A Molecular and Phylogenetic Analysis of Cryobiosis in Nematodes of the Genus *Panagrolaimus*

Lorraine Marie McGill, B. Sc.

Thesis submitted to the National University of Ireland Maynooth in fulfilment of the requirements for the Degree of Doctor of Philosophy

NUI MAYNOOTH

Ollscoil na hÉireann Má Nuad

Department of Biology
National University of Ireland
Maynooth
Co. Kildare

Supervisor: Prof. Ann Burnell
Head of Department: Prof. Kay Ohlendieck

October 2011

Table of Contents

Abbreviations

A	Adenine
AFGP	Antifreeze glycoprotein
AFP	Antifreeze protein
Ala	Alanine
ANOVA	Analysis of variance
Asn	Asparagine
BSA	Bovine serum albumin
C	Cytosine
Ca^{2+}	Calcium
cDNA	Complementary DNA
COI	Cytochrome c oxidase subunit 1
COR	Cold-regulated
CRD	Carbohydrate recognition domain
CRT	C-repeat
CS	Chilling sensitive
Csp	Cold shock protein
CT	Chilling tolerant
Cys	Cysteine
ddH_20	Distilled deionised water
DE	Differential expression
DEPC	Diethylpyrocarbonate
dH_20	Distilled water
DJ	Dauer Juvenile
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DRE	Dehydration-responsive element
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionisation
EST	Expressed sequence tag
ETOH	Ethanol
G	Guanine
G6P	Glucose 6 phosphate
gDNA	Genomic DNA
Gly	Glycine
GPK	Glycogen phosphorylase kinase
GTR	General time reversible
h	Hour
HPLC	High performance liquid chromatography
HSF	Heat shock factor
Hsp	Heat shock protein
IBP	Ice-binding protein
INP	Ice-nucleation protein
IPTG	Isopropyl-beta-D-thiogalactopyranoside
ITS	Internal transcribed spacer
K_2HPO_4	Dipotassium phosphate
KEGG	Kyoto Encyclopedia of Genes and Genomes
KH_2PO_4	Monopotassium phosphate

L	Litre
LB	Liquid broth
LC-MS/MS	Liquid chromatography-mass spectrometry mass spectrometry
LEA	Late embryogenesis abundant
LRR	Leucine rich repeat
LSU	Large subunit
MCMC	Markov Chain Monte Carlo
ME	Minimum evolution
min	Minute
ml	Millilire
ML	Maximum likelihood
MS	Mass spectrometry
MYA	Million years ago
NaCl	Sodium Chloride
NADPH	Nicotinamide adenine dinucleotide phosphate
NaOCL	Sodium hypochlorite
NaOH	Sodium hydroxide
NGM	Nematode growth media
NJ	Neighbour joining
NMR	Nuclear magnetic resonance
NNI	Nearest Neighbor Interchange
^{o}C	Degrees celsius
PCR	Polymerase chain reaction
PGIP	Polygalacturonase inhibitor proteins
Phe	Phenylalanine
PP	Posterior probability
PP1	Protein phosphatase 1
PR	Pathogenesis-related
QPCR	Quantitative PCR
rDNA	Ribosomal DNA
RH	Relative humidity
RIP	Recrystallisation inhibition protein
RNA	Reoxyribonucleic acid
RNase	Ribonuclease
ROS	Reactive oxygen species
RPM	Revolutions per minute
rRNA	Ribosomal RNA
RT	Room temperature
SAS	Sialic acid synthase
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE	Standard error
Ser	Serine
sHsp	Small Hsp
SOD	Superoxide dismutase
sp.	Species
T	Thymine
T6PS	Trehalose 6 phosphate synthase
TAE	Tris Acetate EDTA buffer
TBR	Tree Bisection-Reconnection
TH	Thermal hysteresis

Thr	Threonine
TPP	Trehalose-phosphatase
Tris	Tris (hydroxy-methyl) amino methane
Trp	Tryptophan
U	Uracil
UDP	Uridine diphosphate
UDPG	Uridine diphosphoglucose
UV	Ultra violet
Val	Valine
μL	Microlitre

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to thank my supervisor, Prof. Ann Burnell, for her excellent guidance throughout the project, which was at times very challenging, and for her endless enthusiasm, encouragement and direction during the writing of this thesis and for helping me see this project through to its conclusion. I would also like to take this opportunity to thank Prof. Burnell for the opportunity to work on such an interesting project.

Many thanks to the members of the Nematode Genetics laboratory both past and present for creating such an enjoyable work environment. I would especially like to thank Dr. Bridget Culleton for all her advice and encouragement and most importantly for all the great dinners. Thanks also to Eoin Mulvihill for putting up with my demanding ways and sharing all your qPCR knowledge (Quantitative PCR instrumentation was funded by Science Foundation Ireland Grant No.: SFI/07/RFP/GEN/F571/EC07).

Thanks to all the technical staff past and present for their kind help throughout my PhD, especially Joe O' Sullivan for his help constructing the ice-binding purification apparatus. Thanks to all of the staff of the Biology Dept. for all their support over the years.

I would also like to thank Cindy Collins (Biology Dept.) for all her kind help for the mass spectrometry experiments, Dr. Davide Pisani for all his guidance with the molecular clock analysis and Dr. Chris Creevey (Animal and Grassland Research and Innovation Centre, Teagasc, Dunsany, Co. Meath) for all his suggestions and help with the RNA-seq analysis. Thank you to Prof. Peter Davies (Queen University, Canada) and everyone in the Davies lab for all of your AFP expertise and making my time there a very valuable and fun experience.

I would also like to acknowledge Science Foundation Ireland and the John and Pat Hume Scholarship for funding.

Thanks to all my friends for there continued support and companionship. A special thanks to all the Fencing crew especially Dr. Fergal Martin (Molecular Evolution Laboratory, TCD) for your friendship over the last 8 years and your help answering all my random Bioinformatics questions. A very special thank you goes to Paddy, for all you patience and support and computer expertise. I couldn't have done this without you! Thank you for always been there to cheer me up.

Finally I would like to thank my parents for being my inspiration and guidance and for their constant love and support. They were always there when needed and I am forever thankful to them.

Thank you all!

Declaration

This thesis has not been submitted in whole or in part, to this, or any other University for any degree, and is, except where stated the original work of the author.

Signed _____

Lorraine M McGill

Abstract

Many organisms are able to survive freezing temperatures through the development of biochemical and physiological adaptations. These biochemical adaptations may include the synthesis of proteins such as antifreeze proteins or cryoprotectants such as trehalose, the elimination of ice nucleators, and the expression of stress associated proteins (such as molecular chaperones, antioxidants, late embryogenesis abundant (LEA) proteins). Physiological adaptations include the ability to undergo cryoprotective dehydration. The molecular mechanisms underlying freezing stress tolerance are poorly understood. One of the main aims of this project was to employ phylogenetic, proteomic, and transcriptomic approaches to gain insights into the adaptations that aid the survival of the free- living cryotolerant nematode *Panagrolaimus superbus*.

Panagrolaimus sp. from temperate, subpolar, polar and continental geographic regions show a range of freezing ability from strains that show high survival upon direct exposure to -80 ^{o}C to those that are freezing sensitive. Acclimation significantly improves the freezing survival of temperate, subpolar, polar and continental *Panagrolaimus* strains and species. They also undergo anhydrobiosis and a correlation exists between desiccation tolerance and freezing tolerance. A phylogenetic study did not find a relationship between freezing phenotype, biogeography and phylogeny. The freezing and desiccation tolerance of ten new tropical *Panagrolaimus* strains was investigated. These strains are desiccation tolerant but not freezing tolerant, suggesting that freezing survival requires some specialised adaptations. A phylogenetic study of all the *Panagrolaimus* strains used in this study showed that the desiccation tolerant tropical *Panagrolaimus* strains are more divergent from the other strains and species in this study.

Protein extracts from freezing tolerant *Panagrolaimus* sp. can inhibit the growth of ice along specific planes of an ice crystal, resulting in hexagonal bipyrimidal ice crystals. This ice faceting capacity was considered most likely to be due to the presence of ice binding proteins. An ice affinity purification method was implemented to purify ice-binding proteins from *P. superbus*. Several proteins found to be enriched in the ice fraction were identified by mass spectrometry. As none of the identified

proteins was an obvious ice binding protein, it was not possible to determine whether these proteins had ice-binding protein activity.

The divergence times for five *Panagrolaimus* strains and species were estimated using the relaxed molecular clock approach. The *Panagrolaimus* sp. were found to have diverged from other nematodes 70.12 million years ago. The Antarctic nematode *P. davidi* diverged from its Californian sister species PS1579 approximately 17.18 million years ago and the Arctic nematode *P. superbus* diverged from its Pennsylvanian sister species *Panagrolaimus sp.* AF36 9.97 million years ago.

The genes that are differentially expressed in response to a period of cold acclimation were determined using the next generation sequencing method RNA-seq. A large number of novel genes were significantly up-regulated (P-value <0.01) including those involved in the oxidative stress response, transporting, membrane modification, metabolism, signalling and cytoskeleton remodelling.

