.27)					
EXF24618B.1	Protocatechuate 4,5-dioxygenase beta chain (EC 1.13.11.8)	COG3384	WP_022873379.1	87	0.00E+00	575
EXF26051.1	Protocatechuate 3,4-dioxygenase beta chain (EC 1.13.11.3)	COG3485	YP_003915644.1	80	1.00E-152	439
EXF26052.1	Protocatechuate 3,4-dioxygenase alpha chain (EC 1.13.11.3)	COG3485	WP_006213520.1	72	4.90E-89	270
EXF26118.1	Thiosulfate sulfurtransferase, rhodanese (EC 2.8.1.1)	COG3118	WP_022873326.1	81	1.50E-61	196
EXF25992.1	Lactoyl glutathione lyase (EC 4.4.1.5)	COG0346	WP_018161746.1	63	1.30E-55	181
EXF25775.1	transcriptional regulator, Crp/Fnr family	COG0664	WP_021809151.1	67	9.90E-98	297
EXF25808.1	Hydroxyacyl glutathione hydrolase (EC 3.1.2.6)	COG0491	WP_006237000.1	74	0.00E+00	726
EXF25878.1	Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)	COG2032	WP_022872077.1	46	3.30E-49	171
EXF25651B.1	NAD-dependent protein deacetylase of SIR2 family	COG0846	WP_022871887.1	71	1.10E-117	351
EXF25463.1	Glutathione S-transferase, omega (EC 2.5.1.18)	COG0435	WP_022872186.1	84	0.00E+00	650
EXF25468.1	Ferredoxin, 2Fe-2S	COG2146	WP_022872199.1	64	4.10E-44	151
EXF25331.1	NAD-dependent glyceraldehyde-3-	COG0057	WP_022872233.1	78	0.00E+00	546

	phosphate dehydrogenase (EC 1.2.1.12)					
EXF25332.1	Superoxide dismutase [Mn] (EC 1.15.1.1)	COG0605	WP_022872235.1	87	2.00E-133	384
EXF25206.1	Zinc uptake regulation protein ZUR	COG0735	WP_022872338.1	80	1.50E-74	230
EXF25073.1	Putative Holliday junction resolvase YggF	COG0816	WP_022870710.1	53	4.50E-43	152
EXF25096.1	NADPH-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.13)	COG0057	WP_022872547.1	77	0.00E+00	771
EXF24928.1	Catalase (EC 1.11.1.6)	COG0753	WP_022872599.1	83	0.00E+00	980
EXF24944.1	Glutaredoxin-like protein NrdH, required for reduction of Ribonucleotide reductase class Ib	COG0695	WP_022872882.1	91	1.60E-44	149
EXF25040.1	Gamma-glutamyl transpeptidase (EC 2.3.2.2)	COG0405	WP_022869606.1	50	0.00E+00	535
EXF24994.1	Lactoyl glutathione lyase and related lyases	COG0346	WP_007624079.1	71	3.60E-64	203
EXF25008.1	Peptide methionine sulfoxide reductase MsrA (EC 1.8.4.11)	COG0225	WP_022872543.1	73	7.90E-73	228
EXF24750.1	Ferroxidase (EC 1.16.3.1)	COG0783	WP_022873050.1	76	4.10E-88	267
EXF24778.1	Vanillate O-demethylase oxidoreductase (EC 1.14.13.-)	COG1018	WP_005632093.1	56	1.30E-108	330
EXF24832.1	Thioredoxin	COG3118	WP_022872705.1	89	5.10E-63	198

EXF24855.1	FIG00733190: hypothetical protein	COG1741	WP_022872747.1	73	0.00E+00	578
EXF24885.1	Glyoxalase family protein	COG0346	YP_007463503.1	75	0.00E+00	519
EXF24854.1	Organic hydroperoxide resistance protein	COG1764	WP_008376862.1	67	2.90E-63	201
EXF24632.1	Ribosomal RNA small subunit methyltransferase E (EC 2.1.1.-)	COG1385	WP_022872819.1	55	1.60E-96	298
EXF24421.1	Alkyl hydroperoxide reductase subunit C-like protein	COG0450	WP_022872273.1	46	6.70E-37	136
EXF24480.1	Thiol peroxidase, Bcp-type (EC 1.11.1.15)	COG1225	WP_022873173.1	78	1.10E-83	254
EXF24492.1	Thioredoxin reductase (EC 1.8.1.9)	COG0492	WP_022872938.1	60	1.10E-123	369
EXF24526.1	Gamma-glutamyl transpeptidase (EC 2.3.2.2)	COG0405	WP_022872604.1	72	0.00E+00	840
EXF24488.1	Rubredoxin	actNOG20548	WP_022873186.1	63	1.60E-82	258
EXF24490.1	Nicotinate phosphoribosyl transferase (EC 2.4.2.11)	COG1488	WP_022873188.1	73	0.00E+00	616
EXF24295.1	Transcriptional regulator, FUR family	COG0735	WP_022871973.1	76	1.70E-63	203
EXF24299.1	Vanillate O-demethylase oxidoreductase (EC 1.14.13.-)	COG1018	WP_003937168.1	62	1.10E-129	386
EXF24102.1	3-phenylpropionate dioxygenase ferredoxin subunit	COG2146	WP_005268380.1	72	4.30E-38	134
EXF24216.1	Ferredoxin reductase	COG0446	WP_020515589.1	37	3.40E-60	211
EXF24104.1	Catalase (EC 1.11.1.6)	COG0753	YP_004240449.1	68	0.00E+00	1003

	EXF24103.1	Phenylpropionate dioxygenase, large terminal subunit	COG4638	YP_003486449.1	61	1.70E-171	497
	EXF24169.1	Transcriptional regulator, IclR family	COG1414	YP_003508459.1	51	6.10E-70	228
	EXF23969.1	NAD-dependent protein deacetylase of SIR2 family	COG0846	WP_022872910.1	67	1.10E-157	456
	EXF23844.1	transcriptional regulator, Crp/Fnr family	COG0664	WP_021809151.1	63	1.90E-98	300
	EXF23832.1	transcriptional regulator, Crp/Fnr family	COG0664	WP_021809151.1	61	1.80E-97	298
	EXF24492B.1	Nicotinamidase (EC 3.5.1.19)	COG1335	WP_022873189.1	67	9.40E-90	273
	EXF24966B.1	Catechol 2,3-dioxygenase (EC 1.13.11.2)	COG0346	YP_008478365.1	86	0.00E+00	644
Osmoprotection	EXF25894.1	Signal transduction histidine-protein kinase/phosphatase mprB	COG0642	WP_005267484.1	50	1.20E-135	413
	EXF25716B.1	Choline dehydrogenase (EC 1.1.99.1)	COG2303	YP_949295.1	67	0.00E+00	711
	EXF25603B.1	High-affinity choline uptake protein BetT	COG1292	YP_003155648.1	69	0.00E+00	928
	EXF25160B.1	Glutamate synthase [NADPH] large chain (EC 1.4.1.13)	COG0069	WP_022873300.1	81	0.00E+00	2579
	EXF25967B.1	Choline-sulfatase (EC 3.1.6.6)	COG3119	WP_022870010.1	72	0.00E+00	773
	EXF25929C.1	Glycine betaine transporter OpuD	COG1292	YP_002486171.1	65	0.00E+00	680
	EXF23940B.1	L-proline glycine betaine ABC transport system	COG1125	YP_007294410.1	67	8.30E-153	446

	permease protein ProV (TC 3.A.1.12.1)					
EXF26032.1	Trehalose-6-phosphate phosphatase (EC 3.1.3.12)	COG1877	WP_022873595.1	62	1.70E-97	298
EXF26033.1	Alpha,alpha-trehalose-phosphate synthase [UDP-forming] (EC 2.4.1.15)	COG0380	WP_022873596.1	75	0.00E+00	772
EXF26244.1	L-proline glycine betaine ABC transport system permease protein ProW (TC 3.A.1.12.1)	COG2113	YP_005329610.1	62	3.10E-118	354
EXF26245.1	L-proline glycine betaine ABC transport system permease protein ProW (TC 3.A.1.12.1)	COG4176	WP_010550407.1	57	1.20E-103	317
EXF26059.1	Osmo-sensitive K+ channel histidine kinase KdpD (EC 2.7.3.-)	COG0642	WP_022873606.1	64	0.00E+00	629
EXF26246.1	L-proline glycine betaine ABC transport system permease protein ProV (TC 3.A.1.12.1)	COG4175	WP_022871921.1	70	0.00E+00	571
EXF25980.1	Na+/H+ antiporter	COG0025	WP_020518412.1	59	5.90E-122	371
EXF25916.1	Ktr system potassium uptake protein C	COG0569	WP_009773753.1	57	5.10E-86	266
EXF25917.1	Ktr system potassium uptake protein B+C108	COG0168	WP_019180102.1	66	0.00E+00	528
EXF25872.1	Na(+) H(+) antiporter subunit G	COG1320	WP_022872071.1	79	1.70E-51	170

EXF25873.1	Na(+) H(+) antiporter subunit F	COG2212	WP_022872072.1	81	2.90E-37	131
EXF25875.1	Na(+) H(+) antiporter subunit D	COG0651	WP_022872074.1	84	0.00E+00	845
EXF25874.1	Na(+) H(+) antiporter subunit E	COG1863	WP_022872073.1	66	4.60E-55	180
EXF25876.1	Na(+) H(+) antiporter subunit C	COG1006	WP_022872075.1	85	1.10E-56	188
EXF25877.1	Na(+) H(+) antiporter subunit A; Na(+) H(+) antiporter subunit B	COG1009	WP_022872076.1	74	0.00E+00	1322
EXF25631.1	L-proline glycine betaine ABC transport system permease protein ProV (TC 3.A.1.12.1)	COG1125	WP_007078894.1	68	4.20E-126	372
EXF25633.1	L-proline glycine betaine binding ABC transporter protein ProX (TC 3.A.1.12.1)	COG1732	WP_010533004.1	60	9.80E-107	325
EXF25718.1	Phosphoglycerol transferase I (EC 2.7.8.20)	COG1368	WP_019310579.1	57	0.00E+00	581
EXF25632.1	Glycine betaine ABC transport system permease protein	COG1174	WP_010533005.1	58	1.80E-78	247
EXF25478.1	Na(+) H(+) antiporter subunit G	COG1320	WP_022872212.1	83	5.70E-54	176
EXF25482.1	Na(+) H(+) antiporter subunit C	COG1006	WP_022872216.1	72	2.70E-63	205
EXF25563.1	FIG152265: Sodium:solute symporter associated protein	COG4327	WP_022873414.1	70	1.10E-47	159

EXF25455.1	HspR, transcriptional repressor of DnaK operon	COG0789	WP_022872174.1	71	1.50E-54	180
EXF25479.1	Na(+) H(+) antiporter subunit F	COG2212	WP_022872213.1	52	1.40E-45	158
EXF25480.1	Na(+) H(+) antiporter subunit E	COG1863	WP_022872214.1	72	1.90E-107	320
EXF25481.1	Na(+) H(+) antiporter subunit D	COG0651	WP_022872215.1	85	0.00E+00	852
EXF25556.1	High-affinity choline uptake protein BetT	COG1292	WP_022873419.1	66	0.00E+00	908
EXF25351.1	High-affinity choline uptake protein BetT	COG1292	WP_022872250.1	66	0.00E+00	821
EXF25160.1	Glutamate synthase [NADPH] small chain (EC 1.4.1.13)	COG0493	WP_022873301.1	83	0.00E+00	847
EXF24759.1	L-2,4-diaminobutyric acid acetyltransferase (EC 2.3.1.-)	COG0456	WP_010838282.1	56	9.90E-54	179
EXF24761.1	L-ectoine synthase (EC 4.2.1.-)	bactNOG36145	WP_022873042.1	85	2.00E-77	236
EXF24783.1	Na(+) H(+) antiporter subunit G	COG1320	YP_008372410.1	59	2.30E-26	104
EXF24760.1	Diaminobutyrate-pyruvate aminotransferase (EC 2.6.1.46)	COG0160	WP_005271380.1	72	0.00E+00	639
EXF24460.1	sodium-solute symporter, putative	COG0591	YP_003917178.1	81	0.00E+00	871
EXF24293.1	Trehalose synthase (EC 5.4.99.16)	COG0366	WP_022871968.1	75	0.00E+00	904
EXF24249.1	Organic hydroperoxide resistance protein	COG1764	YP_001360611.1	76	2.00E-74	230

EXF24139.1	Betaine aldehyde dehydrogenase (EC 1.2.1.8)	COG1012	YP_003155642.1	75	0.00E+00	758
EXF24140.1	Choline dehydrogenase (EC 1.1.99.1)	COG2303	WP_018155108.1	75	0.00E+00	801
EXF24227.1	L-proline glycine betaine binding ABC transporter protein ProX (TC 3.A.1.12.1)	COG1732	WP_022873555.1	38	1.60E-43	161
EXF24163.1	Osmotically inducible protein Y precursor	COG2823	WP_022886157.1	52	3.20E-66	215
EXF24123.1	Osmo-sensitive K+ channel histidine kinase KdpD (EC 2.7.3.-)	COG0642	YP_007466195.1	66	1.20E-163	475
EXF24138.1	High-affinity choline uptake protein BetT	COG1292	YP_003155641.1	60	0.00E+00	841
EXF24045.1	Ktr system potassium uptake protein B	COG0168	WP_022873041.1	72	0.00E+00	619
EXF24046.1	Ktr system potassium uptake protein C	COG0569	WP_022873040.1	84	1.70E-135	391
EXF23929.1	Glutamate synthase [NADPH] large chain (EC 1.4.1.13)	COG5279	WP_019618514.1	47	1.20E-36	133
EXF23938.1	L-proline glycine betaine binding ABC transporter protein ProX (TC 3.A.1.12.1)	COG1732	YP_702555.1	55	7.30E-103	315
EXF23939.1	Glycine betaine ABC transport system permease protein	COG1174	YP_007294408.1	64	9.50E-82	255
EXF24009.1	Osmo-sensitive K+ channel histidine kinase KdpD (EC 2.7.3.-)	COG0642	WP_021010967.1	52	1.40E-96	304

	EXF23940.1	L-proline glycine betaine ABC transport system permease protein ProW (TC 3.A.1.12.1)	COG1174	YP_007294409.1	65	3.10E-84	262
	EXF25631B.1	L-proline glycine betaine ABC transport system permease protein ProW (TC 3.A.1.12.1)	COG1174	WP_022893326.1	56	3.60E-65	211
	EXF25443B.1	Proline dehydrogenase (EC 1.5.99.8) (Proline oxidase)	COG1012	WP_022872163.1	71	0.00E+00	1627
	EXF25482B.1	Na(+) H(+) antiporter subunit A; Na(+) H(+) antiporter subunit B	COG1009	WP_022872217.1	71	0.00E+00	1382
General stress	EXF25955.1	Universal stress protein	COG0589	WP_005881911.1	33	8.30E-29	123
	EXF25897.1	Serine phosphatase RsbU, regulator of sigma subunit	COG3920	WP_022871874.1	74	0.00E+00	785
	EXF25307.1	RNA polymerase sigma-70 factor, ECF subfamily	COG1595	WP_019182024.1	50	7.70E-44	154
	EXF25088.1	Hemoglobin-like protein HbO	COG2346	WP_022872554.1	63	1.40E-85	263
	EXF24956.1	Universal stress protein family	COG0589	WP_022873438.1	62	1.20E-121	364
	EXF25045.1	Universal stress protein family	COG0589	WP_022871460.1	41	5.30E-45	166
	EXF24980.1	Serine phosphatase RsbU, regulator of sigma subunit	COG2208	WP_021201695.1	59	5.20E-158	461
	EXF24981.1	Anti-sigma F factor antagonist (spoIIAA-2); Anti-sigma B factor antagonist RsbV	COG1366	NP_737814.1	62	7.20E-39	136
	EXF24982.1	Serine-protein kinase RsbW (EC 2.7.11.1)	COG2172	WP_021201692.1	46	9.20E-26	105

	EXF24860.1	Universal stress protein	COG0589	YP_002489855.1	56	3.20E-40	142
	EXF24277.1	Universal stress protein family	COG0589	WP_022873438.1	54	8.80E-97	302
	EXF24206.1	universal stress protein family	COG0589	YP_001853940.1	40	5.60E-59	201
	EXF24043.1	Serine phosphatase RsbU, regulator of sigma subunit	COG2208	WP_021201695.1	53	5.40E-143	423
	EXF23991.1	Universal stress protein family	COG0589	WP_009376415.1	39	1.80E-47	174

Appendix T 3. COG functional classification of positively selected core gene families in *Nesterenkonia* sp. AN1.

Gene families	LRT	P	BH	COG	Description	Class
***Nesterenkonia* sp. AN1**						
Nest_832	12.28375	0.000457	0.01012	*	*	*
Nest_644	12.74046	0.000358	0.008416	COG1249	Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes	C
Nest_960	14.87192	0.000115	0.003696	COG2609	Pyruvate dehydrogenase complex, dehydrogenase (E1) component	C
Nest_1064	8.181586	0.004232	0.037956	COG2838	Monomeric isocitrate dehydrogenase	C
Nest_1083	7.22822	0.007177	0.053312	COG0479	Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit	C
Nest_1084	11.28614	0.000781	0.013324	COG1053	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	C
Nest_27	8.474932	0.003601	0.03498	COG0045	Succinyl-CoA synthetase, beta subunit	C
Nest_395	9.54011	0.00201	0.023354	COG1009	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na+/H+ antiporter, MnhA subunit	C
Nest_498	11.14211	0.000844	0.013371	COG0421	Spermidine synthase	E

Nest_507	10.39296	0.001265	0.016986	COG0065	3-isopropylmalate dehydratase large subunit	E
Nest_518	8.606614	0.003349	0.033251	COG0129	Dihydroxyacid dehydratase/phosphogluconate dehydratase	E
Nest_527	7.791556	0.005249	0.045856	COG0134	Indole-3-glycerol phosphate synthase	E
Nest_578	13.47077	0.000242	0.006454	COG0128	5-enolpyruvylshikimate-3-phosphate synthase	E
Nest_615	8.46538	0.00362	0.03498	COG0624	Acetylornithine deacetylase/Succinyl-diaminopimelatedesuccinylase and related deacylases	E
Nest_630	21.88803	2.89E-06	0.000351	COG0308	Aminopeptidase N	E
Nest_658	10.50883	0.001188	0.016692	COG1063	Threonine dehydrogenase and related Zn-dependent dehydrogenases	E
Nest_766	14.23283	0.000162	0.004523	COG0160	4-aminobutyrate aminotransferase and related aminotransferases	E
Nest_1036	9.273824	0.002325	0.026258	COG0498	Threonine synthase	E
Nest_1038	10.72447	0.001057	0.016035	COG0019	Diaminopimelate decarboxylase	E
Nest_1187	11.13997	0.000845	0.013371	COG0137	Argininosuccinate synthase	E
Nest_1195	8.934758	0.002798	0.028555	COG0548	Acetylglutamate kinase	E
Nest_1206	10.40725	0.001255	0.016986	COG4448	L-asparaginase II	E
Nest_1216	8.341226	0.003876	0.036112	COG0757	3-dehydroquinate dehydratase II	E
Nest_203	20.38206	6.34E-06	0.000479	COG1104	Cysteine sulfinatedesulfinase/cysteine desulfurase and related enzymes	E
Nest_285	14.38154	0.000149	0.004405	COG4166	ABC-type oligopeptide transport system, periplasmic component	E
Nest_298	7.673338	0.005604	0.046363	COG2902	NAD-specific glutamate dehydrogenase	E
Nest_314	40.24974	2.23E-10	1.22E-07	COG1505	Serine proteases of the peptidase family S9A	E
Nest_341	10.17188	0.001426	0.017899	COG0527	Aspartokinases	E
Nest_1057	9.153276	0.002483	0.027111	COG2812	DNA polymerase III, gamma/tau subunits	F
Nest_1203	17.5522	2.8E-05	0.001387	COG0150	Phosphoribosylaminoimidazole (AIR) synthetase	F
Nest_453	7.411122	0.006482	0.050561	COG0021	Transketolase	G

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_455	21.06122	4.45E-06	0.000479	COG0166	Glucose-6-phosphate isomerase	G
Nest_73	12.71753	0.000362	0.008416	COG0205	6-phosphofructokinase	G
Nest_757	11.81657	0.000587	0.011053	COG1070	Sugar (pentulose and hexulose) kinases	G
Nest_838	18.14771	2.04E-05	0.001142	COG1929	Glycerate kinase	G
Nest_840	14.25844	0.000159	0.004523	COG3386	Gluconolactonase	G
Nest_924	16.748	4.27E-05	0.001793	COG1554	Trehalose and maltose hydrolases (possible phosphorylases)	G
Nest_125	16.03564	6.22E-05	0.002263	COG0063	Predicted sugar kinase	G
Nest_132	11.49127	0.000699	0.012726	COG1472	Beta-glucosidase-related glycosidases	G
Nest_1113	7.683438	0.005573	0.046363	COG1494	Fructose-1,6-bisphosphatase/sedoheptulose 1,7-bisphosphatase and related proteins	G
Nest_222	14.651	0.000129	0.003924	COG1109	Phosphomannomutase	G
Nest_780	9.027928	0.002659	0.027918	COG1653	ABC-type sugar transport system, periplasmic component	G
Nest_49	13.15986	0.000286	0.007263	COG1169	Isochorismate synthase	H
Nest_50	10.69863	0.001072	0.016038	COG2226	Methylase involved in ubiquinone/menaquinone biosynthesis	H
Nest_482	9.744466	0.001799	0.021823	COG0801	7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase	H
Nest_534	7.645668	0.005691	0.046376	COG0054	Riboflavin synthase beta-chain	H
Nest_614	17.24145	3.29E-05	0.001498	COG0095	Lipoate-protein ligase A	H
Nest_646	10.60206	0.00113	0.016231	COG0321	Lipoate-protein ligase B	H
Nest_647	11.49969	0.000696	0.012726	COG0320	Lipoate synthase	H
Nest_706	9.040996	0.00264	0.027918	COG0351	Hydroxymethylpyrimidine/phosphomethylpyrimidine kinase	H
Nest_853	8.17183	0.004255	0.037956	COG0635	Coproporphyrinogen III oxidase and related Fe-S oxidoreductases	H
Nest_895	9.308086	0.002281	0.026225	COG0192	S-adenosylmethioninesynthetase	H
Nest_988	20.50753	5.94E-06	0.000479	COG0285	Folylpolyglutamate synthase	H
Nest_1013	12.67232	0.000371	0.008443	COG1488	Nicotinic acid phosphoribosyltransferase	H

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_1171	7.740358	0.0054	0.045943	COG0499	S-adenosylhomocysteine hydrolase	H
Nest_632	20.31367	6.57E-06	0.000479	COG0183	Acetyl-CoA acetyltransferase	I
Nest_845	9.71209	0.001831	0.021967	COG1960	Acyl-CoA dehydrogenases	I
Nest_141	8.513214	0.003526	0.034686	COG4799	Acetyl-CoA carboxylase, carboxyltransferase component (subunits alpha and beta)	I
Nest_1133	7.450516	0.006342	0.050183	COG0761	Penicillin tolerance protein	I
Nest_1158	8.163042	0.004275	0.037956	COG1022	Long-chain acyl-CoA synthetases (AMP-forming)	I
Nest_1217	11.89705	0.000562	0.010963	COG0365	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	I
Nest_565	7.248796	0.007095	0.053312	COG1022	Long-chain acyl-CoA synthetases (AMP-forming)	I
Nest_538	7.613742	0.005793	0.04679	COG0223	Methionyl-tRNAformyltransferase	J
Nest_622	20.51185	5.93E-06	0.000479	COG2519	tRNA(1-methyladenosine) methyltransferase and related methyltransferases	J
Nest_77	7.350566	0.006704	0.051195	COG0081	Ribosomal protein L1	J
Nest_737	7.646844	0.005687	0.046376	COG0349	Ribonuclease D	J
Nest_91	8.268756	0.004033	0.036703	COG0088	Ribosomal protein L4	J
Nest_828	8.060388	0.004524	0.039843	COG0359	Ribosomal protein L9	J
Nest_871	7.744014	0.005389	0.045943	COG0242	N-formylmethionyl-tRNAdeformylase	J
Nest_1009	17.24636	3.28E-05	0.001498	COG0689	RNase PH	J
Nest_1056	9.15621	0.002479	0.027111	COG1189	Predicted rRNAmethylase	J
Nest_252	11.22996	0.000805	0.013371	COG0442	Prolyl-tRNAsynthetase	J
Nest_294	10.2891	0.001338	0.017396	COG0290	Translation initiation factor 3 (IF-3)	J
Nest_321	11.36866	0.000747	0.013156	COG0423	Glycyl-tRNAsynthetase (class II)	J
Nest_84	8.990538	0.002714	0.027965	COG0085	DNA-directed RNA polymerase, beta subunit/140 kD subunit	K
Nest_1016	10.47817	0.001208	0.016697	COG1061	DNA or RNA helicases of superfamily II	K

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_1116	7.709718	0.005492	0.046136	COG1316	Transcriptional regulator	K
Nest_1227	16.64774	4.5E-05	0.00182	COG0846	NAD-dependent protein deacetylases, SIR2 family	K
Nest_303	9.067462	0.002602	0.027918	COG1309	Transcriptional regulator	K
Nest_383	11.9524	0.000546	0.010836	COG1309	Transcriptional regulator	K
Nest_469	7.602922	0.005827	0.04679	COG4581	Superfamily II RNA helicase	L
Nest_486	15.6311	7.7E-05	0.002712	COG0749	DNA polymerase I - 3'-5' exonuclease and polymerase domains	L
Nest_542	10.20954	0.001397	0.017899	COG0587	DNA polymerase III, alpha subunit	L
Nest_604	7.427378	0.006424	0.050467	COG0468	RecA/RadA recombinase	L
Nest_638	10.00769	0.001559	0.019127	COG0187	Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit	L
Nest_712	10.7504	0.001043	0.016035	COG0513	Superfamily II DNA and RNA helicases	L
Nest_863	9.057358	0.002616	0.027918	COG1381	Recombinational DNA repair protein (RecF pathway)	L
Nest_932	8.684244	0.00321	0.032156	COG0550	Topoisomerase IA	L
Nest_971	12.05849	0.000516	0.010623	COG0164	Ribonuclease HII	L
Nest_301	44.86406	2.11E-11	2.31E-08	COG0513	Superfamily II DNA and RNA helicases	L
Nest_551	9.585414	0.001961	0.02328	COG0773	UDP-N-acetylmuramate-alanine ligase	M
Nest_796	16.0515	6.16E-05	0.002263	COG3764	Sortase (surface protein transpeptidase)	M
Nest_1153	8.439636	0.003671	0.035167	COG1207	N-acetylglucosamine-1-phosphate uridyltransferase (contains nucleotidyltransferase and I-patch acetyltransferase domains)	M
Nest_1213	11.14054	0.000845	0.013371	COG0744	Membrane carboxypeptidase (penicillin-binding protein)	M
Nest_447	19.88947	8.21E-06	0.00056	COG3590	Predicted metalloendopeptidase	O
Nest_585	7.739464	0.005403	0.045943	COG0691	tmRNA-binding protein	O
Nest_717	18.10379	2.09E-05	0.001142	COG1391	Glutamine synthetaseadenylyltransferase	O

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_954	14.73658	0.000124	0.003857	COG0719	ABC-type transport system involved in Fe-S cluster assembly, permease component	O
Nest_1105	16.86953	4E-05	0.001749	COG1331	Highly conserved protein containing a thioredoxin domain	O
Nest_1160	10.62441	0.001116	0.016231	COG1928	Dolichyl-phosphate-mannose--protein O-mannosyl transferase	O
Nest_37	10.73576	0.001051	0.016035	COG1333	ResB protein required for cytochrome c biosynthesis	O
Nest_376	12.9011	0.000328	0.00801	COG0435	Predicted glutathione S-transferase	O
Nest_414	19.25258	1.15E-05	0.000736	COG3221	ABC-type phosphate/phosphonate transport system, periplasmic component	P
Nest_437	25.50147	4.42E-07	9.65E-05	COG0573	ABC-type phosphate transport system, permease component	P
Nest_438	17.83354	2.41E-05	0.001254	COG0226	ABC-type phosphate transport system, periplasmic component	P
Nest_135	9.253544	0.00235	0.026258	COG4606	ABC-type enterochelin transport system, permease component	P
Nest_1060	7.73125	0.005427	0.045943	COG0704	Phosphate uptake regulator	P
Nest_392	7.397202	0.006533	0.050592	COG1863	Multisubunit Na+/H+ antiporter, MnhE subunit	P
Nest_404	8.362412	0.003831	0.036061	COG1116	ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component	P
Nest_260	8.418662	0.003714	0.035266	COG4147	Predicted symporter	R
Nest_450	12.24204	0.000467	0.01012	COG0730	Predicted permeases	R
Nest_579	8.328762	0.003902	0.036112	COG1162	Predicted GTPases	R
Nest_621	11.19978	0.000818	0.013371	COG0637	Predicted phosphatase/phosphohexomutase	R
Nest_818	22.72581	1.87E-06	0.000255	COG0728	Uncharacterized membrane protein, putative virulence factor	R
Nest_837	13.70476	0.000214	0.00584	COG5271	AAA ATPase containing von Willebrand factor type A (*vWA*) domain	R
Nest_929	13.30935	0.000264	0.006866	COG3393	Predicted acetyltransferase	R
Nest_1004	29.40008	5.89E-08	2.14E-05	COG0579	Predicted dehydrogenase	R
Nest_1088	12.15316	0.00049	0.010291	COG1473	Metal-dependent amidase/aminoacylase/carboxypeptidase	R

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_1185	10.35527	0.001291	0.016986	COG3973	Superfamily I DNA and RNA helicases	R
Nest_1214	14.93167	0.000111	0.003689	COG1408	Predicted phosphohydrolases	R
Nest_190	16.07006	6.1E-05	0.002263	COG3173	Predicted aminoglycoside phosphotransferase	R
Nest_207	10.61837	0.00112	0.016231	COG0714	MoxR-like ATPases	R
Nest_361	9.556806	0.001992	0.023354	COG0820	Predicted Fe-S-cluster redox enzyme	R
Nest_490	11.86175	0.000573	0.010977	COG1216	Predicted glycosyltransferases	R
Nest_625	25.76991	3.85E-07	9.65E-05	*	*	S
Nest_891	9.248828	0.002356	0.026258	*	*	S
Nest_1173	10.36909	0.001281	0.016986	*	*	S
Nest_479	12.89146	0.00033	0.00801	COG3428	Predicted membrane protein	S
Nest_669	7.470522	0.006272	0.049991	COG3463	Predicted membrane protein	S
Nest_705	8.842094	0.002944	0.029763	COG2357	Uncharacterized protein conserved in bacteria	S
Nest_748	11.36929	0.000747	0.013156	COG5282	Uncharacterized conserved protein	S
Nest_858	11.98588	0.000536	0.010836	COG1385	Uncharacterized protein conserved in bacteria	S
Nest_131	19.06439	1.26E-05	0.000767	COG2170	Uncharacterized conserved protein	S
Nest_1091	7.291728	0.006927	0.052532	COG0586	Uncharacterized membrane-associated protein	S
Nest_1103	12.22061	0.000473	0.01012	COG2120	Uncharacterized proteins, LmbE homologs	S
Nest_1169	10.18276	0.001418	0.017899	COG1300	Uncharacterized membrane protein	S
Nest_171	8.990024	0.002715	0.027965	COG0586	Uncharacterized membrane-associated protein	S
Nest_192	15.29882	9.18E-05	0.003132	COG5282	Uncharacterized conserved protein	S
Nest_194	24.34489	8.05E-07	0.000147	COG1615	Uncharacterized conserved protein	S
Nest_251	10.09049	0.00149	0.018494	COG0779	Uncharacterized protein conserved in bacteria	S
Nest_29	8.298636	0.003967	0.036408	COG1814	Uncharacterized membrane protein	S
Nest_1142	23.57389	1.2E-06	0.000188	COG4805	Uncharacterized protein conserved in bacteria	S