Chapter I General Introduction

1.1 Nematodes

The Phylum Nematoda is an extremely successful and diverse group. There are approximately 26,000 described nematode species (Hugot et al., 2001) but estimates of the actual number in existence range from 40,000 to 10 million (Blaxter, 1998). Nematodes are also numerically abundant with up to millions of individuals per square meter. Many species of nematodes are parasitic, threatening the health of plants, animals and humans. There are also free-living species found in marine and freshwater sediments and soil ecosystems in high numbers. These nematodes feed on bacteria, fungi, and other nematodes.

1.1.1 Habitat and ecology

Free-living nematodes typically have a simple life cycle. The *Caenorhabditis elegans* life cycle comprises the following stages: embryogenesis (includes fertilisation and hatching), which normally lasts 14 h and has four larval stages, known as L1-L4. Each larval phase is separated by a moult before the final moult occurs to produce an adult nematode. The life cycle of *C. elegans* takes approximately 3 days. Many nematode species may also form a dauer juvenile (DJ) stage under stressful conditions such as over crowding, limited food and high temperatures (Riddle and Albert, 1997).

1.1.2 Nematode morphology

Nematodes are structurally simple organisms. Free-living and plant parasitic nematodes are typically only a few millimetres long, whereas animal parasitic nematodes tend to be larger. The largest nematode ever observed is *Placentonema gigantisma*, discovered in the placenta of a sperm whale which was 8 m long (Gubanov, 1951). In their morphology nematodes resemble a tube within a tube. The inner tube refers to the alimentary canal that extends from the mouth on the anterior to the anus located near the tail. The outer tube refers to the body wall. The space between the inner and outer tubes is occupied by a pseudocoelom. The pseudocoelom is filled with fluid that bathes the nematode tissues and serves as the transport system for oxygen, food and metabolites. Nematodes possess digestive, nervous, excretory

and reproductive systems but lack circulatory or respiratory systems, instead using diffusion for gas exchange.

1.1.3 Molecular taxonomy of the Phylum Nematoda

Species-level identification of nematodes has traditionally relied on detailed morphological analysis. This is very difficult and time-consuming, mainly because of the high phenotypic variability among populations, and requires specialist taxonomic expertise. As a consequence, despite their large abundance in the animal kingdom nematode species descriptions are relatively under-represented. In an effort to improve and standardise species identification in nematodes and several other taxa there has been a shift towards the use of molecular taxonomy. Here species are characterised based on molecular markers, a strategy known as a "DNA barcoding". The DNA barcode is the DNA sequence of a small region of the genome that should carry sufficient information for taxonomic identification and phylogenetics (Floyd *et al.*, 2002; Blaxter *et al.*, 2005). DNA barcoding has been used to perform surveys of nematodes (Floyd *et al.*, 2002; Bhadury *et al.*, 2006; Powers *et al.*, 2009) as well as tardigrades (Blaxter et al., 2004) and other meiofauna in terrestrial and marine habitats (Floyd *et al.*, 2002; Blaxter *et al.*, 2004; Bhadury *et al.*, 2006; Hamilton *et al.*, 2009; Powers *et al.*, 2009). Theoretically DNA barcoding should allow for a robust, rapid and high-throughput identification at a significantly lower cost than using morphological characters. The main difficulty in molecular taxonomy is to find a suitable gene that will give sufficient information for taxonomic identification and phylogenetics. The RNA-sequence of ribosomal small subunit, the large subunit, the internal transcribed spacer, and the mitochondrial cytochrome c oxidase subunit 1 have all been proposed as candidates for barcoding (Rogrigues Da Silva et al., 2010). These regions will be discussed in the following sections.

1.1.3.1 Ribosomal small subunit gene

The ribosomal small subunit (SSU) 18S rRNA gene (Figure 1.1) has proved to be very useful in species identification. Ribosomal RNA genes exist in high copy number in most eukaryotic cells, in *C. elegans* the rRNA unit is tandemly repeated approximately 55 times at the end of chromosome 1 (The *C. elegans* Sequencing Consortium, 1998). The SSU has conserved flanking regions providing ease of

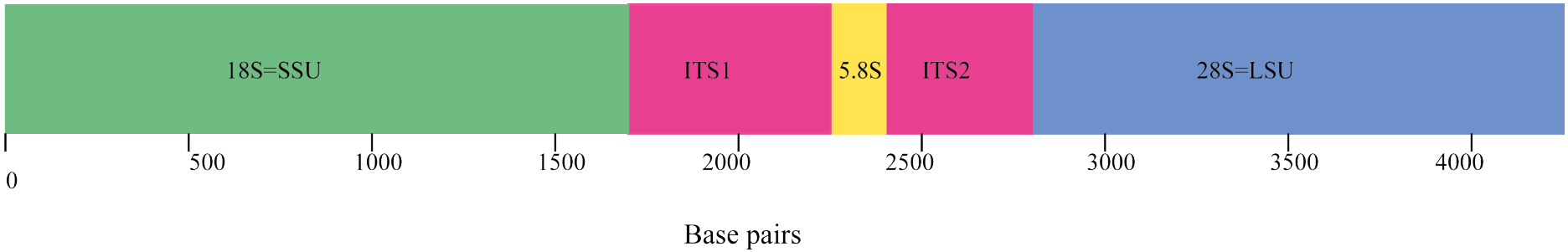


Figure1.1. Ribosomal DNA molecular markers and their relative sizes. The 18S gene encodes the small ribosomal subunit (SSU) and 28S gene encodes the large ribosomal subunit (LSU). The Internal Transcribed Spacer (ITS) region is located in the 18S, 5.8S and 28S ribosomal RNA genes.

amplification using the polymerase chain reaction (Floyd *et al.*, 2002; Blaxter *et al.*, 2005). The gene has sufficient semi-conserved regions that can give information on the deep phylogenetic relationships within the phylum while also having variable regions that enable species to be distinguished and identified. This has made the SSU rRNA the most widely utilised gene for barcoding. The high phylogenetic content of SSU rRNA, with its small amounts of polymorphism has worked well for resolving the relationships between different nematodes (Blaxter *et al.*, 1998; Holterman *et al.*, 2006; Holterman *et al.*, 2008a; van Megen *et al.*, 2009).

1.1.3.2 Ribosomal large subunit gene

The ribosomal large subunit (LSU) 28S rRNA gene (Figure 1.1), particularly the region which spans the D2 and D3 expansion segments, has been used for several years as a diagnostic marker sequence in nematodes (Rogrigues Da Silva et al., 2010). In nematodes, the D2/D3 region is 600-1000 bp in length and is located at the 5' end of the 28S rRNA gene. The D3/D3 segment contains flanking regions that are very conserved across the Phylum Nematoda allowing for selection of very robust PCR primers. As with other expansion regions of the LSU sequence divergence of D3 between related species is often high, so it also provides good separation of the more closely related groups of nematodes (de Ley et al., 1999). D2 or D3 sequences have been used in several phylogenetic studies, alone or in combination (Litvaitis et al., 2000).

1.1.3.3 rRNA internal transcribed spacer

The internal transcribed spacer (ITS) region is another nuclear genetic marker that has been used in nematode identification and phylogeny (Powers et al., 1997). The ITS1 region is located between the 18S of the SSU and the 5.8S rDNA genes. The ITS2 region is located between the 5.8S and the 28S of the LSU (Figure 1.1). In several nematode groups such as the marine nematodes the ITS region has not been considered a good marker for identification or phylogenetics (Rogrigues Da Silva et al., 2010). These spacer regions often have high intra-individual variation and large insertion and deletion (INDEL) events are frequently present in closely related cryptic species, this reduces the sequencing signal making multiple sequence alignments very difficult.

1.1.3.4 Mitochondrial cytochrome c oxidase subunit 1

The mitochondrial gene, cytochrome c oxidase subunit 1 (COI) has been widely used for barcoding animals. Mitochondrial genes are useful as barcodes since they may also provide further information on gene-flow patterns and the diversity among the nematodes. Unfortunately the COI gene has proved extremely difficult to amplify in nematodes (Bhadury et al., 2006). There are no phylum-wide universal primers for the COI gene, and PCR success rates are below 50% for several of the nematode species tested (De Ley et al., 2005). The problems associated with the COI gene are likely due to the emerging evidence that the nematode mitochondrial genomes are highly diverse and have unusual properties such as recombination, insertion editing and multipartitioning (Rogrigues Da Silva et al., 2010).

1.1.4 *Panagrolaimus* sp.