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_83	10.50222	0.001192	0.016692	COG2197	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	T
Nest_859	20.62628	5.58E-06	0.000479	COG1702	Phosphate starvation-inducible protein PhoH, predicted ATPase	T
Nest_215	7.370372	0.006631	0.050991	COG0653	Preprotein translocase subunit SecA (ATPase, RNA helicase)	U
Nest_823	7.238884	0.007134	0.053312	COG1131	ABC-type multidrug transport system, ATPase component	V
Nest_1215	11.28975	0.000779	0.013324	COG1132	ABC-type multidrug transport system, ATPase and permease components	V
***Nesterenkonia* sp. F**						
Nest_513	7.973176	0.004748	0.042999	COG0473	Isocitrate/isopropylmalate dehydrogenase	C
Nest_716	7.77324	0.005303	0.045108	COG2141	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	C
Nest_741	7.428596	0.00642	0.048453	COG1048	Aconitase A	C
Nest_749	7.227218	0.007181	0.05052	COG0221	Inorganic pyrophosphatase	C
Nest_770	9.757336	0.001786	0.022556	COG0584	Glycerophosphoryldiester phosphodiesterase	C
Nest_843	13.36102	0.000257	0.008016	COG2025	Electron transfer flavoprotein, alpha subunit	C
Nest_868	7.414138	0.006471	0.048453	COG1622	Heme/copper-type cytochrome/quinol oxidases, subunit 2	C
Nest_876	8.048444	0.004554	0.042507	COG1290	Cytochrome b subunit of the bc complex	C
Nest_913	10.15445	0.00144	0.019898	COG2141	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	C
Nest_1026	11.94902	0.000547	0.011852	COG0056	F0F1-type ATP synthase, alpha subunit	C
Nest_1231	14.81141	0.000119	0.004805	COG4988	ABC-type transport system involved in cytochrome bd biosynthesis, ATPase and permease components	C

Nest_331	12.72296	0.000361	0.009408	COG1071	Pyruvate/2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, eukaryotic type, alpha subunit	C
Nest_395	30.60712	3.16E-08	1.72E-05	COG1009	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na+/H+ antiporter, MnhA subunit	C
Nest_587	11.1819	0.000826	0.014717	COG2884	Predicted ATPase involved in cell division	D
Nest_813	13.1926	0.000281	0.008295	COG1192	ATPases involved in chromosome partitioning	D
Nest_140	7.805174	0.00521	0.045053	COG0424	Nucleotide-binding protein implicated in inhibition of septum formation	D
Nest_199	27.11215	1.92E-07	5.24E-05	COG1196	Chromosome segregation ATPases	D
Nest_683	10.93673	0.000943	0.015366	COG1063	Threonine dehydrogenase and related Zn-dependent dehydrogenases	E
Nest_52	8.824476	0.002972	0.032684	COG0493	NADPH-dependent glutamate synthase beta chain and related oxidoreductases	E
Nest_58	15.42759	8.57E-05	0.003901	COG0626	Cystathionine beta-lyases/cystathionine gamma-synthases	E
Nest_497	8.811692	0.002993	0.032684	COG0436	Aspartate/tyrosine/aromatic aminotransferase	E
Nest_238	8.982712	0.002725	0.031002	COG0028	Thiamine pyrophosphate-requiring enzymes [acetolactate synthase, pyruvate dehydrogenase (cytochrome), glyoxylatecarboligase, phosphonopyruvate decarboxylase]	E
Nest_518	7.374822	0.006614	0.048803	COG0129	Dihydroxyacid dehydratase/phosphogluconate dehydratase	E
Nest_526	7.607968	0.005811	0.04666	COG0133	Tryptophan synthase beta chain	E
Nest_527	12.8842	0.000331	0.009408	COG0134	Indole-3-glycerol phosphate synthase	E
Nest_531	11.92604	0.000554	0.011852	COG0107	Imidazoleglycerol-phosphate synthase	E
Nest_578	9.095044	0.002563	0.029461	COG0128	5-enolpyruvylshikimate-3-phosphate synthase	E
Nest_600	7.665708	0.005628	0.046013	COG0253	Diaminopimelate epimerase	E
Nest_692	8.315046	0.003932	0.03939	COG0014	Gamma-glutamyl phosphate reductase	E

Nest_718	23.17412	1.48E-06	0.000269	COG0174	Glutamine synthetase	E
Nest_775	12.19653	0.000479	0.010892	COG0665	Glycine/D-amino acid oxidases (deaminating)	E
Nest_1003	14.15463	0.000168	0.005746	COG0747	ABC-type dipeptide transport system, periplasmic component	E
Nest_1035	13.2924	0.000266	0.008083	COG0083	Homoserine kinase	E
Nest_1037	10.56842	0.00115	0.017693	COG0460	Homoserine dehydrogenase	E
Nest_707	22.06441	2.64E-06	0.000411	COG0405	Gamma-glutamyltransferase	E
Nest_1187	10.05806	0.001517	0.020252	COG0137	Argininosuccinate synthase	E
Nest_1195	7.255776	0.007067	0.050441	COG0548	Acetylglutamate kinase	E
Nest_1197	7.15041	0.007495	0.05052	COG0002	Acetylglutamatesemialdehyde dehydrogenase	E
Nest_1206	15.84916	6.86E-05	0.003567	COG4448	L-asparaginase II	E
Nest_165	16.36653	5.22E-05	0.003167	COG1003	Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain	E
Nest_279	7.36037	0.006668	0.048867	COG0436	Aspartate/tyrosine/aromatic aminotransferase	E
Nest_298	8.914562	0.002829	0.031849	COG2902	NAD-specific glutamate dehydrogenase	E
Nest_314	30.95931	2.63E-08	1.72E-05	COG1505	Serine proteases of the peptidase family S9A	E
Nest_380	7.413188	0.006475	0.048453	COG0141	Histidinol dehydrogenase	E
Nest_384	7.164594	0.007436	0.05052	COG0147	Anthranilate/para-aminobenzoate synthases component I	E
Nest_702	11.51154	0.000692	0.013487	COG0209	Ribonucleotide reductase, alpha subunit	F
Nest_1152	12.7926	0.000348	0.009408	COG0462	Phosphoribosylpyrophosphatesynthetase	F
Nest_1203	10.33866	0.001303	0.018968	COG0150	Phosphoribosylaminoimidazole (AIR) synthetase	F
Nest_318	8.109518	0.004403	0.041813	COG0232	dGTPtriphosphohydrolase	F
Nest_344	17.6792	2.61E-05	0.002133	COG0046	Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain	F
Nest_453	13.98958	0.000184	0.006083	COG0021	Transketolase	G

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_73	8.402232	0.003748	0.038607	COG0205	6-phosphofructokinase	G
Nest_730	8.233796	0.004112	0.040451	COG0362	6-phosphogluconate dehydrogenase	G
Nest_924	7.514646	0.00612	0.048427	COG1554	Trehalose and maltose hydrolases (possible phosphorylases)	G
Nest_985	7.896892	0.004952	0.043964	COG1134	ABC-type polysaccharide/polyol phosphate transport system, ATPase component	G
Nest_1144	9.104278	0.00255	0.029461	COG1109	Phosphomannomutase	G
Nest_50	12.52552	0.000401	0.009564	COG2226	Methylase involved in ubiquinone/menaquinone biosynthesis	H
Nest_700	14.96632	0.000109	0.004597	COG0095	Lipoate-protein ligase A	H
Nest_772	15.89157	6.71E-05	0.003567	COG0476	Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 2	H
Nest_853	15.60219	7.82E-05	0.003711	COG0635	Coproporphyrinogen III oxidase and related Fe-S oxidoreductases	H
Nest_1070	7.97379	0.004746	0.042999	COG0190	5,10-methylene-tetrahydrofolate dehydrogenase/Methenyl tetrahydrofolate cyclohydrolase	H
Nest_1171	10.63258	0.001111	0.017334	COG0499	S-adenosylhomocysteine hydrolase	H
Nest_20	8.829786	0.002964	0.032684	COG1932	Phosphoserine aminotransferase	H
Nest_300	10.39991	0.00126	0.018851	COG0373	Glutamyl-tRNA reductase	H
Nest_1065	9.179272	0.002448	0.028741	COG0111	Phosphoglycerate dehydrogenase and related dehydrogenases	H
Nest_777	8.731762	0.003127	0.03381	COG0352	Thiamine monophosphate synthase	H
Nest_898	9.964756	0.001596	0.020743	COG2267	Lysophospholipase	I
Nest_946	9.302128	0.002289	0.027467	COG1028	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	I
Nest_141	10.30641	0.001326	0.019048	COG4799	Acetyl-CoA carboxylase, carboxyltransferase component (subunits alpha and beta)	I
Nest_1158	7.272012	0.007004	0.050316	COG1022	Long-chain acyl-CoA synthetases (AMP-forming)	I

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_1217	20.06417	7.49E-06	0.001022	COG0365	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	I
Nest_565	13.82991	0.0002	0.006428	COG1022	Long-chain acyl-CoA synthetases (AMP-forming)	I
Nest_622	7.659872	0.005646	0.046013	COG2519	tRNA(1-methyladenosine) methyltransferase and related methyltransferases	J
Nest_76	9.704308	0.001838	0.022556	COG0080	Ribosomal protein L11	J
Nest_77	11.09117	0.000867	0.014717	COG0081	Ribosomal protein L1	J
Nest_737	7.432694	0.006405	0.048453	COG0349	Ribonuclease D	J
Nest_759	7.391456	0.006553	0.048683	COG0172	Seryl-tRNAsynthetase	J
Nest_90	18.42089	1.77E-05	0.001758	COG0087	Ribosomal protein L3	J
Nest_820	11.76226	0.000604	0.012223	COG0617	tRNA nucleotidyltransferase/poly(A) polymerase	J
Nest_107	9.233168	0.002377	0.02821	COG0200	Ribosomal protein L15	J
Nest_112	11.26266	0.000791	0.014393	COG0099	Ribosomal protein S13	J
Nest_973	12.66277	0.000373	0.009408	COG0335	Ribosomal protein L19	J
Nest_1056	7.672808	0.005606	0.046013	COG1189	Predicted rRNAmethylase	J
Nest_1182	14.27099	0.000158	0.005746	COG0566	rRNAmethylases	J
Nest_1198	7.827784	0.005145	0.044946	COG0072	Phenylalanyl-tRNAsynthetase beta subunit	J
Nest_243	8.36966	0.003815	0.038938	COG0030	Dimethyladenosine transferase (rRNA methylation)	J
Nest_321	10.22236	0.001387	0.019677	COG0423	Glycyl-tRNAsynthetase (class II)	J
Nest_351	11.1155	0.000856	0.014717	COG0154	Asp-tRNAAsn/Glu-tRNAGlnamidotransferase A subunit and related amidases	J
Nest_606	11.45713	0.000712	0.013644	COG1846	Transcriptional regulators	K
Nest_75	7.987334	0.004711	0.042999	COG0250	Transcription antiterminator	K
Nest_814	8.639928	0.003289	0.034867	COG1475	Predicted transcriptional regulators	K
Nest_161	9.800004	0.001745	0.02242	COG1802	Transcriptional regulators	K

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_198	10.17461	0.001424	0.019898	COG1842	Phage shock protein A (IM30), suppresses sigma54-dependent transcription	K
Nest_466	11.1597	0.000836	0.014717	COG0322	Nuclease subunit of the excinuclease complex	L
Nest_469	11.83276	0.000582	0.011991	COG4581	Superfamily II RNA helicase	L
Nest_486	17.59359	2.74E-05	0.002133	COG0749	DNA polymerase I - 3'-5' exonuclease and polymerase domains	L
Nest_638	12.63258	0.000379	0.009408	COG0187	Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit	L
Nest_675	12.73772	0.000358	0.009408	COG4581	Superfamily II RNA helicase	L
Nest_932	14.29195	0.000157	0.005746	COG0550	Topoisomerase IA	L
Nest_943	28.24589	1.07E-07	3.89E-05	COG0188	Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), A subunit	L
Nest_1007	8.12259	0.004372	0.041813	COG2176	DNA polymerase III, alpha subunit (gram-positive type)	L
Nest_1054	19.68422	9.14E-06	0.001108	COG0497	ATPase involved in DNA repair	L
Nest_1080	23.76921	1.09E-06	0.000237	COG0708	Exonuclease III	L
Nest_149	11.02439	0.000899	0.014878	COG0272	NAD-dependent DNA ligase (contains BRCT domain type II)	L
Nest_316	11.07284	0.000876	0.014717	COG0358	DNA primase (bacterial type)	L
Nest_47	7.966714	0.004765	0.042999	COG4948	L-alanine-DL-glutamate epimerase and related enzymes of enolase superfamily	M
Nest_503	17.1496	3.45E-05	0.002515	COG1181	D-alanine-D-alanine ligase and related ATP-grasp enzymes	M
Nest_556	10.77655	0.001028	0.016508	COG0769	UDP-N-acetylmuramyl tripeptide synthase	M
Nest_557	7.764358	0.005329	0.045108	COG0768	Cell division protein FtsI/penicillin-binding protein 2	M
Nest_792	12.70651	0.000364	0.009408	COG0768	Cell division protein FtsI/penicillin-binding protein 2	M
Nest_126	11.84716	0.000577	0.011991	COG0787	Alanine racemase	M
Nest_1200	8.069974	0.0045	0.042367	COG0438	Glycosyltransferase	M

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_1213	7.341486	0.006738	0.048877	COG0744	Membrane carboxypeptidase (penicillin-binding protein)	M
Nest_22	8.330678	0.003898	0.03939	COG0791	Cell wall-associated hydrolases (invasion-associated proteins)	M
Nest_255	7.426512	0.006427	0.048453	COG0750	Predicted membrane-associated Zn-dependent proteases 1	M
Nest_53	7.876402	0.005008	0.044106	COG0501	Zn-dependent protease with chaperone function	O
Nest_443	7.1559	0.007472	0.05052	COG1066	Predicted ATP-dependent serine protease	O
Nest_447	7.155608	0.007473	0.05052	COG3590	Predicted metalloendopeptidase	O
Nest_585	8.164214	0.004272	0.041288	COG0691	tmRNA-binding protein	O
Nest_680	9.970708	0.001591	0.020743	COG1585	Membrane protein implicated in regulation of membrane protease activity	O
Nest_717	9.74122	0.001802	0.022556	COG1391	Glutamine synthetaseadenylyltransferase	O
Nest_130	15.6929	7.45E-05	0.003698	COG0533	Metal-dependent proteases with possible chaperone activity	O
Nest_1105	16.4323	5.04E-05	0.003167	COG1331	Highly conserved protein containing a thioredoxin domain	O
Nest_37	7.4904	0.006203	0.048453	COG1333	ResB protein required for cytochrome c biosynthesis	O
Nest_374	8.449956	0.003651	0.037965	COG0443	Molecular chaperone	O
Nest_436	7.216254	0.007225	0.05052	COG1117	ABC-type phosphate transport system, ATPase component	P
Nest_438	14.15455	0.000168	0.005746	COG0226	ABC-type phosphate transport system, periplasmic component	P
Nest_54	10.41535	0.00125	0.018851	COG1283	Na+/phosphate symporter	P
Nest_571	7.731756	0.005426	0.045577	COG1108	ABC-type Mn2+/Zn2+ transport systems, permease components	P
Nest_582	12.11269	0.000501	0.011161	COG1121	ABC-type Mn/Zn transport systems, ATPase component	P
Nest_1132	7.518814	0.006106	0.048427	COG0471	Di- and tricarboxylate transporters	P
Nest_133	10.67913	0.001084	0.017148	*	*	Q
Nest_511	8.522358	0.003508	0.036835	COG0179	2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase (catechol pathway)	Q
Nest_169	11.37377	0.000745	0.013787	*	*	R

Nest_477	15.01319	0.000107	0.004597	COG4552	Predicted acetyltransferase involved in intracellular survival and related acetyltransferases	R
Nest_72	10.05328	0.001521	0.020252	COG0354	Predicted aminomethyltransferase related to GcvT	R
Nest_694	8.174828	0.004248	0.041288	COG0536	Predicted GTPase	R
Nest_753	7.627026	0.00575	0.046511	COG0561	Predicted hydrolases of the HAD superfamily	R
Nest_776	7.067754	0.007848	0.052579	COG0693	Putative intracellular protease/amidase	R
Nest_794	8.68556	0.003207	0.034338	COG0515	Serine/threonine protein kinase	R
Nest_818	7.471392	0.006269	0.048453	COG0728	Uncharacterized membrane protein, putative virulence factor	R
Nest_837	16.58262	4.66E-05	0.003167	COG5271	AAA ATPase containing von Willebrand factor type A (vWA) domain	R
Nest_928	7.336022	0.006759	0.048877	COG1205	Distinct helicase family with a unique C-terminal domain including a metal-binding cysteine cluster	R
Nest_930	7.794798	0.00524	0.045053	COG1054	Predicted sulfurtransferase	R
Nest_950	7.412256	0.006478	0.048453	COG0488	ATPase components of ABC transporters with duplicated ATPase domains	R
Nest_153	11.57448	0.000669	0.013275	COG2374	Predicted extracellular nuclease	R
Nest_207	14.47088	0.000142	0.005551	COG0714	MoxR-like ATPases	R
Nest_940	7.690914	0.00555	0.046013	COG0596	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	R
Nest_32	11.38009	0.000742	0.013787	*	*	S
Nest_624	12.51884	0.000403	0.009564	*	*	S
Nest_1119	7.231174	0.007165	0.05052	*	*	S
Nest_1167	16.2489	5.55E-05	0.003192	*	*	S
Nest_479	9.71227	0.00183	0.022556	COG3428	Predicted membrane protein	S

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_698	12.45582	0.000417	0.009681	COG4243	Predicted membrane protein	S
Nest_786	7.19391	0.007315	0.05052	COG2353	Uncharacterized conserved protein	S
Nest_821	10.36471	0.001284	0.018955	COG5650	Predicted integral membrane protein	S
Nest_1169	10.09841	0.001484	0.020252	COG1300	Uncharacterized membrane protein	S
Nest_251	7.933328	0.004853	0.043441	COG0779	Uncharacterized protein conserved in bacteria	S
Nest_29	9.347612	0.002233	0.027091	COG1814	Uncharacterized membrane protein	S
Nest_1142	17.7101	2.57E-05	0.002133	COG4805	Uncharacterized protein conserved in bacteria	S
Nest_178	7.177742	0.007381	0.05052	COG0745	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	T
Nest_15	18.58567	1.62E-05	0.001758	COG0642	Signal transduction histidine kinase	T
Nest_1072	8.243618	0.00409	0.040451	COG0842	ABC-type multidrug transport system, permease component	V
***Nesterenkonia* sp. NP1**						
Nest_574	9.940122	0.001617	0.023878	COG0567	2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, and related enzymes	C
Nest_716	16.30191	5.4E-05	0.003932	COG2141	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	C
Nest_741	11.66448	0.000637	0.015123	COG1048	Aconitase A	C
Nest_749	8.39257	0.003768	0.038094	COG0221	Inorganic pyrophosphatase	C
Nest_770	8.97491	0.002737	0.032845	COG0584	Glycerophosphoryldiester phosphodiesterase	C
Nest_843	12.88973	0.00033	0.012466	COG2025	Electron transfer flavoprotein, alpha subunit	C
Nest_876	19.36492	1.08E-05	0.00131	COG1290	Cytochrome b subunit of the bc complex	C
Nest_1027	10.21034	0.001397	0.022442	COG0712	F0F1-type ATP synthase, delta subunit (mitochondrial oligomycin sensitivity protein)	C

Nest_1028	11.28094	0.000783	0.017102	COG0711	F0F1-type ATP synthase, subunit b	C
Nest_1071	8.315236	0.003931	0.039386	COG2141	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	C
Nest_1097	9.29956	0.002292	0.029825	COG1252	NADH dehydrogenase, FAD-containing subunit	C
Nest_156	10.20909	0.001397	0.022442	COG2224	Isocitratelyase	C
Nest_315	11.68252	0.000631	0.015123	COG2352	Phosphoenolpyruvate carboxylase	C
Nest_395	17.60313	2.72E-05	0.002286	COG1009	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na+/H+ antiporter, MnhA subunit	C
Nest_610	12.81218	0.000344	0.012466	COG1674	DNA segregation ATPase FtsK/SpoIIIE and related proteins	D
Nest_199	34.49992	4.26E-09	4.65E-06	COG1196	Chromosome segregation ATPases	D
Nest_521	9.527774	0.002024	0.028334	COG0493	NADPH-dependent glutamate synthase beta chain and related oxidoreductases	E
Nest_527	9.822748	0.001724	0.024766	COG0134	Indole-3-glycerol phosphate synthase	E
Nest_532	10.95339	0.000934	0.018271	COG0040	ATP phosphoribosyltransferase	E
Nest_630	8.803804	0.003006	0.034054	COG0308	Aminopeptidase N	E
Nest_643	7.916654	0.004898	0.04611	COG0260	Leucyl aminopeptidase	E
Nest_658	18.75738	1.48E-05	0.001474	COG1063	Threonine dehydrogenase and related Zn-dependent dehydrogenases	E
Nest_718	10.94819	0.000937	0.018271	COG0174	Glutamine synthetase	E
Nest_727	14.82283	0.000118	0.006787	COG0031	Cysteine synthase	E
Nest_707	20.90656	4.82E-06	0.000658	COG0405	Gamma-glutamyltransferase	E
Nest_1137	9.027248	0.00266	0.032341	COG0436	Aspartate/tyrosine/aromatic aminotransferase	E
Nest_165	8.951354	0.002773	0.03291	COG1003	Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain	E

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_203	8.894796	0.00286	0.03306	COG1104	Cysteine sulfinatedesulfinase/cysteine desulfurase and related enzymes	E
Nest_285	7.515076	0.006118	0.053882	COG4166	ABC-type oligopeptide transport system, periplasmic component	E
Nest_298	10.13973	0.001451	0.022825	COG2902	NAD-specific glutamate dehydrogenase	E
Nest_314	30.63145	3.12E-08	1.70E-05	COG1505	Serine proteases of the peptidase family S9A	E
Nest_51	10.04651	0.001526	0.023476	COG0537	Diadenosinetetraphosphate (Ap4A) hydrolase and other HIT family hydrolases	F
Nest_484	9.70963	0.001833	0.025996	COG0634	Hypoxanthine-guanine phosphoribosyltransferase	F
Nest_66	8.63604	0.003296	0.036353	COG0207	Thymidylate synthase	F
Nest_1090	11.20546	0.000816	0.017147	COG0213	Thymidine phosphorylase	F
Nest_1207	8.524508	0.003504	0.036792	COG0151	Phosphoribosylamine-glycine ligase	F
Nest_344	15.77706	7.13E-05	0.004864	COG0046	Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain	F
Nest_453	8.884456	0.002876	0.03306	COG0021	Transketolase	G
Nest_768	10.62387	0.001116	0.019176	COG3839	ABC-type sugar transport systems, ATPase components	G
Nest_132	8.414878	0.003722	0.038094	COG1472	Beta-glucosidase-related glycosidases	G
Nest_222	8.214446	0.004156	0.041256	COG1109	Phosphomannomutase	G
Nest_672	13.03441	0.000306	0.012466	COG2182	Maltose-binding periplasmic proteins/domains	G
Nest_780	12.56637	0.000393	0.012466	COG1653	ABC-type sugar transport system, periplasmic component	G
Nest_563	24.35912	7.99E-07	0.000175	COG0142	Geranylgeranyl pyrophosphate synthase	H
Nest_614	10.00666	0.00156	0.023656	COG0095	Lipoate-protein ligase A	H
Nest_853	12.14241	0.000493	0.013744	COG0635	Coproporphyrinogen III oxidase and related Fe-S oxidoreductases	H
Nest_1171	11.81494	0.000588	0.014969	COG0499	S-adenosylhomocysteine hydrolase	H
Nest_181	12.62233	0.000381	0.012466	COG0276	Protohemeferro-lyase (ferrochelatase)	H

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_1065	9.294532	0.002298	0.029825	COG0111	Phosphoglycerate dehydrogenase and related dehydrogenases	H
Nest_631	9.08983	0.00257	0.031896	COG1960	Acyl-CoA dehydrogenases	I
Nest_632	7.722662	0.005453	0.049376	COG0183	Acetyl-CoA acetyltransferase	I
Nest_864	23.25621	1.42E-06	0.000258	COG0020	Undecaprenyl pyrophosphate synthase	I
Nest_962	13.91246	0.000192	0.009507	COG0331	(acyl-carrier-protein) S-malonyltransferase	I
Nest_514	13.17719	0.000283	0.012466	COG0143	Methionyl-tRNAsynthetase	J
Nest_622	15.58375	7.89E-05	0.00507	COG2519	tRNA(1-methyladenosine) methyltransferase and related methyltransferases	J
Nest_661	7.716724	0.005471	0.049376	COG0441	Threonyl-tRNAsynthetase	J
Nest_697	10.68858	0.001078	0.019176	COG1530	Ribonucleases G and E	J
Nest_722	10.12446	0.001463	0.022825	COG0024	Methionine aminopeptidase	J
Nest_93	7.520574	0.0061	0.053882	COG0090	Ribosomal protein L2	J
Nest_828	7.980996	0.004727	0.044887	COG0359	Ribosomal protein L9	J
Nest_1033	9.414102	0.002153	0.029764	COG0216	Protein chain release factor A	J
Nest_1039	10.41863	0.001248	0.020641	COG0018	Arginyl-tRNAsynthetase	J
Nest_1182	26.239	3.02E-07	0.00011	COG0566	rRNAmethylases	J
Nest_205	8.15812	0.004287	0.042174	COG0482	Predicted tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase, contains the PP-loop ATPase domain	J
Nest_249	8.533618	0.003486	0.036792	COG0532	Translation initiation factor 2 (IF-2; GTPase)	J
Nest_350	14.84846	0.000117	0.006787	COG0227	Ribosomal protein L28	J
Nest_501	7.7422	0.005395	0.049376	COG1316	Transcriptional regulator	K
Nest_1193	11.53144	0.000684	0.015282	COG1438	Arginine repressor	K
Nest_1227	13.552	0.000232	0.011017	COG0846	NAD-dependent protein deacetylases, SIR2 family	K
Nest_542	11.57004	0.00067	0.015282	COG0587	DNA polymerase III, alpha subunit	L

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_675	9.167186	0.002464	0.031165	COG4581	Superfamily II RNA helicase	L
Nest_932	8.482652	0.003585	0.037289	COG0550	Topoisomerase IA	L
Nest_1007	8.119888	0.004378	0.042576	COG2176	DNA polymerase III, alpha subunit (gram-positive type)	L
Nest_1080	11.70097	0.000625	0.015123	COG0708	Exonuclease III	L
Nest_149	11.52757	0.000686	0.015282	COG0272	NAD-dependent DNA ligase (contains BRCT domain type II)	L
Nest_301	19.01366	1.3E-05	0.001417	COG0513	Superfamily II DNA and RNA helicases	L
Nest_1229	12.48038	0.000411	0.012466	COG0350	Methylated DNA-protein cysteine methyltransferase	L
Nest_424	9.939008	0.001618	0.023878	COG0791	Cell wall-associated hydrolases (invasion-associated proteins)	M
Nest_554	8.792358	0.003025	0.034054	COG0472	UDP-N-acetylmuramylpentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase	M
Nest_556	11.17012	0.000831	0.017147	COG0769	UDP-N-acetylmuramyl tripeptide synthase	M
Nest_557	7.72749	0.005439	0.049376	COG0768	Cell division protein FtsI/penicillin-binding protein 2	M
Nest_677	7.891276	0.004967	0.046362	COG0463	Glycosyltransferases involved in cell wall biogenesis	M
Nest_792	11.08029	0.000873	0.017644	COG0768	Cell division protein FtsI/penicillin-binding protein 2	M
Nest_123	8.892058	0.002864	0.03306	COG0449	Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains	M
Nest_717	12.98266	0.000314	0.012466	COG1391	Glutamine synthetaseadenylyltransferase	O
Nest_958	7.530402	0.006067	0.053882	COG0450	Peroxiredoxin	O
Nest_1105	10.58994	0.001137	0.019176	COG1331	Highly conserved protein containing a thioredoxin domain	O
Nest_14	8.091044	0.004448	0.042611	COG0459	Chaperonin GroEL (HSP60 family)	O
Nest_414	9.867284	0.001682	0.024496	COG3221	ABC-type phosphate/phosphonate transport system, periplasmic component	P
Nest_438	10.81005	0.00101	0.019007	COG0226	ABC-type phosphate transport system, periplasmic component	P

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_451	14.1088	0.000173	0.008972	COG1119	ABC-type molybdenum transport system, ATPase component/photorepair protein PhrA	P
Nest_54	12.64444	0.000377	0.012466	COG1283	Na+/phosphate symporter	P
Nest_581	8.554704	0.003446	0.036792	COG0803	ABC-type metal ion transport system, periplasmic component/surface adhesin	P
Nest_583	21.99981	2.73E-06	0.000425	COG1108	ABC-type Mn2+/Zn2+ transport systems, permease components	P
Nest_135	8.406578	0.003739	0.038094	COG4606	ABC-type enterochelin transport system, permease component	P
Nest_1060	8.561966	0.003433	0.036792	COG0704	Phosphate uptake regulator	P
Nest_1063	10.90528	0.000959	0.018371	COG2239	Mg/Co/Ni transporter MgtE (contains CBS domain)	P
Nest_1190	16.92524	3.89E-05	0.003033	COG1118	ABC-type sulfate/molybdate transport systems, ATPase component	P
Nest_289	18.3161	1.87E-05	0.001703	COG1230	Co/Zn/Cd efflux system component	P
Nest_392	8.7698	0.003063	0.034126	COG1863	Multisubunit Na+/H+ antiporter, MnhE subunit	P
Nest_830	12.86085	0.000336	0.012466	*	*	Q
Nest_621	12.86881	0.000334	0.012466	COG0637	Predicted phosphatase/phosphohexomutase	R
Nest_818	8.569948	0.003418	0.036792	COG0728	Uncharacterized membrane protein, putative virulence factor	R
Nest_915	10.76483	0.001034	0.019146	COG1279	Lysine efflux permease	R
Nest_928	10.58286	0.001141	0.019176	COG1205	Distinct helicase family with a unique C-terminal domain including a metal-binding cysteine cluster	R
Nest_1088	9.370276	0.002205	0.029825	COG1473	Metal-dependent amidase/aminoacylase/carboxypeptidase	R
Nest_1106	8.108544	0.004406	0.042576	COG1272	Predicted membrane protein, hemolysin III homolog	R
Nest_1185	9.153126	0.002483	0.031165	COG3973	Superfamily I DNA and RNA helicases	R
Nest_153	9.276184	0.002322	0.029825	COG2374	Predicted extracellular nuclease	R
Nest_361	11.80907	0.000589	0.014969	COG0820	Predicted Fe-S-cluster redox enzyme	R
Nest_371	12.10284	0.000503	0.013744	COG0220	Predicted S-adenosylmethionine-dependent methyltransferase	R