Pangarolaimus species are free-living and inhabit the soils of diverse geographical regions including temperate, tropical, arid and polar regions. These nematodes belong to the order Panagrolaimidae (Figure 1.2, clade IVa). *Pangarolaimus* species are bacteriovores and can be readily cultured in the laboratory on a lawn of *Escherichia coli.* The life cycle of these nematodes is completed in approximately 8 days at 20 ^{o}C. Members of the genus *Pangarolaimus* exhibit different reproductive modes including gonochoristic, hermaphroditic and parthenogenetic (Lewis et al., 2009). Various *Pangarolaimus* species are capable of anhydrobiosis meaning they can survive 'life without water' (Shannon et al., 2005). In their work Shannon *et al.* (2005) investigated the desiccation tolerance of a range of *Pangarolaimus* species isolated from the soil of desiccation prone environments and found that many of these species were capable of anhydrobiosis. Within the genus *Pangarolaimus* they found that slow and fast dehydration strategists can exist. *P. superbus* is a fast dehydration strategist that is capable of surviving immediate exposure to 0% relative humidity (RH) while PS1159 is an example of a slow dehydration strategist that requires a 24 h preconditioning period of 98% RH prior to exposure to 0% RH (Shannon et al., 2005). Freezing tolerance has also been found in the genus *Panagrolaimus.* The Antarctic nematode *P. davidi* can survive internal ice formation and temperatures down to -80 ^{o}C (Wharton and Brown, 1991; Wharton and Ferns, 1995). The objective of this project was to compare the relationship between biogeography, phylogeny and

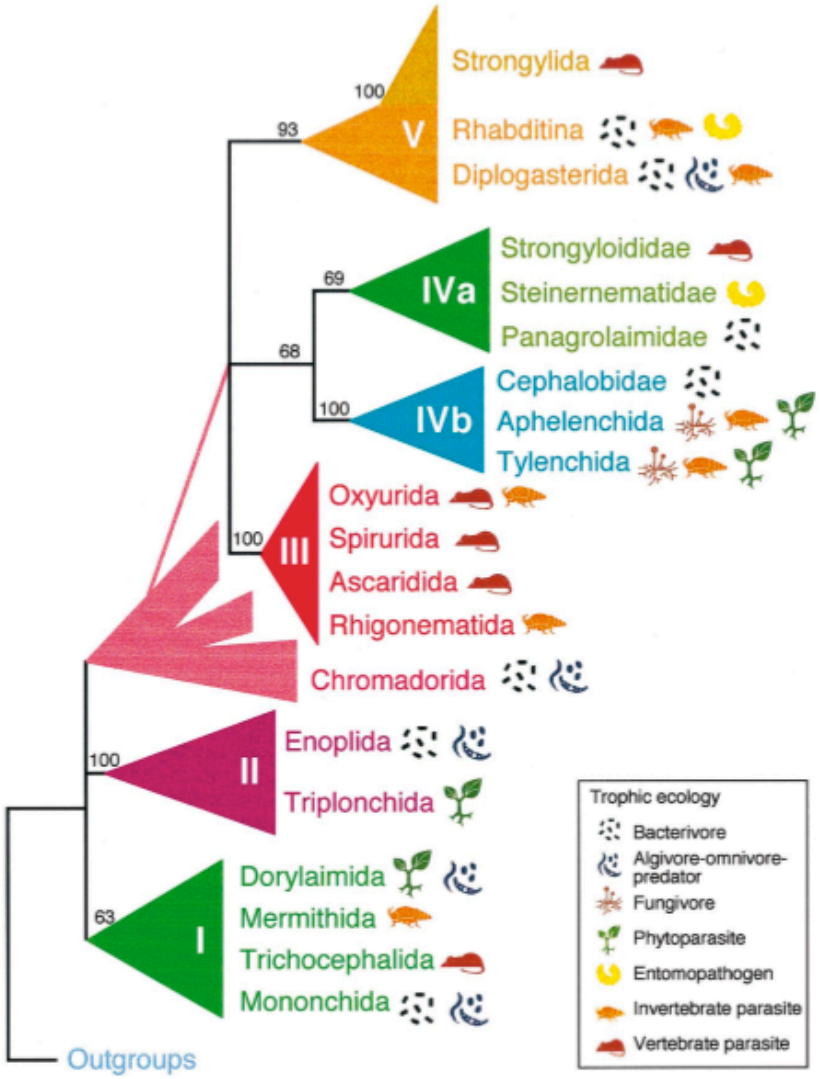


Figure 1.2. The phylogenetic structure of the Phylum Nematoda showing the clade groupings. This maximum parsimony tree was constructed using full-length sequences of the SSU rDNA molecule. The icons on the right indicate the trophic ecology of each clade. The source of this illustration is Dorris *et al.* (1999).

the cryotolerant and anhydrobiotic phenotypes of these nematodes and to get a better understanding of the molecular mechanisms underlying these phenotypes.

1.2 Phylogenetics

Phylogenetics is the science of reconstructing the evolutionary relationships between taxa, in the case of molecular phylogeny this is based on the comparison between DNA or protein sequences. Comparing several sequences along their lengths can show the parts that are evolving rapidly and which sequences show evidence of being shaped by natural selection. These comparative analyses rely on a phylogenetic tree that describes the evolutionary relationships between the sequences. The types of applications for phylogenetics include the detection of orthology and paralogy, estimates of divergence times, reconstructing ancestral proteins, finding sequences that are important in natural selection, detecting recombination points, identifying mutations associated with disease and determining the identity of new species (Holder and Lewis, 2003).

1.2.1 Methods used in phylogenetic studies

The main techniques used for constructing phylogenetics trees are distance, minimum evolution parsimony, likelihood and Bayesian frameworks and algorithms. The distance based method Neighbour-Joining (NJ) algorithm is an extremely popular method due to its speed (Saitou and Nei, 1987). It performs well when the divergence between sequences is low. The DNA or protein sequences are converted into a distance matrix that represents the evolutionary distances between sequences. The major disadvantage of this method is that the observed differences are not an accurate reflection of the evolutionary distances between them because multiple substitutions at the same site can make sequences seem artificially close to each other. Therefore it is not a reliable method for divergent sequences or when inferring older relationships. However NJ trees may often be treated as the starting point for a computationally intensive search for the best phylogeny. Minimum evolution (ME) methods, like the NJ approach compresses the data into a distance matrix so it is also has problems with accurately estimating the relationship between divergent sequences (Rzhetsky and Nei, 1992). Parsimony is a method that assumes that shared mutations between sequences must mean they have a common ancestor. It builds a tree on the basis of these shared mutations and assigns a score to each tree. The score is the minimum

number of mutations that could possibly produce the data (Fitch, 1970). However judging the best tree by the minimum number of mutations may not truly reflect the evolutionary history of a clade. Parsimony also fails to take into account the fact that the number of mutations is unlikely to be equal on all branches in a tree and it can perform poorly if there is substantial variation in branch lengths (Felsentein, 1978; Kim, 1996). For the reasons outlined above NJ, ME and parsimony methods were not used in this study. The Maximum Likelihood and Bayesian methods were used in this study; these can accurately reconstruct the relationships between divergent or rapidly evolving sequences and are described in detail in the following sections.

1.2.1.1 Maximum likelihood methods

Maximum likelihood (ML) is long established statistical method (Edwards, 1972) and it is considered the most statistically robust method to infer phylogenetic relationships. The likelihood value *L,* used in phylogenetic inference is the probability of observing the data under a given phylogenetic tree and a specific model of evolving. The tree describes the topology of the evolutionary relationships between the sequences, and a set of branch lengths estimate how evolution has occurred in different regions of the tree. The model contains the parameters that describe the pattern and processes of evolution, for example, the rate of mutation. In ML methods the optimal topology is that with the highest likelihood, and finding this requires calculating the likelihood of all topologies individually. In practice, even for a small number of sequences this is computationally impractical. Heuristic searches for the ML tree are widely used but they are not certain to find the optimal topology. With this approach, the hill-climbing optimisation techniques such as Nearest Neighbour Interchange (NNI) or Tree Bisection-Reconnection (TBR) are used. These algorithms function by making small rearrangements from a given candidate tree to make new topologies. The likelihood of the new topologies is calculated and the highest value is selected for generating the next set of topologies. This process is repeated until no further improvements in topology can be found (reviewed by Kosiol *et al.,* 2006). In this study the ML method was performed using the MEGA 5 program (Tamura et al., 2011). The bootstrap values used in this analysis are a method for testing the reliability of the observed tree and can be defined as the frequency of which a branch appears in the dataset (Felsenstein, 1985).

1.2.1.2 Bayesian methods

The Bayesian method for inferring phylogeny is a relatively new and computationally intensive method, which is growing in popularity due to its relative speed compared to the ML method and because of its ability to produce reliable results (Yang and Rannala, 1997; Huelsenbeck and Ronquist, 2001; Huelsenbeck and Rannala, 2004). The Bayesian method is based upon a quantity called the posterior probability distribution of trees, which is the probability of a tree given the observations (alignments of sequences) (Huelsenbeck and Rannala, 2004). This is accomplished using Bayes' theorem. As the posterior probability is impossible to calculate by analytical methods, phylogenetic programs that employ Bayesian approaches use a mathematical simulation technique known as Markov Chain Monte Carlo (MCMC) to approximate the posterior probabilities along with a user selected model of evolution (Ronquist and Huelsenbeck, 2003). This method results in the creation of a series of trees that are represented using a consensus tree. This consensus tree contains information on branch length and support values known as posterior probabilities (PP). PP values are the probability the tree is accurate, assuming the specified model is correct (Huelsenbeck and Rannala, 2004). These values are similar to bootstrap value supports, although they are not usually considered to be as stringent as bootstrap values (Simmons et al., 2004).

1.2.2 The fossil record and evolutionary timescales

In the 1790s George Cuvier in France and William Smith in England first showed that there is a link between the order of rocks and fossils and the history of the earth. The older rocks with the older fossils occur in the deeper layers with the younger rocks with more recent fossils lie in successive layers on top of these. Specifically, they noticed that certain fossils appeared to characterise particular rock units and they occurred in the same order everywhere they were seen. Geologists then used this observation to correlate rock units from place to place, to name them (e.g. Carboniferous) and to match stratigraphic units on the first geological maps. Around the 1800s the "father of English geology" William Smith, mapped correlatable geological units across England and hypothesised that these units extended over Europe. Robert Murchison and others in the 1830s drove these ideas forward and named all of the major divisions of geological time: the Palaeozoic, Mesozoic, and the Cenozoic eras and all of the periods within these. This timescale proved to be

applicable across the world and since 1840 this international geologic timescale has not been substantially revised.

A link between the fossil record in the rocks and organisational phylogeny could not be made until the concept of evolution could be understood. Charles Darwin was the first to understand that the evolution of life was a branching tree linking all living and fossil species and that all species are united backwards in time to a single common ancestor. His first branching tree was the only illustration in *The Origin of Species* (Darwin, 1859). In the early to mid-twentieth century, palaeontologists used trees to calculate the rates of evolution, although they had uncertain timescales but were still able to calculate rates in a relative way, as shown by George Gaylord Simpson in his work, *Tempo and Mode of Evolution* (Simpson, 1944). Since then a wide range of methods have been developed to study the relationships between species, determine their common ancestors and their diversification over time. The molecular clock approach is a common tool used in these studies. This method relates the number of fixed mutations or substitutions in nucleotide or amino acid sequences to divergence time of taxa. The fossil evidence may allow the rate of evolution to be calibrated. The introduction of the molecular clock concept was originally by Zuckerkandl and Pauling (Zuckerkandl and Pauling, 1962). They found that amino acid differences in mammalian α and β chains of haemoglobin were roughly proportional to divergence times inferred from paleontological data. This work was published in a landmark paper where they introduced the term molecular clock, describing its stochastic nature as a Poisson process (Zuckerkandl and Pauling, 1965).

1.2.2.1 Molecular clock models

Widely used methods for estimating divergence times assume that all evolutionary lineages experience sequence change at an identical rate (Figure 1.3 (i)) (Zuckerkandl and Pauling, 1962, 1965). Since the rate at which sequences change depends on a variety of factors such as natural selection, population size, generation time and mutation pattern it seems biologically unrealistic that these factors are identical across all evolutionary lineages. This “strict molecular clock” can give misleading divergence times and phylogenetic inferences. Problems with the strict molecular

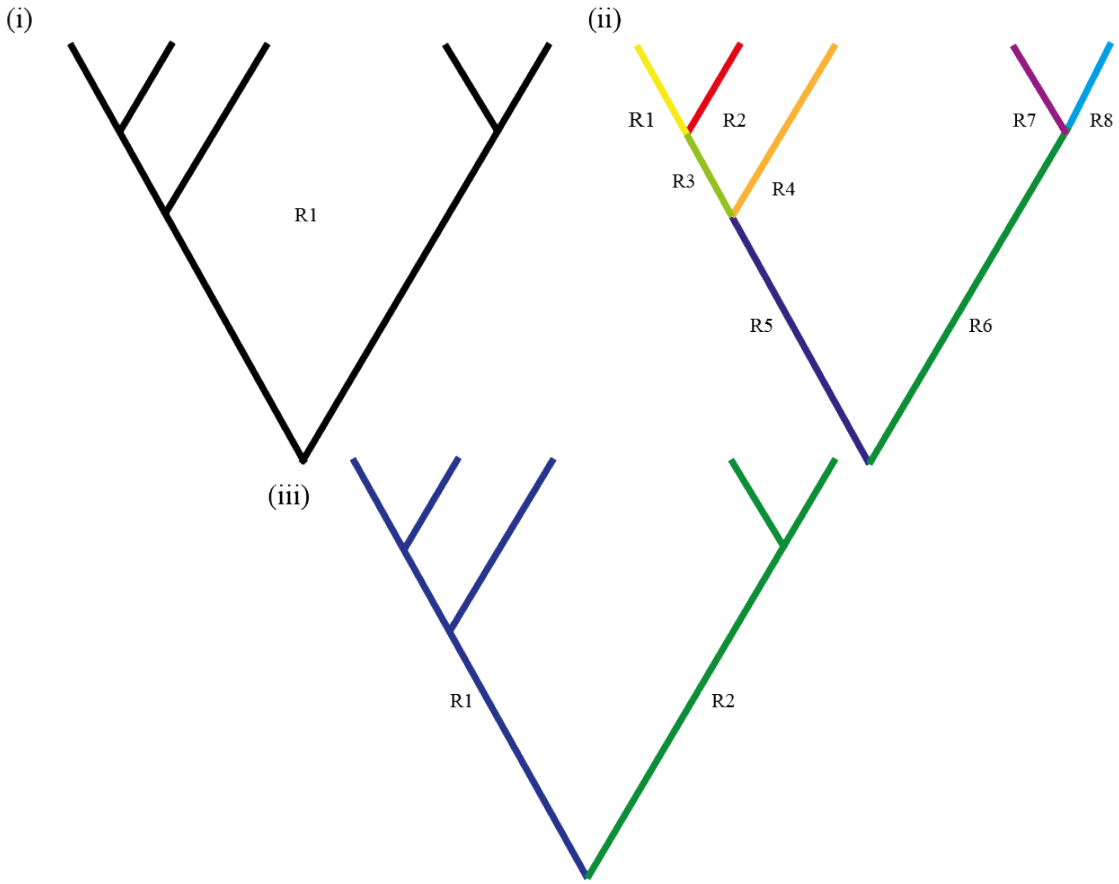


Figure 1.3. Phylogenies with different molecular clock models based on rate heterogeneity. (i) Strict molecular clock tree with a single rate (R1). (ii) Relaxed clock tree with different rates for different branches (R1-R6). (iii) Molecular clock tree with two different rates (local clocks; R1, R2). This image was modified from Wilke *et al.* (2009).

clock have resulted in it being almost entirely abandoned in favour of "relaxed molecular clock" models. These models assume that every branch has an independent heterogeneous rate of molecular evolution (Figure 1.3 (ii)). Using a form of the relaxed clock model, lineages within a clade may be allowed to share the same evolutionary rates since closely related evolutionary lineages typically evolve at similar rates forming local clocks (Figure 1.3 (iii)). These molecular clock models are described as being "autocorrelated" and include the lognormal (Kishino et al., 2001) and the "CIR" processes (Lepage et al., 2007). The CIR model is a continuous Markov process that is widely used in finance to model interest rates and it is named after its creators Cox, Ingersoll and Ross (1985). Others have rejected the assumption that closely related evolutionary lineages typically evolve at similar rates. They recommend methods that allow independent rates at each point, and then evaluate whether the reconstructed rates are autocorrelated. These methods are known as "uncorrelated" and include the uncorrelated gamma model, Ugamma (Drummond et al., 2006). Both these methods will be used in this study.

1.2.2.2 Molecular clock calibration

In order to calculate the divergence times within a given phylogeny, calibration points within the phylogeny are required. The known dates of the origin of ancestral DNA/RNA extracted from preserved material, fossils, or biogeographical events are used as calibration points for estimating rates of evolution in a given phylogeny. It is still debated whether an analysis based on many genes and few dates or few genes and many dates is preferable. In relaxed clock methods multiple calibration points throughout the tree are useful as they act as anchor points, allowing the method to estimate the rate variation more accurately. This allows divergence time to be calculated between or beyond calibration points to other parts of the tree that may be poorly calibrated (Benton and Donoghue, 2007).

Ideally, ancestral DNA/RNA of known age would be used as calibration points. However this is difficult to obtain and ancestral DNA/RNA is often too recent to estimate older phylogenetic events. Fossil and biogeographical data are most often used as the source of calibration dates. There are problems with the use of fossil data. The fossil record is largely incomplete; there are many gaps, with several species and phyla including nematodes poorly preserved. The fossil record also does not provide

actual age estimates for divergence events, but precise minimum constraints and looser maximum constraints that may be used for the calibration of molecular clocks (Benton and Donoghue, 2007). The minimum constraints on calibration are the oldest fossils belonging to a crown clade, these may be calculated with relative precision and may be treated as hard bounds (because the oldest fossil in a given clade will always be younger than the origin of this clade). The maximum constraints are soft bounds that may be represented by probability distributions that reflect the distribution of fossilliferous rocks around the same time (Benton et al., 2009) (Figure 1.4). According to Benton *et al*. (2007) "an older fossil deposit that ought to contain fossils of the clade in question, but does not, can mark an ultimate maximum bound". The accuracy of the fossil data can significantly affect the phylogeny so they must be selected with great care.

Calibration dates may be based on biogeographical data, but the phylogenetic event must be associated with a major biogeographical event. For example the divergence of the Pacific/Caribbean germinate species is associated with the closure of the Isthmus of Panama approximately 3.0-2.5 million years ago (MYA). But this date may not be an accurate estimate as it possible that some species separated before the Isthmus closed or gene flow may have continued after the closure of the Isthmus (Knowlton and Weigt, 1998). The other problem with the use of biogeographical data is that the timings of the major events may not be certain. Determining the calibration points for molecular clocks require a specific assessment of a geographical event in the context of the biology of the species.