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_625	25.57288	4.26E-07	0.000116	*	*	S
Nest_685	9.280082	0.002317	0.029825	*	*	S
Nest_799	10.58678	0.001139	0.019176	*	*	S
Nest_855	14.49802	0.00014	0.007661	*	*	S
Nest_879	9.337268	0.002245	0.029825	*	*	S
Nest_200	12.47181	0.000413	0.012466	*	*	S
Nest_1142	11.16794	0.000832	0.017147	COG4805	Uncharacterized protein conserved in bacteria	S
Nest_176	9.023364	0.002666	0.032341	COG1576	Uncharacterized conserved protein	S
Nest_752	12.18994	0.00048	0.013744	COG1966	Carbon starvation protein, predicted membrane protein	T
Nest_859	12.03566	0.000522	0.013901	COG1702	Phosphate starvation-inducible protein PhoH, predicted ATPase	T
Nest_961	12.73298	0.000359	0.012466	COG2508	Regulator of polyketide synthase expression	T
Nest_1074	12.43042	0.000422	0.012466	COG1136	ABC-type antimicrobial peptide transport system, ATPase component	V
Nest_1215	10.64386	0.001104	0.019176	COG1132	ABC-type multidrug transport system, ATPase and permease components	V
***N. alba* DSM 19423**						
Nest_574	9.580902	0.001966	0.020845	COG0567	2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, and related enzymes	C
Nest_741	9.040058	0.002641	0.025752	COG1048	Aconitase A	C
Nest_770	19.88761	8.21E-06	0.000815	COG0584	Glycerophosphoryldiester phosphodiesterase	C
Nest_868	12.08023	0.00051	0.009532	COG1622	Heme/copper-type cytochrome/quinol oxidases, subunit 2	C
Nest_876	10.36838	0.001282	0.015907	COG1290	Cytochrome b subunit of the bc complex	C

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_913	42.49536	7.08E-11	7.74E-08	COG2141	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	C
Nest_960	7.370386	0.006631	0.047325	COG2609	Pyruvate dehydrogenase complex, dehydrogenase (E1) component	C
Nest_1064	13.27248	0.000269	0.006684	COG2838	Monomeric isocitrate dehydrogenase	C
Nest_1125	10.17064	0.001427	0.017123	COG0604	NADPH:quinone reductase and related Zn-dependent oxidoreductases	C
Nest_27	7.272716	0.007001	0.049007	COG0045	Succinyl-CoA synthetase, beta subunit	C
Nest_315	10.54958	0.001162	0.01529	COG2352	Phosphoenolpyruvate carboxylase	C
Nest_375	17.40625	3.02E-05	0.00157	COG1012	NAD-dependent aldehyde dehydrogenases	C
Nest_549	19.32548	1.1E-05	0.00086	COG0206	Cell division GTPase	D
Nest_610	14.96977	0.000109	0.003409	COG1674	DNA segregation ATPase FtsK/SpoIIIE and related proteins	D
Nest_813	17.60502	2.72E-05	0.00157	COG1192	ATPases involved in chromosome partitioning	D
Nest_1220	8.028066	0.004606	0.037298	COG1192	ATPases involved in chromosome partitioning	D
Nest_683	13.87119	0.000196	0.005235	COG1063	Threonine dehydrogenase and related Zn-dependent dehydrogenases	E
Nest_52	14.26598	0.000159	0.004561	COG0493	NADPH-dependent glutamate synthase beta chain and related oxidoreductases	E
Nest_506	10.45145	0.001226	0.015561	COG0066	3-isopropylmalate dehydratase small subunit	E
Nest_507	11.40251	0.000733	0.011124	COG0065	3-isopropylmalate dehydratase large subunit	E
Nest_512	12.22305	0.000472	0.009532	COG0115	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase	E
Nest_518	12.92951	0.000323	0.007515	COG0129	Dihydroxyacid dehydratase/phosphogluconate dehydratase	E
Nest_521	19.57679	9.66E-06	0.000823	COG0493	NADPH-dependent glutamate synthase beta chain and related oxidoreductases	E

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_527	8.598702	0.003364	0.029866	COG0134	Indole-3-glycerol phosphate synthase	E
Nest_532	13.15452	0.000287	0.006915	COG0040	ATP phosphoribosyltransferase	E
Nest_630	13.56975	0.00023	0.005976	COG0308	Aminopeptidase N	E
Nest_692	9.878566	0.001672	0.018444	COG0014	Gamma-glutamyl phosphate reductase	E
Nest_718	14.03196	0.00018	0.005032	COG0174	Glutamine synthetase	E
Nest_766	8.377358	0.003799	0.032668	COG0160	4-aminobutyrate aminotransferase and related aminotransferases	E
Nest_1035	8.774432	0.003055	0.027998	COG0083	Homoserine kinase	E
Nest_1038	7.612142	0.005798	0.043662	COG0019	Diaminopimelate decarboxylase	E
Nest_163	7.804858	0.005211	0.04007	COG1984	Allophanate hydrolase subunit 2	E
Nest_165	8.761398	0.003077	0.027998	COG1003	Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain	E
Nest_282	10.20029	0.001404	0.017123	COG0444	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	E
Nest_298	21.4136	3.7E-06	0.000577	COG2902	NAD-specific glutamate dehydrogenase	E
Nest_314	16.99317	3.75E-05	0.001707	COG1505	Serine proteases of the peptidase family S9A	E
Nest_384	16.87052	4E-05	0.001748	COG0147	Anthranilate/para-aminobenzoate synthases component I	E
Nest_575	9.422698	0.002143	0.022289	*	*	E
Nest_51	7.53351	0.006056	0.044685	COG0537	Diadenosinetetraphosphate (Ap4A) hydrolase and other HIT family hydrolases	F
Nest_326	10.73124	0.001053	0.014561	COG0044	Dihydroorotase and related cyclic amidohydrolases	F
Nest_892	9.973774	0.001588	0.018122	COG0194	Guanylate kinase	F
Nest_1057	17.99445	2.22E-05	0.001423	COG2812	DNA polymerase III, gamma/tau subunits	F
Nest_1145	9.043602	0.002636	0.025752	COG0274	Deoxyribose-phosphate aldolase	F
Nest_1207	9.374722	0.0022	0.022664	COG0151	Phosphoribosylamine-glycine ligase	F

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_344	13.86386	0.000197	0.005235	COG0046	Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain	F
Nest_453	11.95594	0.000545	0.009532	COG0021	Transketolase	G
Nest_456	7.418272	0.006456	0.046692	COG0364	Glucose-6-phosphate 1-dehydrogenase	G
Nest_730	11.41956	0.000727	0.011124	COG0362	6-phosphogluconate dehydrogenase	G
Nest_840	9.235134	0.002374	0.024005	COG3386	Gluconolactonase	G
Nest_125	9.445952	0.002116	0.02222	COG0063	Predicted sugar kinase	G
Nest_780	10.70194	0.00107	0.014609	COG1653	ABC-type sugar transport system, periplasmic component	G
Nest_535	9.205194	0.002413	0.024024	COG0108	3,4-dihydroxy-2-butanone 4-phosphate synthase	H
Nest_563	10.02987	0.00154	0.018085	COG0142	Geranylgeranyl pyrophosphate synthase	H
Nest_740	7.553798	0.005988	0.044635	COG1154	Deoxyxylulose-5-phosphate synthase	H
Nest_772	10.63692	0.001109	0.014945	COG0476	Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 2	H
Nest_853	11.71939	0.000619	0.009933	COG0635	Coproporphyrinogen III oxidase and related Fe-S oxidoreductases	H
Nest_631	8.905502	0.002843	0.027065	COG1960	Acyl-CoA dehydrogenases	I
Nest_1158	11.95805	0.000544	0.009532	COG1022	Long-chain acyl-CoA synthetases (AMP-forming)	I
Nest_42	8.017384	0.004633	0.037298	COG0318	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II	I
Nest_565	7.302758	0.006885	0.048505	COG1022	Long-chain acyl-CoA synthetases (AMP-forming)	I
Nest_601	8.458382	0.003634	0.031685	COG0324	tRNA delta(2)-isopentenylpyrophosphate transferase	J
Nest_602	8.453994	0.003642	0.031685	COG0621	2-methylthioadenine synthetase	J
Nest_696	17.43415	2.97E-05	0.00157	COG0261	Ribosomal protein L21	J
Nest_697	17.26106	3.26E-05	0.001616	COG1530	Ribonucleases G and E	J
Nest_92	14.46236	0.000143	0.004337	COG0089	Ribosomal protein L23	J
Nest_820	9.200114	0.00242	0.024024	COG0617	tRNA nucleotidyltransferase/poly(A) polymerase	J

Nest_105	15.05382	0.000104	0.003356	COG0098	Ribosomal protein S5	J
Nest_110	8.041854	0.004571	0.037298	COG0361	Translation initiation factor 1 (IF-1)	J
Nest_113	8.447232	0.003656	0.031685	COG0100	Ribosomal protein S11	J
Nest_118	8.978548	0.002732	0.026398	COG0102	Ribosomal protein L13	J
Nest_1129	12.39436	0.000431	0.008873	COG0012	Predicted GTPase, probable translation factor	J
Nest_147	15.49596	8.27E-05	0.002949	COG0064	Asp-tRNAAsn/Glu-tRNAGlnamidotransferase B subunit (PET112 homolog)	J
Nest_1182	11.72398	0.000617	0.009933	COG0566	rRNAmethylases	J
Nest_214	7.985302	0.004716	0.037317	COG1544	Ribosome-associated protein Y (PSrp-1)	J
Nest_266	9.88585	0.001666	0.018444	COG0052	Ribosomal protein S2	J
Nest_351	32.73737	1.05E-08	5.76E-06	COG0154	Asp-tRNAAsn/Glu-tRNAGlnamidotransferase A subunit and related amidases	J
Nest_491	11.85998	0.000574	0.009786	COG1316	Transcriptional regulator	K
Nest_501	15.16876	9.83E-05	0.003253	COG1316	Transcriptional regulator	K
Nest_84	16.07282	6.1E-05	0.002377	COG0085	DNA-directed RNA polymerase, beta subunit/140 kD subunit	K
Nest_914	12.47999	0.000411	0.00875	COG0583	Transcriptional regulator	K
Nest_955	12.0713	0.000512	0.009532	COG2345	Predicted transcriptional regulator	K
Nest_161	8.025628	0.004612	0.037298	COG1802	Transcriptional regulators	K
Nest_45	7.822962	0.005159	0.039952	COG0819	Putative transcription activator	K
Nest_469	23.64935	1.16E-06	0.000295	COG4581	Superfamily II RNA helicase	L
Nest_472	8.03618	0.004585	0.037298	COG0556	Helicase subunit of the DNA excision repair complex	L
Nest_486	14.33782	0.000153	0.004509	COG0749	DNA polymerase I - 3'-5' exonuclease and polymerase domains	L
Nest_675	8.6138	0.003336	0.029862	COG4581	Superfamily II RNA helicase	L
Nest_712	9.726044	0.001817	0.019642	COG0513	Superfamily II DNA and RNA helicases	L

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_932	20.84002	4.99E-06	0.000682	COG0550	Topoisomerase IA	L
Nest_967	12.64701	0.000376	0.008215	COG4974	Site-specific recombinase XerD	L
Nest_1054	10.08167	0.001498	0.017775	COG0497	ATPase involved in DNA repair	L
Nest_1080	20.23725	6.84E-06	0.000795	COG0708	Exonuclease III	L
Nest_316	11.25058	0.000796	0.011589	COG0358	DNA primase (bacterial type)	L
Nest_336	8.765752	0.003069	0.027998	COG0353	Recombinational DNA repair protein (RecF pathway)	L
Nest_424	15.50061	8.25E-05	0.002949	COG0791	Cell wall-associated hydrolases (invasion-associated proteins)	M
Nest_432	16.27076	5.49E-05	0.002221	COG0451	Nucleoside-diphosphate-sugar epimerases	M
Nest_503	8.091134	0.004448	0.037298	COG1181	D-alanine-D-alanine ligase and related ATP-grasp enzymes	M
Nest_523	8.837158	0.002952	0.027548	COG0682	Prolipoproteindiacylglyceryltransferase	M
Nest_792	11.93831	0.00055	0.009532	COG0768	Cell division protein FtsI/penicillin-binding protein 2	M
Nest_941	11.16476	0.000834	0.011978	COG0451	Nucleoside-diphosphate-sugar epimerases	M
Nest_267	15.4724	8.37E-05	0.002949	COG0438	Glycosyltransferase	M
Nest_443	11.82462	0.000585	0.00982	COG1066	Predicted ATP-dependent serine protease	O
Nest_447	25.89354	3.61E-07	0.000131	COG3590	Predicted metalloendopeptidase	O
Nest_679	7.547736	0.006009	0.044635	COG0330	Membrane protease subunits, stomatin/prohibitin homologs	O
Nest_717	7.705682	0.005505	0.041744	COG1391	Glutamine synthetaseadenylyltransferase	O
Nest_954	21.60518	3.35E-06	0.000577	COG0719	ABC-type transport system involved in Fe-S cluster assembly, permease component	O
Nest_130	11.7524	0.000608	0.009933	COG0533	Metal-dependent proteases with possible chaperone activity	O
Nest_1105	11.49365	0.000698	0.011038	COG1331	Highly conserved protein containing a thioredoxin domain	O
Nest_1230	16.39265	5.15E-05	0.002162	COG0542	ATPases with chaperone activity, ATP-binding subunit	O
Nest_37	9.967666	0.001593	0.018122	COG1333	ResB protein required for cytochrome c biosynthesis	O
Nest_346	10.98888	0.000917	0.012832	COG0526	Thiol-disulfide isomerase and thioredoxins	O

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_374	10.46724	0.001215	0.015561	COG0443	Molecular chaperone	O
Nest_438	18.87473	1.4E-05	0.001016	COG0226	ABC-type phosphate transport system, periplasmic component	P
Nest_451	11.97409	0.000539	0.009532	COG1119	ABC-type molybdenum transport system, ATPase component/photorepair protein PhrA	P
Nest_54	10.60017	0.001131	0.015059	COG1283	Na+/phosphate symporter	P
Nest_581	7.72896	0.005434	0.041498	COG0803	ABC-type metal ion transport system, periplasmic component/surface adhesin	P
Nest_686	9.845134	0.001703	0.018595	COG0753	Catalase	P
Nest_951	7.51642	0.006114	0.044808	COG1122	ABC-type cobalt transport system, ATPase component	P
Nest_980	20.11827	7.28E-06	0.000795	COG0004	Ammonia permease	P
Nest_400	9.347728	0.002233	0.022785	COG0474	Cation transport ATPase	P
Nest_290	10.38164	0.001273	0.015907	COG1228	Imidazolonepropionase and related amidohydrolases	Q
Nest_133	12.45599	0.000417	0.00875	*	*	Q
Nest_991	13.31289	0.000264	0.006684	*	*	Q
Nest_1139	12.6573	0.000374	0.008215	*	*	Q
Nest_477	12.11606	0.0005	0.009532	COG4552	Predicted acetyltransferase involved in intracellular survival and related acetyltransferases	R
Nest_72	7.999404	0.004679	0.037298	COG0354	Predicted aminomethyltransferase related to GcvT	R
Nest_829	11.30344	0.000774	0.011416	COG0596	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	R
Nest_905	8.004452	0.004666	0.037298	COG2252	Permeases	R
Nest_915	8.62815	0.00331	0.029862	COG1279	Lysine efflux permease	R
Nest_929	8.900912	0.00285	0.027065	COG3393	Predicted acetyltransferase	R
Nest_1004	10.17541	0.001423	0.017123	COG0579	Predicted dehydrogenase	R