1.2.3 Nematode evolution

Their diversity and the absence of a fossil record have made it difficult to define a clear evolutionary framework for the Phylum Nematoda. Traditionally the classification of the nematodes has relied on morphological traits. Traits most commonly used are buccal and pharyngeal structures, but the cuticle, lip region, intestine, reproductive system, sense organs and tail are also used, in addition to traits such as parasitic hosts (Dorris et al., 1999). However observational identification of nematode species is difficult, leading to bias and errors. In recent years, phylogenetic methods using the small and large subunit (SSU and LSU) ribosomal RNA genes described in Section 1.1.3 have reliably resolved the internal relationships within the

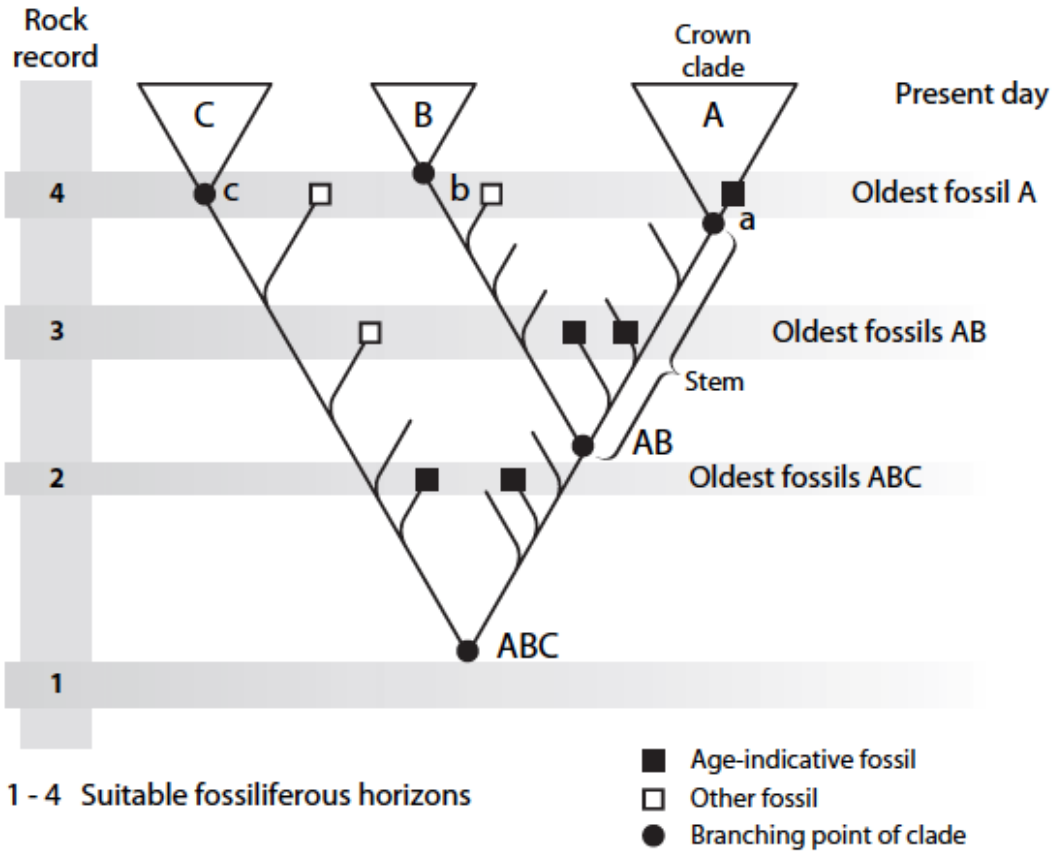


Figure 1.4. Definition of terms used in assigning fossils to clades. The crown clade consists of all living species and their most recent common ancestor and any other fossils located within this grouping. The crown clade is preceded by an extinct stem lineage of forms that are closer to their crown clade than to another crown clade. Fossiliferous horizon 1 that contains no fossils assignable to the clade ABC marks the maximum constraint (soft bound) on the age of the clade; similarly fossiliferous horizon 2 marks a maximum constraint for the clade AB. The minimum constraint on the age of a clade is the age of the geological formation that contains the oldest phylogenetically secure fossil belonging to that clade. Hence fossiliferous horizon 2 marks a minimum constraint for the clade ABC, fossiliferous horizon 3 marks a minimum constraint for clade AB and fossiliferous horizon 4 marks a minimum constraint for clade A. The source of this illustration is Benton and Donoghue (2007).

Phylum Nematoda (Figure 1.2) (Blaxter *et al.*, 1998; Meldal *et al.*, 2007; Holterman *et al.*, 2008b).

Two nematode classes have traditionally been recognised, the predominantly marine Adenophorea and terrestrial Secernentea (Chitwood, 1933). A bioinformatic analysis of the SSU ribosomal RNAs has shown no evidence of an adenophorean ancestry (Blaxter et al., 1998). These authors have identified five major nematode clades as shown in Figure 1.2. Clades I and II are purely adenophorean. Clade I includes the Trichocephalida (*Trichinella* and *Trichuris*) and the Mermithida (insect parasites). Three major clades (III-V) are found within the Secernentea. These clades do not support the classic divisions where the taxa are grouped by trophic ecology. Rather, the plant and animal-parasitic orders are integrated within the free-living groups (Dorris et al., 1999). Clade III contains animal parasites including *Brugia malayi* whose genome has been sequenced (Ghedin et al., 2007). Clade IV can be split into two subclades IVa and IVb (Dorris et al., 1999). This clade comprises animal and plant parasites as well as bacterivores and fungivores. The Panagrolaimidae nematodes used in this study belong to the Strongyloididae within clade IVa of the Blaxter *et al.* (1998) phylogeny. Clade V contains free-living, microbivores and vertebrate parasitic members. The free-living model organism *C. elegans* is a member of clade V. A more recent study with a larger dataset (Holterman et al., 2006) confirms the phylogenetic relationships described by Blaxter *et al.* (1998) but has subdivided the nematodes into 12 clades with the most basal being Enoplia (Holterman *et al.*, 2006; van Megen *et al.*, 2009). Holterman *et al.* (2006) also observed that a small number of morphological and physiological adaptations are required for major habitat transitions in groups of nematodes with various taxonomic levels.

1.3 Freezing survival

Temperature is one of the most important factors that determines the activity and distribution of living organisms. The majority of organisms will experience temperatures below their optimum temperature. Outside of the tropics and temperate waters the temperature falls below 0 ^{o}C on a seasonal and occasional basis (Figure 1.5). Low temperatures and freezing conditions affect the lives of all organisms in

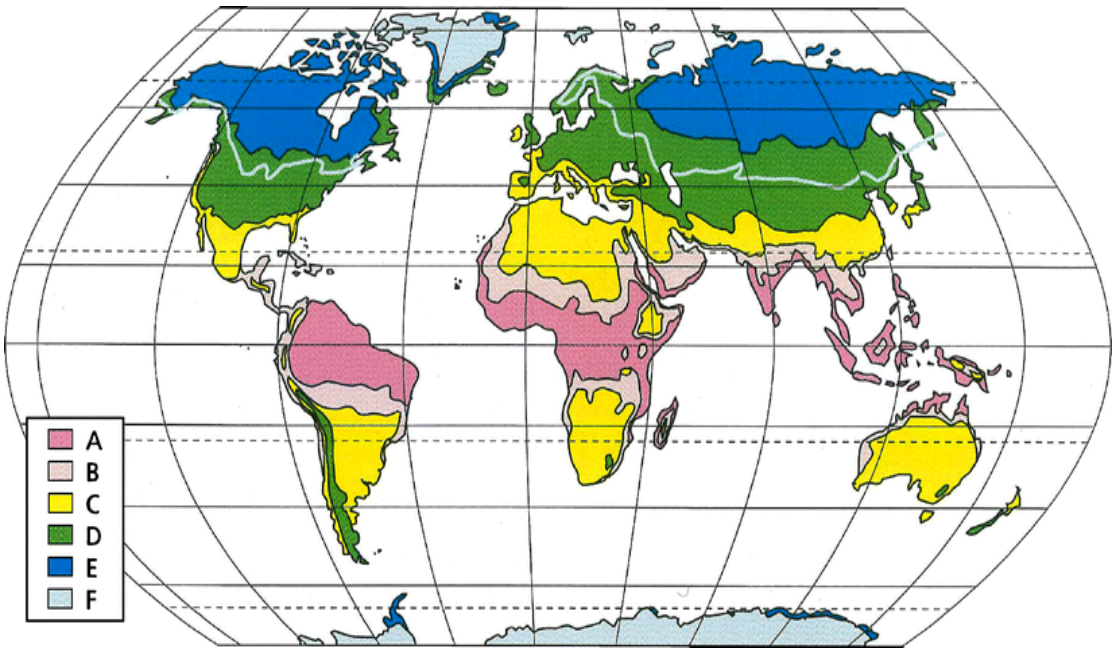


Figure 1.5. The occurrence of low temperatures and frost on the Earth. *A* annual minimum temperatures above +5 °C, *B* annual minimum temperatures above 0 °C, *C* episodic frost with temperatures down to -10 °C, *D* regions with cold winters and mean annual minimum temperatures between -10 and -40 °C (*white lines* -30 °C minimum isotherm), *E* mean annual minimum temperatures below -40 °C, F polar ice. The source of this illustration is Margesin *et al.* (2007).

multiple ways. They may reduce the rates of biochemical reactions, cause changes in the cell membrane fluidity and protein conformation, reduce nutrient availability or hinder successful reproduction (Margesin et al., 2007). To escape or cope with these low temperatures, microbes, plants and animals have evolved many physiological and behavioural strategies. Some species such as the monarch butterflies undertake a long distance migratory flight to escape winter (Dingle, 1996). Others seek refuge by spending winter in thermally buffered microclimates such as underwater, underground or under a snowpack, eliminating the probability of encountering freezing temperatures (Storey and Storey, 1996). The insulation of the snowpack can maintain the temperature under the snow close to 0 ^{o}C. Organisms that cannot escape the cold have developed biochemical and physiological adaptations to survive. Some examples of organisms that survive exposure to low temperatures are shown in Figure 1.6.

1.3.1 Definition and occurrence of freezing-adapted organisms

1.3.1.1 Microorganisms

Cold-adapted microbes may be divided into two groups, the psychrophilic (cold-loving) and the psychrotolerant (cold-tolerant). Psychrophiles can be defined as microbes whose cardinal growth temperatures (i.e. minimum, optimum, and maximum) are at or below 0, 15 and 20 ^{o}C respectively (Morita, 1975). They predominantly exist in permanently cold habitats, such as the polar regions, high altitudes and deep in the sea. Psychrotolerant microbes grow over a much wider temperature range and grow fastest above 20 ^{o}C (Margesin et al., 2007), making them favourable to environments that may have periodic or seasonal temperature fluctuations. A wide diversity of microbes occur in the cold ecosystems. Bacteria dominate, and are found in great abundance in the polar environment while the Archaea are widespread in cold, deep ocean water (Karner et al., 2001).

1.3.1.2 Plants

The ability to tolerate low temperatures and freezing is essential for the recruitment, survival, productivity and distribution of plants. Based on their response to cold plants

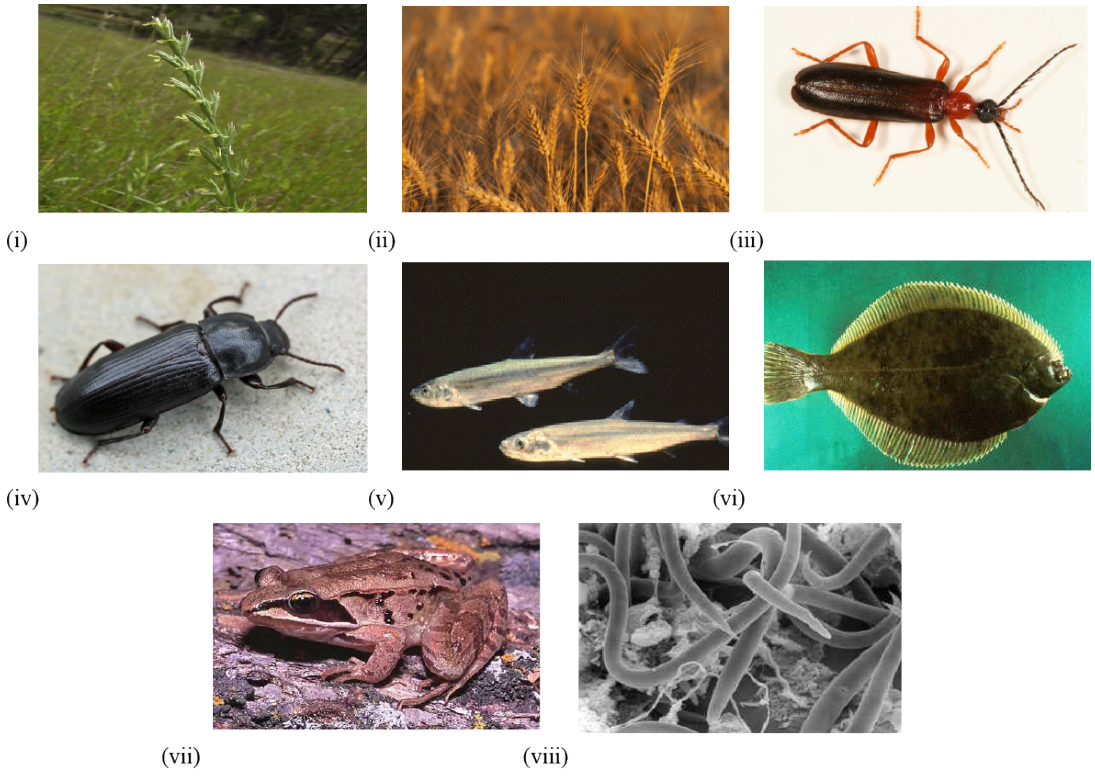


Figure 1.6. Examples of species that can survive freezing: (i) *Lolium perenne*, the perennial ryegrass (ii) *Triticum aestivum,* a winter wheat (iii) *Dendroides canadensis*, the pyrochoroid beetle (iv) *Tenebrio molitor*, the yellow mealworm beetle (v) *Osmerus mordax*, the rainbow smelt (vi) *Pseudopleuronectes americanus*, the winter flounder (vii) *Rana sylvatica*, the wood frog (viii) *Panagrolaimus davidi*, a nematode. The sources of the images are as follows:

(i) www.rbgsyd.nsw.gov.au/science/Evolutionary_Ecology_Research/Ecology_of_Cumberland_Plain_Woodland/woodland_plants/lolium_perenne

(ii) www.uniprot.org/taxonomy/4565

(iii) www.pbase.com/tmurray74/image/61621171

(iv) www.pbase.com/image/32057744

(v) www.iowagis.org/iris/fishatlas/IA162041.html

(vi) www.marine.rutgers.edu/navesink/about.htm

(vii) www.rbnc.org/herps.htm

(viii) www.nematodes.org/nematodegenomes/index.php/Panagrolaimus_davidi

can be divided into those that are "chilling sensitive (CS)" and "chilling tolerant (CT)" (Raison and Orr, 1990). CS plants do not tolerate ice formation in their tissues and will suffer damage and die when exposed to temperatures between +12 ^{o}C and 0 ^{o}C. They include tropical rainforest species, such as the horticultural plant *Saintpaulia ionantha* (Bodner and Larcher, 1987). Plants from outside of the tropics and from high tropical mountains are usually chilling tolerant and remain undamaged by exposure to subzero temperatures. However the extent of tolerance varies, some may suffer damage just below 0 ^{o}C and may not be able to complete their life cycle, while others may survive exposure to liquid nitrogen (-196 ^{o}C). CT plants have evolved a mechanism known as cold acclimation to increase their resistance to freezing temperatures, it is induced by exposure to low non-freezing temperatures or shortening day length (Levitt, 1980; Sakai, 1987).

1.3.1.3 Animals

To survive temperatures below the freezing point (Fp) of their body fluids (-5 ^{o}C for most terrestrial and fresh water animals, and -1.9 ^{o}C for marine invertebrates) ectothermic animals employ one of two basic strategies: freeze-avoidance or freeze-tolerance (Storey and Storey, 1996; Block, 2003; Sinclair *et al.*, 2003). Freeze-avoidance is widely found among terrestrial insects and other arthropods such spiders and mites. These arthropods may take advantage of existing factors such as a waterproof cuticle and protective silk cocoon as well as their ability to eliminate gut contents (a potent source of ice nucleators) in autumn and to introduce metabolic changes upon entry to winter dormancy. Freeze-avoiding animals are adapted for supercooling. Supercooling is a metabolic state in which the animals may maintain body fluids as liquid at temperatures below the equilibrium freezing point of these fluids. Freeze-avoiding animals may also produce antifreeze proteins and high concentrations of cryoprotectants or undergo partial dehydration to enhance supercooling (Davies and Sykes, 1997; Ewart *et al.*, 1999; Duman, 2001; Davies *et al.*, 2002; Duman *et al.*, 2004).

Freeze-tolerance has developed in many animal species. It occurs in species of terrestrial insects, marine invertebrates, nematodes and certain terrestrially hibernating amphibians and reptiles (Costanzo *et al.*, 1995; Wharton and Ferns, 1995;

Storey and Storey, 1996; Duman, 2001; Block, 2003; Wharton, 2003). Freeze-tolerant animals defend the liquid state of the cytoplasm while allowing ice to form in the extracellular spaces of their bodies (up to 70% of total body water may freeze). Ice nucleation proteins (INPs), recrystallisation inhibition proteins (RIPs) and cryoprotectants are synthesised by freeze-tolerant animals to control ice growth, maintain the critical minimum cell volume and stabilise the lipid bilayer (Zachariassen and Kristiansen, 2000; Duman, 2001; Wharton *et al.*, 2003; Wharton *et al.*, 2005; Yancey, 2005). Although uncommon, there are reports of animals that survive intracellular freezing, the best studied example is the nematode *Panagrolaimus davidi* (Figure 1.6(vii)) (Wharton and Ferns, 1995).