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_1042	10.49475	0.001197	0.015561	COG2329	Uncharacterized enzyme involved in biosynthesis of extracellular polysaccharides	R
Nest_1088	18.64037	1.58E-05	0.001077	COG1473	Metal-dependent amidase/aminoacylase/carboxypeptidase	R
Nest_1140	7.330624	0.006779	0.048069	COG3173	Predicted aminoglycoside phosphotransferase	R
Nest_153	23.352	1.35E-06	0.000295	COG2374	Predicted extracellular nuclease	R
Nest_1224	8.85611	0.002921	0.027499	COG2358	TRAP-type uncharacterized transport system, periplasmic component	R
Nest_190	17.17767	3.4E-05	0.001616	COG3173	Predicted aminoglycoside phosphotransferase	R
Nest_207	13.12564	0.000291	0.006915	COG0714	MoxR-like ATPases	R
Nest_275	15.41195	8.64E-05	0.00295	COG1611	Predicted Rossmann fold nucleotide-binding protein	R
Nest_479	12.12768	0.000497	0.009532	COG3428	Predicted membrane protein	S
Nest_821	11.99687	0.000533	0.009532	COG5650	Predicted integral membrane protein	S
Nest_900	11.46931	0.000708	0.011038	COG1307	Uncharacterized protein conserved in bacteria	S
Nest_131	17.53213	2.82E-05	0.00157	COG2170	Uncharacterized conserved protein	S
Nest_194	12.64861	0.000376	0.008215	COG1615	Uncharacterized conserved protein	S
Nest_251	7.93767	0.004842	0.037765	COG0779	Uncharacterized protein conserved in bacteria	S
Nest_29	9.92031	0.001635	0.018402	COG1814	Uncharacterized membrane protein	S
Nest_200	7.495402	0.006186	0.045032	*	*	S
Nest_855	8.1053	0.004414	0.037298	*	*	S
Nest_866	9.647934	0.001896	0.020295	*	*	S
Nest_888	7.384738	0.006578	0.047257	*	*	S
Nest_891	7.96486	0.004769	0.037469	*	*	S
Nest_917	11.30536	0.000773	0.011416	*	*	S
Nest_259	19.55059	9.8E-06	0.000823	COG3275	Putative regulator of cell autolysis	T
Nest_817	9.976974	0.001585	0.018122	*	*	T

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Nest_653	8.109882	0.004402	0.037298	COG0342	Preprotein translocase subunit SecD	U
Nest_1072	11.07823	0.000873	0.012387	COG0842	ABC-type multidrug transport system, permease component	V

Appendix T 4. List of differentially expressed genes at 5 ºC relative to 21 ºC in *Nesterenkonia* sp. AN1; columns represent Locus tag, Fold change, Bonferroni P-value, FDR, Gene, Description and COG number.

Locus_tag	Fold change	FDR	Gene name	Gene products	COG
Upregulated					
CELLULAR PROCESSES AND SIGNALING					
Cell cycle control, cell division, chromosome partitioning					
EXF25644.1	1.79	2.35E-08	*mrp*	ATPases involved in chromosome partitioning	COG0489
WP_036474664.1	2.57	1.51E-10	*ywqD*	ATPases involved in chromosome partitioning	COG0489
EXF25185.1	1.59	5.56E-09	*ftsZ*	Cell division GTPase	COG0206
WP_036476043.1	2.27	1.29E-17	-	Cell division protein	COG3115
Cell wall/membrane/envelope biogenesis					
EXF26067.1	1.85	7.98E-16	*yddH*	Cell wall-associated hydrolases (invasion-associated proteins)	COG0791
EXF25633.1	1.85	7.66E-09	*osmF*	Periplasmic glycine betaine/choline-binding (lipo)protein of an ABC-type transport system (osmoprotectant binding protein)	COG1732
EXF25429.1	1.97	8.40E-13	*ybjT*	Predicted nucleoside-diphosphate-sugar epimerases	COG0702
EXF25430.1	2.12	6.05E-28	*mpb83*	Secreted and surface protein containing fasciclin-like repeats	COG2335
EXF25351.1	1.90	5.65E-19	*betT*	Choline-glycine betaine transporter	COG1292
EXF23952.1	1.56	3.59E-07	*pbpG*	Membrane carboxypeptidase (penicillin-binding protein)	COG0744
EXF23918.1	2.00	8.37E-09	*wapA*	Rhs family protein	COG3209
EXF23827.1	3.08	5.68E-12	-	Sortase (surface protein transpeptidase)	COG3764
Defense mechanisms					
EXF26056.1	1.64	1.68E-05	-	Uncharacterized protein involved in methicillin resistance	COG2348

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Intracellular trafficking, secretion, and vesicular transport					
EXF26180.1	1.75	1.64E-11	*secY*	Preprotein translocase subunit SecY	COG0201
Posttranslational modification, protein turnover, chaperones					
EXF24480.1	2.06	1.15E-06	*bcp*	Peroxiredoxin	COG1225
EXF24910.1	1.90	9.73E-15	*ybbK*	Membrane protease subunits, stomatin/prohibitin homologs	COG0330
EXF24416.1	1.90	1.06E-15	*sufB*	ABC-type transport system involved in Fe-S cluster assembly, permease component	COG0719
EXF24497.1	1.81	1.97E-08	-	Thioredoxin domain-containing protein	COG3118
EXF25987.1	1.73	3.90E-05	*degP*	Trypsin-like serine proteases, typically periplasmic, contain C-terminal PDZ domain	COG0265
EXF26118.1	1.61	6.20E-06	*trxC*	Thioredoxin domain-containing protein	COG3118
EXF25584.1	1.51	9.34E-07	*clpC*	ATPases with chaperone activity, ATP-binding subunit	COG0542
EXF24911.1	1.86	4.59E-07	-	Membrane protein implicated in regulation of membrane protease activity	COG1585
EXF24478.1	1.74	1.50E-07	*clpP2*	Protease subunit of ATP-dependent Clp proteases	COG0740
Signal transduction mechanisms					
EXF25956.1	3.28	2.55E-07	*yxiE*	Universal stress protein UspA and related nucleotide-binding proteins	COG0589
EXF23991.1	17.20	1.09E-09	-	Universal stress protein UspA and related nucleotide-binding proteins	COG0589
WP_036473428.1	1.74	3.48E-06	*tcrY*	Signal transduction histidine kinase	COG0642
EXF26060.1	1.69	9.44E-07	*tcrX*	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	COG0745
EXF25955.1	3.14	1.94E-12	-	Universal stress protein UspA and related nucleotide-binding proteins	actNOG02640

EXF25908.1	1.73	2.48E-08	*ygeK*	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	COG2197
EXF24956.1	1.65	1.14E-04	-	Universal stress protein UspA and related nucleotide-binding proteins	COG0589
EXF24801.1	1.86	1.31E-10	-	FOG: FHA domain	COG1716
WP_036476411.1	2.01	1.02E-11	*odhI*	FOG: FHA domain	COG1716
EXF24277.1	1.57	6.45E-07	-	Universal stress protein UspA and related nucleotide-binding proteins	COG0589
EXF24043.1	2.81	2.18E-18	*rsbU*	Serine phosphatase RsbU, regulator of sigma subunit	COG2208
EXF23949.1	1.73	8.65E-11	*crp*	cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases	COG0664
EXF23992.1	2.79	1.86E-13	*dosT*	Signal transduction histidine kinase	COG4585
INFORMATIONSTORAGEANDPROCESSING					
Chromatin structure and dynamics					
EXF25802.1	2.16	5.36E-10	*hdaH*	Deacetylases, including yeast histone deacetylase and acetoin utilization protein	COG0123
Transcription					
EXF24043.1	2.81	2.18E-18	*rsbU*	Serine phosphatase RsbU, regulator of sigma subunit	COG2208
EXF26060.1	1.69	9.44E-07	*tcrX*	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	COG0745
EXF26117.1	1.68	1.44E-07	*ptsJ*	Transcriptional regulators containing a DNA-binding HTH domain	COG1167
EXF26020.1	1.51	1.01E-04	-	RNA Polymerase Sigma Factor Sigma-70 Family	actNOG12377
EXF25908.1	1.73	2.48E-08	*ygeK*	Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	COG2197

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

WP_036474496.1	1.60	3.47E-06	*sigK*	DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog	COG1595
EXF25256.1	2.37	4.21E-14	-	Predicted transcriptional regulators	COG1396
EXF25064.1	1.74	5.17E-09	*hrdB*	DNA-directed RNA polymerase, sigma subunit (sigma70/sigma32)	COG0568
EXF24743.1	1.87	2.50E-06	*cytR*	Transcriptional regulators	COG1609
Translation, ribosomal structure and biogenesis					
EXF26151.1	1.52	4.31E-06	*rplK*	Ribosomal protein L11	COG0080
EXF26158.1	1.57	1.53E-07	*rpsG*	Ribosomal protein S7	COG0049
WP_036473242.1	1.59	2.42E-08	*rplC*	Ribosomal protein L3	COG0087
EXF26001.1	1.81	1.02E-15	*yhbH*	Ribosome-associated protein Y (PSrp-1)	COG1544
EXF25352.1	1.82	1.19E-05	*rlmG*	16S RNA G1207 methylase RsmC	COG2813
WP_036475302.1	1.81	1.24E-14		Ribosomal protein S4 and related proteins	COG0522
EXF24513.1	1.64	1.03E-05	*prfA*	Protein chain release factor A	COG0216
METABOLISM					
Amino acid transport and metabolism					
EXF26117.1	1.68	1.44E-07	*ptsJ*	Transcriptional regulators containing a DNA-binding HTH domain	COG1167
WP_036473225.1	1.77	3.12E-06	*aspC*	Aspartate/tyrosine/aromatic aminotransferase	COG0436
EXF26229.1	2.25	3.87E-06	*dagA*	Na+/alanine symporter	COG1115
EXF25751.1	2.46	2.64E-17	*poxB*	Thiamine pyrophosphate-requiring enzymes [acetolactate synthase, pyruvate dehydrogenase (cytochrome)	COG0028
WP_036474538.1	1.67	1.87E-06	*ilvE*	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase	COG0115
EXF25724.1	1.70	9.94E-08		5-enolpyruvylshikimate-3-phosphate synthase	COG0128
EXF24993.1	2.21	3.04E-22	*glnA*	Glutamine synthetase	COG0174
EXF24848.1	2.55	1.46E-11	*glsA*	Glutaminase	COG2066
EXF24472.1	1.57	3.36E-05	*gcvH*	Glycine cleavage system H protein (lipoate-binding)	COG0509

EXF24150.1	1.62	7.20E-06	*yfbQ*	Aspartate/tyrosine/aromatic aminotransferase	COG0436
EXF24151.1	1.72	7.11E-11	*yfbQ*	Aspartate/tyrosine/aromatic aminotransferase	COG0436
EXF26083.1	2.50	3.93E-18	-	Na+/glutamate symporter	COG0786
EXF24018.1	2.46	4.02E-07	*bdhA*	Threonine dehydrogenase and related Zn-dependent dehydrogenases	COG1063
Carbohydrate transport and metabolism					
EXF26033.1	1.81	4.90E-11	*otsA*	Trehalose-6-phosphate synthase	COG0380
EXF26148.1	2.02	8.82E-09	*pfp*	6-phosphofructokinase	COG0205
EXF25375.1	1.57	5.07E-05	*glpF*	Glycerol uptake facilitator and related permeases (Major Intrinsic Protein Family)	COG0580
EXF25416.1	1.63	9.34E-06	*mcl2*	Citrate lyase beta subunit	COG2301
EXF25429.1	1.97	8.40E-13	*ybjT*	Predicted nucleoside-diphosphate-sugar epimerases	COG0702
EXF25469.1	3.31	4.57E-91	*nagD*	Predicted sugar phosphatases of the HAD superfamily	COG0647
EXF25511.1	1.99	8.10E-20	*fba*	Fructose/tagatose bisphosphate aldolase	COG0191
EXF24741.1	2.76	8.82E-18	*xylA*	Xylose isomerase	COG2115
EXF24551.1	1.57	4.64E-06	*ppnK*	Predicted sugar kinase	COG0061
EXF24359.1	2.02	2.03E-09	*glpX*	Fructose-1,6-bisphosphatase/sedoheptulose 1,7-bisphosphatase and related proteins	COG1494
EXF24201.1	2.49	7.80E-17	*prpB*	PEP phosphonomutase and related enzymes	COG2513
EXF25937.1	3.41	7.96E-27	*qacA*	Arabinose efflux permease	COG2814
EXF25330.1	1.86	2.79E-08	*pgk*	3-phosphoglycerate kinase	COG0126
Coenzyme transport and metabolism					
EXF26110.1	1.92	7.99E-06	*ubiE*	Methylase involved in ubiquinone/menaquinone biosynthesis	COG2226
EXF25977.1	1.66	2.41E-06	*abgT*	Putative p-aminobenzoyl-glutamate transporter	COG2978
EXF25739.1	2.74	4.69E-08	*mog*	Molybdopterin biosynthesis enzymes	COG0521
EXF25742.1	1.66	1.03E-04	*moaA*	Molybdenum cofactor biosynthesis enzyme	COG2896

EXF25751.1	2.46	2.64E-17	*poxB*	Thiamine pyrophosphate-requiring enzymes [acetolactate synthase, pyruvate dehydrogenase (cytochrome)	COG0028
WP_036474538.1	1.67	1.87E-06	*ilvE*	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase	COG0115
WP_036475297.1	1.50	1.44E-04	*lipA*	Lipoate synthase	COG0320
Energy production and conversion					
EXF26029.1	1.84	3.89E-07	*aceB*	Malate synthase	COG2225
EXF26095.1	1.73	1.04E-05	*ydiJ*	FAD/FMN-containing dehydrogenases	COG0277
EXF26243.1	3.69	3.49E-21	*icl*	Isocitrate lyase	COG2224
EXF26308.1	3.81	4.64E-06	*aceB*	Malate synthase	COG2225
EXF25860.1	2.20	3.92E-23	*gltA2*	Citrate synthase	COG0372
EXF25862.1	2.05	6.12E-11	-	Ferredoxin	COG1146
EXF25459.1	1.57	6.63E-06	*gabD*	NAD-dependent aldehyde dehydrogenases	COG1012
EXF26249.1	1.58	9.76E-07	*pyc*	Pyruvate carboxylase	COG1038
WP_036475437.1	1.61	9.62E-06	*adhC*	Zn-dependent alcohol dehydrogenases, class III	COG1062
EXF24647.1	1.57	9.42E-09	*ctaE*	Heme/copper-type cytochrome/quinol oxidase, subunit 3	COG1845
EXF24548.1	1.73	1.07E-07	*gltA*	Citrate synthase	COG0372
EXF24326.1	1.60	6.82E-07	*sdhB*	Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit	COG0479
EXF24327.1	1.78	5.39E-13	*sdhA*	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	COG1053
EXF24328.1	2.08	6.97E-15	-	Succinate dehydrogenase, hydrophobic anchor subunit	COG2142
EXF24329.1	2.03	5.16E-14	-	Succinate dehydrogenase/fumarate reductase, cytochrome b subunit	COG2009
EXF25374.1	1.59	6.77E-06	*glpK*	Glycerol kinase	COG0554
Inorganic ion transport and metabolism					
EXF24102.1	5.70	2.95E-05	*hcaC*	Ferredoxin subunits of nitrite reductase and ring-hydroxylating dioxygenases	COG2146