1.3.2 Why study freeze tolerance?

Freezing stress has a serious effect on the survival and success of microbes, plants and animals. The low temperatures may reduce the nutrient availability and the opportunity to reproduce successfully. The lack of available water in freezing conditions may impair biochemical reaction rates and cause changes in the membrane fluidity and protein conformation (Margesin et al., 2007). In plants, freezing temperatures are a major factor in determining the locations that are suitable for growing crops and horticultural plants, and freezing injury accounts for significant losses in plant productivity (Thomashow, 1999). An understanding of the molecular mechanisms that confer survival in freezing tolerant organisms is important for the development of biotechnological, agricultural and medical applications, such as improved freeze tolerance in plants and cold hardiness in animals, organ cryopreservation in medicine, using cold-active enzymes for detergent and food industries and in biological decontamination (Margesin et al., 2007).

1.4 Freezing injury

Freezing can cause significant injury to the cells and tissues of freezing-sensitive organisms. The damage may arise during the freezing and/or the thawing stage. When ice crystallisation occurs, the growth of ice crystals can cause severe physical damage. Intracellular freezing is generally lethal to a cell as the growing ice crystals cause physical damage to the subcellular architecture. The growth of ice crystals in the extracellular spaces will also cause physical damage since water expands when it

crystallises. In animals growing ice crystals may break capillaries, leading to a loss in vascular integrity when the cells thaw; this is the major barrier in attempts to freeze organs (Rubinsky *et al.*, 1990a; Rubinsky *et al.*, 1990b).

The water potential of ice is lower than that of liquid water. When ice crystals grow outside the cell this causes an osmotic gradient between the unfrozen extracellular fraction and the cell cytoplasm, forcing an efflux of water from the cell (Gusta et al., 1975). The cell becomes dehydrated and the concentration of cytoplasmic components such as ions, sugars and proteins increases. As the temperature decreases the water potential of water reduces, thus dehydration of the cell will increase as the temperature lowers. The increase in concentration of solutes increases the viscosity of cellular fluids, affecting enzyme function and changing the cellular pH, both of which can severely damage cells and tissues.

Freezing stress results in the production of reactive oxygen species (ROS) (Park et al., 1998). ROS accumulation in cells leads to the oxidative modification of proteins, lipids, DNA and other macromolecules. Proteins become more susceptible to aggregation, unfolding and degradation. ROS causes substantial damage to membrane lipids decreasing membrane fluidity and increasing membrane leakage. ROS will also damage DNA by causing base and nucleotide modifications and strand breaks (Temple et al., 2005). In thawing, the rapid reintroduction of oxygen may also initiate a further burst of damage by reactive oxygen species (Storey and Storey, 1996).

In addition to the oxidative damage to the membrane lipids, freezing-induced dehydration has devastating effects on the lipid bilayer. The consequence is a loss of compartmentalisation resulting in electrolyte and solute leakage. The membrane can only withstand limited osmotic shock and cell dehydration before the stress leads to a phase change in the membrane lipids, from a bilayer to a gel state, making the membrane unstable (Storey and Storey, 1996).

1.5 Mechanisms of freezing tolerance

Freezing tolerant organisms utilise diverse mechanisms to prevent intracellular ice formation and to protect membrane integrity in response to subzero temperatures.

These include the synthesis of cryoprotectant molecules, antifreeze proteins, ice-nucleating proteins, heat shock proteins, alterations to phospholipid bilayers and the expression of cold responsive genes. These adaptive strategies will be discussed in the following sections.

1.5.1 Accumulation of colligative cryoprotectants

1.5.1.1 Colligative cryoprotectants

Colligative cryoprotectants are typically sugars or polyhydric alcohols that confer freezing tolerance by accumulating intracellularly in high concentrations (0.2-2 M). The high concentration of cryoprotectants increases osmotic pressure in the cells reducing the percentage of total water that can accumulate as extracellular ice. This prevents a reduction in the cell volume below the critical minimum cell volume. Some species may also use a combination of cryoprotectants (Joanisse and Storey, 1994; Walters *et al.*, 2009a). Cryoprotectants produced by microbes, plants and animals in response to exposure to low temperatures include the sugars glucose and trehalose, the polyhydric alcohols, glycerol, sorbitol, ribitol, erythritol, threitol and ethylene glycol. Other cryoprotectants such, as proline and glycine betaine, accumulate in plants (Figure 1.7) (Storey, 1997).

1.5.1.2 Glycerol

The most commonly employed colligative cryoprotectant is the three-carbon polyhydric alcohol, glycerol (Figure 1.7(i)). The widespread occurrence of glycerol as a cryoprotectant is derived from several factors: its high solubility, its compatibility with other biological macromolecules, it is produced by constitutive and centrally located biosynthetic and catabolic pathways, it has an optimal output of osmotically active particles (1 hexose unit from glycogen produces 2 glycerol molecules) and it has optimal conversion efficiency (Storey and Storey, 1991; Storey, 1997). Well-studied models that accumulate glycerol include the gall moth, *Epiblema scudderiana* (Storey and Storey, 1996) caterpillars and the rainbow smelt, *Osmerus mordax* (Raymond, 1992; Raymond, 1995; Storey and Storey, 1996).

Glycerol is produced from the catabolism of glycogen stores. In insects, a glycogen phosphorylase activates glycerol production via a signal transduction cascade system

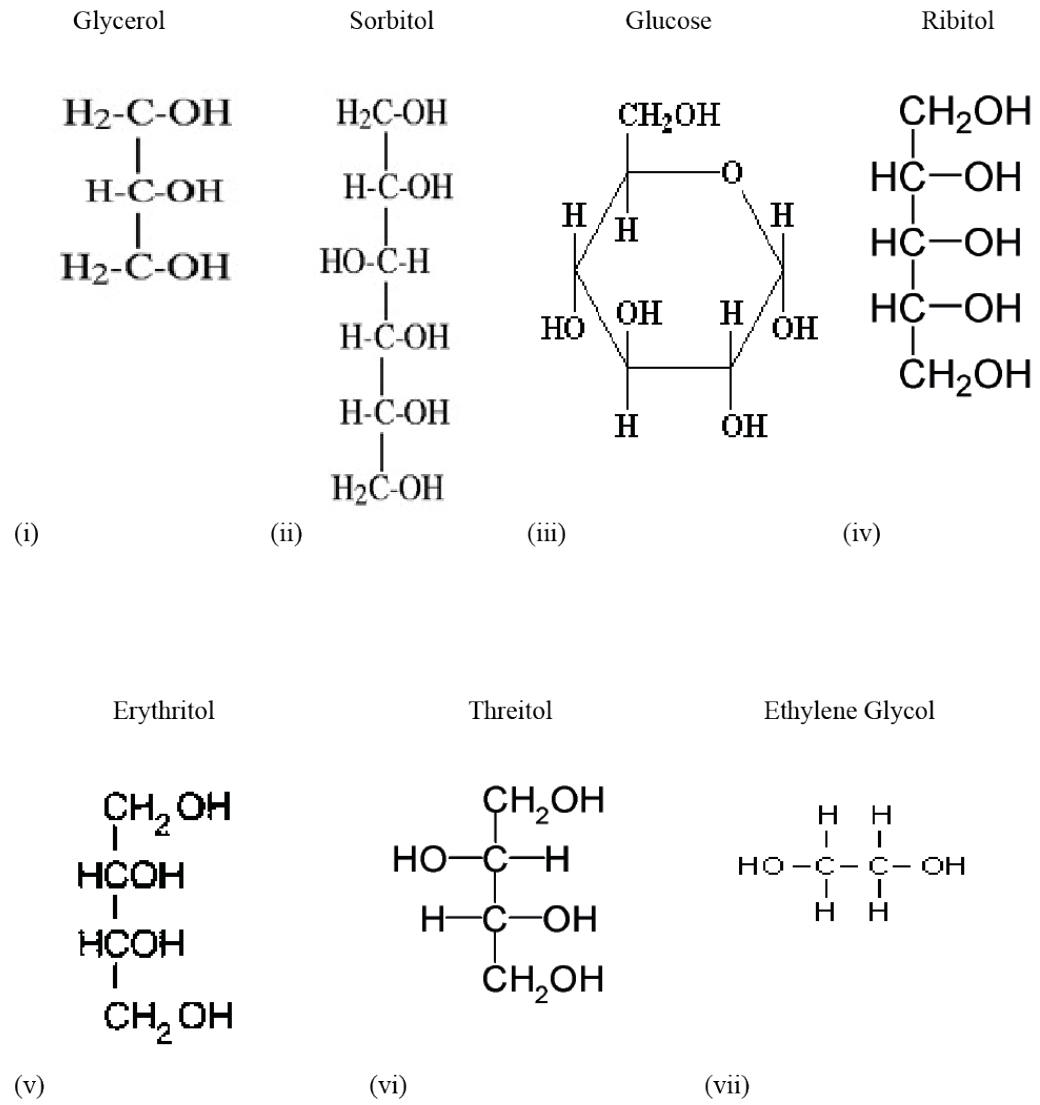


Figure 1.7. The chemical structures of colligative cryoprotectants. (i) Glycerol, a polyol; (ii) Sorbitol, a polyol; (iii) Glucose, a carbohydrate; (iv) Ribitol, a polyol; (v) Erythritol, a polyol; (vi) Threitol, a polyol; (vii) Ethylene glycol, a polyol.
This image was modified from Yancey (2005).

activated at temperatures in the range of 5 ^{o}C to -5 ^{o}C (Clark and Worland, 2008). The production of glycerol from glycogen is inter-convertible and glycogen phosphorylase activity is regulated by the differential cold-responsiveness of glycogen phosphorylase kinase (GPK) and protein phosphatase 1 (PP1) (Hayakawa, 1985).