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

EXF25332.1	1.81	1.27E-12	*sodA*	Superoxide dismutase	COG0605
EXF24837.1	2.78	8.74E-05	-	Ion Transport 2 Domain-Containing Protein	aproNOG25208
EXF24295.1	2.36	2.54E-08	*fur*	Fe2+/Zn2+ uptake regulation proteins	COG0735
EXF24358.1	2.08	1.70E-06	*cynT*	Carbonic anhydrase	COG0288
EXF24046.1	1.58	4.82E-05	*ktrC*	K+ transport systems, NAD-binding component	COG0569
Lipid transport and metabolism					
EXF25091.1	1.72	9.00E-07	*tesB*	Acyl-CoA thioesterase	COG1946
Nucleotide transport and metabolism					
EXF26181.1	1.82	1.61E-08	*adk*	Adenylate kinase and related kinases	COG0563
EXF26237.1	1.66	9.56E-07	*guaB*	IMP dehydrogenase/GMP reductase	COG0516
WP_036475832.1	1.59	7.82E-06	*rutG*	Xanthine/uracil permeases	COG2233
EXF24660.1	2.09	6.52E-06	*pyrR*	Pyrimidine operon attenuation protein/uracil phosphoribosyltransferase	COG2065
EXF24335.1	1.77	2.60E-08	*add*	Adenosine deaminase	COG1816
WP_036477438.1	1.84	2.26E-10	*guaA*	GMP synthase, PP-ATPase domain/subunit	COG0519
Secondary metabolites biosynthesis, transport and catabolism					
EXF25802.1	2.16	5.36E-10	*hdaH*	Deacetylases, including yeast histone deacetylase and acetoin utilization protein	COG0123
EXF24465.1	1.78	3.28E-06	-	Uncharacterized protein Rv2286c/MT2344	actNOG01646
Function unknown					
EXF26251.1	1.85	3.19E-05	-	Hypothetical Protein Xcel	actNOG21102
EXF26271.1	1.96	1.95E-08	-	NUDIX Hydrolase	actNOG07785
EXF25960.1	1.58	1.07E-05	-	Uncharacterized conserved protein	COG3253
EXF25978.1	1.58	6.43E-05	-	Uncharacterized conserved protein	COG5282
EXF25786.1	3.20	1.07E-07	*yhaI*	Predicted membrane protein	COG3152
EXF25798.1	2.43	3.95E-05	-	Conserved Hypothetical Protein	actNOG15050
EXF25871.1	2.16	3.04E-12	*yfkH*	Predicted membrane protein	COG1295
EXF25643.1	2.48	1.76E-15	*ymgE*	Predicted membrane protein	COG2261
EXF25694.1	2.00	2.39E-11	*yuaG*	Uncharacterized protein conserved in bacteria	COG2268

WP_036474483.1	2.38	1.59E-06	*phnB*	Uncharacterized protein conserved in bacteria	COG2764
EXF25488.1	2.08	1.05E-05	-	Hypothetical Protein	actNOG36132
WP_036474701.1	3.15	1.45E-43	-	Predicted membrane protein	COG4129
EXF25276.1	2.95	2.45E-52	-	Hypothetical Protein Namu	actNOG21786
EXF25182.1	2.57	2.19E-27	*sepF*	Uncharacterized protein conserved in bacteria	COG1799
EXF25215.1	1.69	1.85E-10	-	Predicted membrane protein	COG2259
EXF25031.1	1.64	8.70E-05	-	Hypothetical Protein	actNOG23531
EXF25050.1	1.85	3.28E-06	-	Uncharacterized protein Rv0883c/MT0906	actNOG06454
EXF24733.1	1.87	1.47E-06	-	Uncharacterized conserved protein	COG5282
EXF24666.1	2.61	2.89E-12	-	Integration Host Factor	actNOG19712
EXF24429.1	1.52	1.21E-04	-	Conserved Hypothetical Protein	actNOG14524
EXF24097.1	2.32	1.03E-15	-	Polyketide Cyclase/Dehydrase	actNOG15323
WP_036477309.1	2.18	6.50E-15	-	Conserved Hypothetical Protein	actNOG09921
EXF23968.1	2.47	4.25E-07	*ycdC*	Uncharacterized protein conserved in bacteria	COG3226
EXF23915.1	2.66	7.62E-14	*yetF*	Predicted membrane protein	COG2323
EXF23828.1	2.25	2.16E-32	-	Hypothetical Protein Namu	actNOG21786
EXF23829.1	2.79	5.87E-08	-	Hypothetical Protein	actNOG34804
General function prediction only					
EXF25997.1	1.87	2.94E-05	*yeaC*	MoxR-like ATPases	COG0714
EXF25800.1	2.25	1.04E-15	*yfcH*	Predicted nucleoside-diphosphate sugar epimerase	COG1090
EXF25808.1	2.21	6.14E-05	*yrkH*	Zn-dependent hydrolases, including glyoxylases	COG0491
EXF25557.1	1.60	9.34E-06	*yqhE*	Aldo/keto reductases, related to diketogulonate reductase	COG0656
EXF25340.1	1.67	1.55E-06	*gph*	Predicted phosphatases	COG0546
EXF25135.1	1.52	8.97E-07	-	Predicted hydrolase of the metallo-beta-lactamase superfamily	COG0595
EXF25116.1	1.97	2.62E-08	-	Archaeal enzymes of ATP-grasp superfamily	COG1938
EXF24738.1	2.09	2.08E-16	*yigL*	Predicted hydrolases of the HAD superfamily	COG0561
EXF24858.1	2.56	3.94E-18	-	SurA Domain-Containing Protein	COG5271

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

EXF24464.1	1.65	5.37E-05	*ydfG*	Short-chain alcohol dehydrogenase of unknown specificity	COG4221
EXF24347.1	1.86	1.62E-10	*yhhT*	Predicted permease	COG0628
EXF24250.1	2.79	1.52E-10	*bpoA2*	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	COG0596
WP_036477178.1	2.07	4.60E-06	*entA*	Short-chain dehydrogenases of various substrate specificities	COG0300
EXF23990.1	7.69	7.47E-06	-	Predicted flavin-nucleotide-binding protein	COG3467
Uncharacterised					
WP_036474711.1	4.85	1.38E-04	-	Hypothetical Protein	
WP_051500849	1.70	1.40E-05	-	Tellurium resistance protein terD	
EXF25904.1	1.62	3.96E-06	-	Hypothetical Protein	
EXF26212.1	2.02	2.61E-12	-	Hypothetical Protein	
EXF25620.1	2.27	6.97E-15	-	Hypothetical Protein	
EXF25181.1	1.79	2.43E-07	-	Integral Membrane Protein	
EXF24898.1	2.20	1.40E-28	-	Hypothetical Protein	
EXF24386.1	2.74	5.89E-09	-	Conserved Hypothetical Protein	
Downregulated					
CELLULAR PROCESSES AND SIGNALING					
Cell wall/membrane/envelope biogenesis					
EXF25454.1	-1.84	3.03E-06	*pbpC*	Cell division protein FtsI/penicillin-binding protein 2	COG0768
EXF24408.1	-12.35	5.79E-12	*rfbP*	Sugar transferases involved in lipopolysaccharide synthesis	COG2148
Cell cycle control, cell division, chromosome partitioning					
EXF23873.1	-2.44	3.30E-19	-	ATPases involved in chromosome partitioning	COG1192
Coenzyme transport and metabolism					
EXF26263.1	-7.56	2.75E-07	*yhjG*	2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases	COG0654

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

EXF24964.1	-2.48	2.20E-09	*pcpB*	2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases	COG0654
Defense mechanisms					
EXF25958.1	-7.72	4.50E-22	*ywjA*	Uncharacterized ABC transporter ATP-binding protein YwjA	COG1132
EXF25041.1	-3.14	1.08E-05	*cydC*	ABC-type multidrug transport system, ATPase and permease components	COG1132
EXF24839.1	-1.64	6.04E-07	*bcrA*	ABC-type multidrug transport system, ATPase component	COG1131
EXF24320.1	-1.54	2.90E-06	*lolD*	ABC-type antimicrobial peptide transport system, ATPase component	COG1136
EXF23954.1	-2.01	9.40E-06	*ywjA*	ABC-type multidrug transport system, ATPase and permease components	COG1132
WP_036478099.1	-5.51	2.89E-76	-	ABC-type multidrug transport system, ATPase and permease components	COG1132
Intracellular trafficking, secretion, and vesicular transport					
EXF25280.1	-3.78	4.09E-05	*gspE*	Flp pilus assembly protein, ATPase CpaF	COG4962
Lipid transport and metabolism					
EXF24427.1	-1.61	3.98E-06	*acpP*	Acyl carrier protein	COG0236
Posttranslational modification, protein turnover, chaperones					
EXF24832.1	-2.11	7.72E-13	*trxA*	Thioredoxin domain-containing protein	COG3118
EXF24712.1	-107.78	2.30E-07	-	Cytochrome c biogenesis protein	COG0785
EXF23876.1	-7.54	4.64E-84	*cydC*	ABC-type transport system involved in cytochrome bd biosynthesis, ATPase and permease components	COG4988
Signal transduction mechanisms					
EXF24945.1	-1.76	2.60E-05	*relA*	Guanosine polyphosphate pyrophosphohydrolases/synthetases	COG0317
EXF24860.1	-2.21	1.34E-04	-	Universal stress protein UspA and related nucleotide-binding proteins	COG0589

EXF25522.1	-3.56	1.02E-04	*ulaC*	Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)	COG1762
EXF24206.1	-13.42	1.02E-26	-	Universal stress protein UspA and related nucleotide-binding proteins	COG0589
INFORMATIONSTORAGEANDPROCESSING					
Replication, recombination and repair					
EXF25970.1	-1.55	7.47E-05	*uvrD*	Superfamily I DNA and RNA helicases	COG0210
EXF25334.1	-4.04	4.93E-14	*uvrC*	Nuclease subunit of the excinuclease complex	COG0322
EXF25369.1	-2.55	1.05E-05		Transposase and inactivated derivatives	COG3547
EXF23881.1	-1.55	9.47E-05	-	Transposase and inactivated derivatives	COG4584
EXF23882.1	-1.83	2.57E-07	-	DNA replication protein	COG1484
EXF23883.1	-2.17	1.16E-04	-	Transposase and inactivated derivatives	COG3464
Transcription					
EXF24604.1	-1.71	1.03E-07	*lysG*	Transcriptional regulator	COG0583
EXF26073.1	-2.06	0.00E+00	*cspA*	Cold shock proteins	COG1278
EXF25926.1	-5.00	1.39E-09	*glpR*	Transcriptional regulators of sugar metabolism	COG1349
EXF25716.1	-6.02	1.14E-04	*feaR*	AraC-type DNA-binding domain-containing proteins	COG2207
EXF24945.1	-1.76	2.60E-05	*relA*	Guanosine polyphosphate pyrophosphohydrolases/synthetases	COG0317
EXF24740.1	-2.63	7.62E-06	*xylR*	Transcriptional regulator/sugar kinase	COG1940
EXF24365.1	-2.12	3.07E-09	*lytR*	Transcriptional regulator	COG1316
EXF24205.1	-3.46	5.81E-17	-	TetR Family Transcriptional Regulator	actNOG23547
Translation, ribosomal structure and biogenesis					
EXF25811.1	-1.90	1.81E-11	*rpsO*	Ribosomal protein S15P/S13E	COG0184
EXF25319.1	-1.60	1.08E-05	*rpmE2*	Ribosomal protein L31	COG0254
EXF24626.1	-1.55	3.69E-05	*rpsT*	Ribosomal protein S20	COG0268
EXF24208.1	-14.93	2.10E-32	*gatA*	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases	COG0154
EXF23878.1	-10.72	1.59E-32	-	Conserved Hypothetical Protein	actNOG00459

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

METABOLISM					
Amino acid transport and metabolism					
EXF26264.1	-1.83	1.67E-07	*gcvP*	Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain	COG1003
EXF26265.1	-2.05	4.22E-05	*gcvT*	Glycine cleavage system T protein (aminomethyltransferase)	COG0404
EXF24920.1	-3.39	1.37E-13	*sdaA*	L-serine deaminase	COG1760
EXF25867.1	-1.94	0.00E+00	*oppA*	ABC-type oligopeptide transport system, periplasmic component	COG4166
EXF25882.1	-3.38	3.11E-16	*yuxL*	Dipeptidyl aminopeptidases/acylaminoacyl-peptidases	COG1506
EXF24140.1	-7.86	2.79E-09	*betA*	Choline dehydrogenase and related flavoproteins	COG2303
EXF25443.1	-2.66	4.81E-05	-	3-Hydroxyacyl-CoA Dehydrogenase	
EXF25101.1	-1.55	7.25E-05	*dcp*	Zn-dependent oligopeptidases	COG0339
EXF24934.1	-2.05	1.13E-05	*proB*	Glutamate 5-kinase	COG0263
WP_036475604.1	-25.56	3.40E-37	*puuP*	Amino acid transporters	COG0531
EXF24986.1	-13.77	4.03E-30	-	Saccharopine dehydrogenase and related proteins	COG1748
EXF24460.1	-4.64	7.32E-13	*panF*	Na+/proline symporter	COG0591
EXF24526.1	-2.62	1.34E-06	*ywrD*	Gamma-glutamyltransferase	COG0405
EXF24140.1	-1.58	7.34E-06	*betA*	Choline dehydrogenase and related flavoproteins	COG2303
EXF24111.1	-10.63	5.00E-84	*puo*	Monoamine oxidase	COG1231
WP_036475604.1	-6.42	1.74E-27	*puuP*	Amino acid transporters	COG0531
EXF24207.1	-20.54	8.90E-27	*aguA*	Peptidylarginine deiminase and related enzymes	COG2957
EXF24209.1	-15.68	2.84E-25	*gabT*	4-aminobutyrate aminotransferase and related aminotransferases	COG0160
EXF23879.1	-7.84	1.94E-31	*ptrB*	Serine proteases of the peptidase family S9A	COG1505
EXF25774.1	-3.79	2.62E-05	-	ABC-type spermidine/putrescine transport system, permease component II	COG1177
Carbohydrate transport and metabolism					
EXF24762.1	-2.07	2.01E-12	*msmX*	ABC-type sugar transport systems, ATPase components	COG3839

EXF24087.1	-2.18	2.32E-06	*aglA*	Glycosidases	COG0366
EXF25924.1	-2.67	1.01E-08	*fruA*	Phosphotransferase system, fructose-specific IIC component	COG1299
EXF25925.1	-5.77	7.67E-10	*fruK*	Fructose-1-phosphate kinase and related fructose-6-phosphate kinase (PfkB)	COG1105
EXF25926.1	-5.00	1.39E-09	*glpR*	Transcriptional regulators of sugar metabolism	COG1349
EXF25384.1	-2.79	1.35E-20	*yhjE*	Arabinose efflux permease	COG2814
EXF25522.1	-3.56	1.02E-04	*ulaC*	Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)	COG1762
EXF25056.1	-5.96	8.72E-07		ABC-type sugar transport systems, permease components	COG1175
EXF25057.1	-18.19	5.64E-05	*mdxG*	ABC-type maltose transport systems, permease component	COG3833
EXF25055.1	-29.98	3.07E-09	-	Fructose-Bisphosphate Aldolase	actNOG02556
EXF25051.1	-7.18	2.81E-11	*malE*	Maltose-binding periplasmic proteins/domains	COG2182
EXF24740.1	-2.63	7.62E-06	*xylR*	Transcriptional regulator/sugar kinase	COG1940
EXF24789.1	-14.44	5.03E-19	*rbsA1*	ABC-type sugar transport system, ATPase component	COG1129
EXF24790.1	-25.28	1.28E-07	*rbsC*	Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components	COG1172
EXF24800.1	-2.92	4.60E-05	*araN*	ABC-type sugar transport system, periplasmic component	COG1653
EXF24090.1	-7.50	1.29E-06	-	ABC-type sugar transport system, periplasmic component	COG1653
Coenzyme transport and metabolism					
EXF26263.1	-7.56	2.75E-07	*yhjG*	2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases	COG0654
EXF25692.1	-2.60	7.08E-17	*paaK*	Coenzyme F390 synthetase	COG1541
EXF24932.1	-1.87	9.27E-07	*nadD*	Nicotinic acid mononucleotide adenylyltransferase	COG1057
EXF24964.1	-2.48	2.20E-09	*pcpB*	2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases	COG0654

WP_036478037.1	-4.47	8.62E-08	*ubiE*	Methylase involved in ubiquinone/menaquinone biosynthesis	COG2226
Energy production and conversion					
EXF26097.1	-1.54	7.27E-05	*ydeP*	Anaerobic dehydrogenases, typically selenocysteine-containing	COG0243
EXF25712.1	-3.76	1.86E-33	*paaE*	Flavodoxin reductases (ferredoxin-NADPH reductases) family 1	COG1018
EXF25717.1	-29.98	3.07E-09	*feaB*	NAD-dependent aldehyde dehydrogenases	COG1012
EXF25719.1	-1.91	1.16E-09	*maoC*	NAD-dependent aldehyde dehydrogenases	COG1012
EXF25720.1	-2.17	2.62E-05	*yhdH*	NADPH:quinone reductase and related Zn-dependent oxidoreductases	COG0604
EXF25723.1	-3.88	1.92E-09	*pdhB*	Pyruvate/2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, eukaryotic type, beta subunit	COG0022
EXF24450.1	-3.95	1.04E-71	*putA*	NAD-dependent aldehyde dehydrogenases	COG1012
EXF24967.1	-3.49	3.13E-12	*hpcC*	NAD-dependent aldehyde dehydrogenases	COG1012
EXF24988.1	-1.90	2.70E-10	*cat1*	Acetyl-CoA hydrolase	COG0427
EXF24603.1	-4.78	4.20E-56	*luxA*	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	COG2141
EXF24204.1	-13.23	5.00E-105	*prr*	NAD-dependent aldehyde dehydrogenases	COG1012
EXF23936.1	-7.48	3.60E-19	*gabD*	NAD-dependent aldehyde dehydrogenases	COG1012
EXF23876.1	-7.54	4.64E-84	*cydC*	ABC-type transport system involved in cytochrome bd biosynthesis, ATPase and permease components	COG4988
Inorganic ion transport and metabolism					
EXF26205.1	-2.56	4.10E-06	*yclQ*	ABC-type enterochelin transport system, periplasmic component	COG4607
EXF26207.1	-4.31	8.04E-05	*yclO*	ABC-type enterochelin transport system, permease component	COG4605
EXF25438.1	-2.64	1.61E-08	*viuB*	Siderophore-interacting protein	COG2375

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

EXF25439.1	-2.08	1.97E-08	*fecB*	ABC-type Fe3+-hydroxamate transport system, periplasmic component	COG0614
EXF25574.1	-3.04	1.34E-10	*pstS3*	ABC-type phosphate transport system, periplasmic component	COG0226
EXF25219.1	-2.25	2.43E-07	*yfeB*	ABC-type Mn/Zn transport systems, ATPase component	COG1121
WP_036476175.1	-5.51	1.06E-04	*tauC*	ABC-type nitrate/sulfonate/bicarbonate transport system, permease component	COG0600
EXF24686.1	-2.53	4.22E-06	*yvdB*	Sulfate permease and related transporters (MFS superfamily)	COG0659
EXF24284.1	-2.84	5.35E-06	*phoU1*	Phosphate uptake regulator	COG0704
EXF24144.1	-1.76	2.86E-05	*sdcS*	Di- and tricarboxylate transporters	COG0471
EXF24212.1	-3.78	4.85E-10	*ybaR*	Sulfate permease and related transporters (MFS superfamily)	COG0659
WP_036477590.1	-8.03	6.52E-06	*sir2*	Sulfite reductase, beta subunit (hemoprotein)	COG0155
EXF23985.1	-8.65	8.25E-19	*hmuU*	ABC-type Fe3+-siderophore transport system, permease component	COG0609
EXF23906.1	-3.31	4.71E-08	-	ABC-type Fe3+-hydroxamate transport system, periplasmic component	COG0614
Lipid transport and metabolism					
EXF26043.1	-2.56	3.62E-09	*ytcI*	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	COG0365
EXF26266.1	-5.97	4.17E-21	*yngJ*	Acyl-CoA dehydrogenases	COG1960
EXF25713.1	-3.50	4.56E-11	*ygfG*	Enoyl-CoA hydratase/carnithine racemase	COG1024
WP_036474437.1	-3.45	2.78E-10	*crt*	Enoyl-CoA hydratase/carnithine racemase	COG1024
EXF25380.1	-1.82	1.40E-04	*pcaF*	Acetyl-CoA acetyltransferase	COG0183
EXF25127.1	-3.21	5.49E-07	*mmgC*	Acyl-CoA dehydrogenases	COG1960
WP_036475261.1	-3.04	7.92E-06	*fadI*	Acetyl-CoA acetyltransferase	COG0183
EXF24426.1	-1.78	2.47E-08	*fabH*	3-oxoacyl-[acyl-carrier-protein] synthase III	COG0332
WP_036476387.1	-1.50	5.17E-05	*oatA*	Predicted acyltransferases	COG1835

WP_036476502.1	-3.44	2.59E-09	*prpE*	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	COG0365
Nucleotide transport and metabolism					
EXF24364.1	-3.24	1.16E-04	*purE*	Phosphoribosylcarboxyaminoimidazole (NCAIR) mutase	COG0041
EXF23933.1	-1.92	7.20E-06	-	Purine-cytosine permease and related proteins	COG1457
Secondary metabolites biosynthesis, transport and catabolism					
EXF25880.1	-3.14	1.37E-09	*ade2*	Imidazolonepropionase and related amidohydrolases	COG1228
EXF25709.1	-3.38	8.59E-09	*paaB*	Uncharacterized enzyme of phenylacetate metabolism	COG3460
EXF25714.1	-5.46	2.61E-21	*mco*	Putative multicopper oxidases	COG2132
EXF24966.1	-2.83	5.64E-07	*hpcG*	2-keto-4-pentenoate hydratase	COG3971
EXF24969.1	-3.35	6.85E-21	*hpaG*	2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase (catechol pathway)	COG0179
EXF24973.1	-2.54	2.43E-07	*tynA*	Cu2+-containing amine oxidase	COG3733
EXF24427.1	-1.61	3.98E-06	*acpP*	Acyl carrier protein	COG0236
POORLYCHARACTERIZED					
Function unknown					
EXF24602.1	-1.98	1.02E-11	-	Conserved Hypothetical Protein	
EXF25567.1	-2.05	1.87E-08		Uncharacterized protein conserved in bacteria	COG3861
EXF25949.1	-3.45	5.42E-22	-	Conserved Hypothetical Protein	actNOG19576
WP_036474354.1	-3.14	6.48E-48	*paaA*	Uncharacterized conserved protein	COG3396
EXF25710.1	-4.26	3.48E-25	*paaC*	Uncharacterized conserved protein	COG3396
WP_036475608.1	-4.99	5.66E-56	-	Uncharacterized protein conserved in bacteria	COG4405
EXF24989.1	-6.74	6.44E-24	-	Integral Membrane Protein	
EXF24601.1	-2.68	2.37E-21	-	Hypothetical Protein ABC	actNOG16955
EXF23877.1	-8.08	5.95E-67	-	Uncharacterized conserved protein	COG1944
EXF23870.1	-2.70	3.16E-05	-	TniB Family Protein	gproNOG15454

WP_036478139.1	-3.79	1.26E-05		Hypothetical protein	
General function prediction only					
EXF24591.1	-1.73	2.27E-07	*lhr*	Lhr-like helicases	COG1201
EXF25950.1	-2.39	2.82E-19	-	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	COG0596
EXF25781.1	-1.90	8.05E-07	*gsiA*	ATPase components of various ABC-type transport systems, contain duplicated ATPase	COG1123
EXF25701.1	-3.76	1.18E-04	*yiaN*	TRAP-type uncharacterized transport system, fused permease components	COG4666
EXF25711.1	-3.26	3.49E-08	*paaD*	Predicted metal-sulfur cluster biosynthetic enzyme	COG2151
EXF25381.1	-2.12	5.05E-05	*cysA*	ABC-type uncharacterized transport system, ATPase component	COG1101
EXF25260.1	-2.60	3.79E-13	*ygeK*	Predicted ATPase	COG3899
EXF24675.1	-5.30	3.42E-06	*comEC*	Predicted hydrolase (metallo-beta-lactamase superfamily)	COG2333
EXF24439.1	-1.72	1.42E-08	-	Predicted RNA-binding protein (contains KH domain)	COG1837
EXF24272.1	-15.40	8.40E-42	-	Predicted phage phi-C31 gp36 major capsid-like protein	COG4653
Uncharacterised					
EXF26312.1	-1.69	2.27E-06	-	Hypothetical Protein AARI	
EXF26218.1	-2.26	2.09E-13	-	Prevent-Host-Death Family Protein	actNOG21586
WP_036474537.1	-1.67	5.92E-06	-	Hypothetical Protein	
EXF26198.1	-2.72	1.66E-07	*bphC*	Manganese-dependent 2,3-dihydroxybiphenyl 1,2-dioxygenase	
EXF24990.1	-8.37	1.47E-58	-	Hypothetical Protein	
EXF24271.1	-9.00	4.08E-08	-	Hypothetical Protein Tpau	
EXF23874.1	-1.58	5.41E-06	-	Hypothetical Protein	
EXF23835.1	-2.32	1.29E-05	-	Transposase	

Appendix T 5. GOEAST annotation of differentially expressed genes associated with molecular functions (GO-term) in *Nesterenkonia* sp. AN1 grown at 5°C and 21°C. The columns from left-right contain the locus tag, Fold change, Bonferroni adjusted P-value, FDR, genes and gene names, respectively

Upregulated					
oxo-acid-lyase activity (GO:0016833)					
EXF26243.1	3.69	1.46E-19	3.49E-21	*icl*	Isocitratelyase
EXF25416.1	1.63	2.47E-03	9.34E-06	*mcl2*	Citrate lyase beta subunit
antioxidant activity (GO:0016209)					
EXF25332.1	1.81	1.14E-10	1.27E-12	*sodA*	Superoxide dismutase
EXF24480.1	2.06	2.42E-04	1.15E-06	*bcp*	Peroxiredoxin
EXF24250.1	2.79	1.73E-08	1.52E-10	*bpoA2*	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)
transferase activity, transferring nitrogenous groups (GO:0016769)					
WP_036473225.1	1.77	7.11E-04	3.12E-06	*aspC*	Aspartate/tyrosine/aromatic aminotransferase
WP_036474538.1	1.67	4.09E-04	1.87E-06	*ilvE*	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase
EXF24150.1	1.62	1.83E-03	7.20E-06	*yfbQ*	Aspartate/tyrosine/aromatic aminotransferase
EXF24151.1	1.72	7.75E-09	7.11E-11	*yfbQ*	Aspartate/tyrosine/aromatic aminotransferase
Downregulated					
oxidoreductase activity, acting on the CH-NH2 group of donors (GO:0016638)					
EXF26264.1	-1.83	2.98E-05	1.67E-07	*gcvP*	Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain
EXF24111.1	-10.63	3.00E-83	5.00E-84	*puo*	Monoamine oxidase
oxidoreductase activity, acting on the aldehyde or oxo group of donors (GO:0016903)					
EXF26097.1	-1.54	2.00E-02	7.27E-05	*ydeP*	Anaerobic dehydrogenases, typically selenocysteine-containing
EXF25717.1	-29.98	4.11E-07	3.07E-09	*feaB*	NAD-dependent aldehyde dehydrogenases
EXF25719.1	-1.91	1.44E-07	1.16E-09	*maoC*	NAD-dependent aldehyde dehydrogenases
EXF25723.1	-3.88	2.44E-07	1.92E-09	*pdhB*	Pyruvate/2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, eukaryotic type, beta subunit
EXF24967.1	-3.49	3.00E-10	3.13E-12	*hpcC*	NAD-dependent aldehyde dehydrogenases

EXF24204.1	-13.23	5.00E-105	1.67E-105	*prr*	NAD-dependent aldehyde dehydrogenases
EXF23936.1	-7.48	1.76E-17	3.60E-19	*gabD*	NAD-dependent aldehyde dehydrogenases
serine hydrolase activity (GO:0017171)					
EXF25882.1	-3.38	1.99E-14	3.11E-16	*yuxL*	Dipeptidyl aminopeptidases/acylaminoacyl-peptidases
EXF23879.1	-7.84	4.45E-30	1.94E-31	*ptrB*	Serine proteases of the peptidase family S9A
small molecule binding (GO:0036094)					
EXF25925.1	-5.77	9.36E-08	7.67E-10	*fruK*	Fructose-1-phosphate kinase and related fructose-6-phosphate kinase (PfkB)
EXF25781.1	-1.9	1.62E-04	8.05E-07	*gsiA*	ATPase components of various ABC-type transport systems, contain duplicated ATPase
EXF25381.1	-2.12	2.00E-02	5.05E-05	*cysA*	ABC-type uncharacterized transport system, ATPase component
EXF25280.1	-3.78	1.00E-02	4.09E-05	*gspE*	Flp pilus assembly protein, ATPase CpaF
EXF25219.1	-2.25	4.49E-05	2.43E-07	*yfeB*	ABC-type Mn/Zn transport systems, ATPase component
EXF24986.1	-13.77	1.01E-28	4.03E-30	-	Saccharopine dehydrogenase and related proteins
EXF24789.1	-14.44	2.52E-17	5.03E-19	*rbsA1*	ABC-type sugar transport system, ATPase component
EXF24839.1	-1.64	1.18E-04	6.04E-07	*bcrA*	ABC-type multidrug transport system, ATPase component
EXF24320.1	-1.54	6.51E-04	2.90E-06	*lolD*	ABC-type antimicrobial peptide transport system, ATPase component
EXF23882.1	-1.83	4.81E-05	2.57E-07	-	DNA replication protein
EXF23870.1	-2.7	9.26E-03	3.16E-05	-	TniB Family Protein
nucleoside phosphate binding (GO:1901265)					
EXF25925.1	-5.77	9.36E-08	7.67E-10	*fruK*	Fructose-1-phosphate kinase and related fructose-6-phosphate kinase (PfkB)
EXF25781.1	-1.9	1.62E-04	8.05E-07	*gsiA*	ATPase components of various ABC-type transport systems, contain duplicated ATPase
EXF25692.1	-2.6	4.32E-15	7.08E-17	*paaK*	Coenzyme F390 synthetase
EXF25381.1	-2.12	2.00E-02	5.05E-05	*cysA*	ABC-type uncharacterized transport system, ATPase component
EXF25280.1	-3.78	1.00E-02	4.09E-05	*gspE*	Flp pilus assembly protein, ATPase CpaF
EXF25219.1	-2.25	4.49E-05	2.43E-07	*yfeB*	ABC-type Mn/Zn transport systems, ATPase component
EXF24986.1	-13.77	1.01E-28	4.03E-30	-	Saccharopine dehydrogenase and related proteins
EXF24789.1	-14.44	2.52E-17	5.03E-19	*rbsA1*	ABC-type sugar transport system, ATPase component
EXF24839.1	-1.64	1.18E-04	6.04E-07	*bcrA*	ABC-type multidrug transport system, ATPase component

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

EXF24320.1	-1.54	6.51E-04	2.90E-06	*lolD*	ABC-type antimicrobial peptide transport system, ATPase component
WP_036478099.1	-5.51	2.02E-75	2.89E-76	-	ABC-type multidrug transport system, ATPase and permease components
EXF23882.1	-1.83	4.81E-05	2.57E-07	-	DNA replication protein
EXF23870.1	-2.7	9.26E-03	3.16E-05	-	TniB Family Protein