Glycerol is used widely in nature as a protectant not only during freezing but also in other natural occurrences of water stress such as desiccation. It may be possible that transporters facilitate glycerol movement across membranes during stress. Alternatively since glycerol is composed of hydrogen and hydroxyl groups that are orientated very similar to those in water that it may simply pass through aquaporin water channels (Storey, 1997).

1.5.1.3 Sorbitol

Sorbitol (Figure 1.7 (ii)) is also frequently produced in combination with glycerol as a cryoprotectant. The larvae of the freezing tolerant gall fly, *Eurosta solidaginis* accumulate about 250 mM glycerol but also produce about 150 mM sorbitol (Storey, 1997). Sorbitol is synthesised from glucose-6-phosphate using the enzymes glucose-6-phosphatase and NADPH-dependent polyol dehydrogenase. It is degraded in spring by NAD-dependent sorbitol dehydrogenase and hexokinase to produce fructose (Storey and Storey, 1991).

1.5.1.4 Glucose

The synthesis of glycerol and sorbitol requires a period of cold acclimation where over days or weeks organisms may synthesise stocks from glycogen and glucose-6-phosphate respectively (Joanisse and Storey, 1995). To date, earthworms and amphibians are the only animals that have been found to utilise glucose as their primary cryoprotectant (Calderon et al., 2009). Some animals that require an immediate response to ice formation may use glucose as a cryoprotectant. Glucose transporter proteins mediate the rapid glucose movement across the lipid bilayer. The earthworm *Dendrobaena octaedra* can tolerate freezing to -20 ^{o}C and glucose can constitute more than 20% of dry mass less than 24 h after freezing has commenced (Calderon et al., 2009). In mammalian cells glucose is a poor cryoprotectant suggesting that organisms that use glucose as a cryoprotectant must have adaptations

in sugar transport to facilitate its quick export across membranes. The wood frog, *Rana sylvatica*, is an extensively studied freeze tolerant species that utilises glucose as a cryoprotectant. It appears that these frogs have increased numbers of glucose transporters in their membranes to deal with the requirement for rapid cryoprotectant movement (King *et al.*, 1993; King *et al.*, 1995).

In *R. sylvatica* production of glucose is triggered by the initiation of freezing in the skin. The liver glycogen stocks are converted to glucose-6-phosphate (G6P) raising the glucose sugar levels rapidly to 200-300 mM before it is exported to core organs to provide cryoprotection (Storey and Storey, 1985; Crerar *et al.*, 1988; Arav *et al.*, 1993). Generally high glucose levels cause metabolic problems including the nonenzymatic glycation of proteins resulting in their impaired function or acting as pro-oxidants stimulating the production of reactive oxygen species (Margesin et al., 2007). However, the wood frog has mechanisms to prevent nonenzymatic glycation damage to their proteins and has highly active antioxidant enzymes to keep glucose-mediated oxidative damage to a minimum (Margesin et al., 2007).

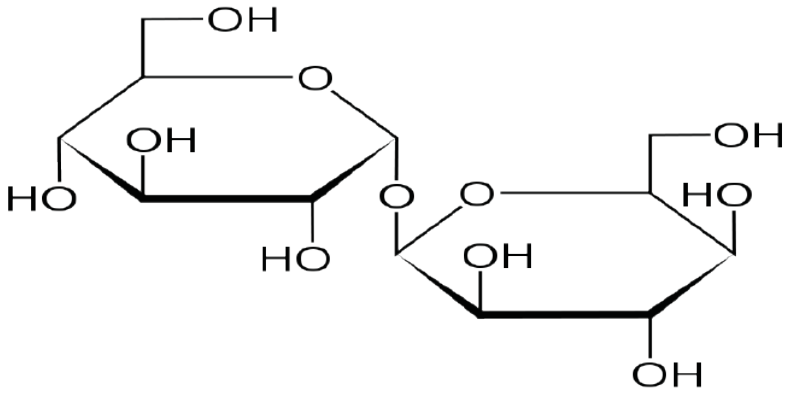
1.5.2 Accumulation of low-molecular weight cryoprotectants

The non-reducing sugar trehalose, the amino acid proline and the quaternary ammonium compound glycine betaine are the compounds that are most commonly associated with protecting membranes (Figure 1.8). These low molecular weight cryoprotectants, which may stabilise membrane bilayers, accumulate in low amounts (<0.2 M). They specifically target and stabilise the membrane bilayer by preventing the irreversible transition to the gel state when the plasma membrane is compressed during cell volume reduction.

1.5.2.1 Trehalose

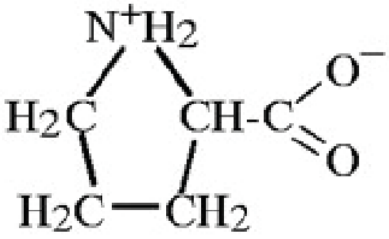
Trehalose is a soluble, non-reducing disaccharide of glucose (Figure 1.8 (i)). It is composed of two glucose molecules and is synthesised by the enzymes trehalose-6-phosphate phosphatase and trehalose-6-phosphate synthase. They link the 1, 1 carbon atoms of the two glucopyranose rings together. The energy required for formation of trehalose is very low, which makes it a very stable, nonreactive molecule (Elbein, 1974). Since its discovery in the embryonic cysts of the brine shrimp, *Artemia salina,*

Trehalose



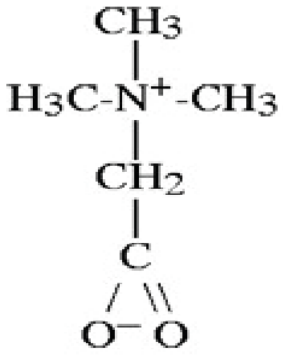
(i)

Proline



(ii)

Glycine Betaine



(iii)

Figure 1.8. Cryoprotectants that may stabilise biological membranes and proteins and also function as osmolytes. (i) Trehalose, a carbohydrate. (ii) Proline, an amino acid. (iii) Glycine betaine, a quaternary ammonium compound. The sources of the images are as follows: (i) www.de.wikipedia.org/wiki/Datei:Trehalose.png;
(ii) & (ii) Yancey (2005).

the role of trehalose in desiccation tolerance has been studied extensively (Clegg, 1965, 1967). As freezing is associated with a loss of available water, the mechanisms employed for trehalose to protect the cell during desiccation also hold true for freezing stress. Trehalose is also synthesised in many organisms in response to high osmolarity, hypoxia and heat shock. Trehalose shields organisms from damage caused by desiccation and freezing stress by the protection of proteins and the cell membrane (Crowe *et al.*, 1984; Carpenter and Crowe, 1988; Carpenter, 1993; Prestrelski *et al.*, 1993).

Trehalose has been shown to be the most efficient carbohydrate for protecting proteins (Prestrelski et al., 1995). It can stabilise the enzyme phosphofructokinase which normally becomes irreversibly damaged when dried (Carpenter et al., 1987). Trehalose molecules can protect proteins by several mechanisms. They may replace accessible "bound" water by hydrogen bonding through its OH groups to the polar residues of the protein during dehydration (Carpenter, 1993). They may also protect proteins by a process known as preferential exclusion (Carpenter and Crowe, 1988). It is hypothesised that the addition of certain solutes such as other sugars to a protein causes a thermodynamically unfavourable effect. This results in these solutes becoming excluded from the immediate vicinity of the protein, making the protein more likely to exist in a compact folded form in order to reduce the available surface area for interaction with the solute (Timasheff, 1992). Trehalose may also protect proteins by acting as a chemical chaperone (Singer and Lindquist, 1998; Crowe, 2007). Unlike traditional molecular chaperones that promote correct folding, chemical chaperones may influence the rate or fidelity of the folding reaction by stabilising the properly folded form (Welch and Brown, 1996). Trehalose has also been shown to be an antioxidant. In the yeast species, *Saccharomyces cerevisiae* it accumulates in response to oxidative stress and it protects against lipid peroxidation (Herdeiro et al., 2006). In the absence of trehalose, *S. cerevisiae* trehalose deficient mutants were shown to develop increased levels of protein oxidation damage upon exposure to ROS (Benaroudj et al., 2001).

Both freezing and desiccation stress cause water molecules to be removed from the lipid bilayers, resulting in a phase transition in the plasma membranes from the liquid crystalline state to a gel phase. The water replacement hypothesis proposes that

trehalose can protect the lipid bilayer by inserting between the phosphate groups in the lipid bilayer (Crowe et al., 1984). Trehalose stabilises and prevents leakage of the intracellular contents by hydrogen bonding to the phosphate heads of the bilayer in place of water. This prevents the increase in van der Waal's interactions between the phosphate groups. Thus the presence of trehalose in drying cells leads to a decrease in phase transition temperature. Due to this the membrane remains in a liquid crystalline state and does not undergo a phase transition during freezing or upon rehydration during melting (Crowe et al., 1996). Trehalose may also contribute to the stability of the lipid bilayer through the formation of glasses. A glass forms on both sides of the membrane and the rigidity of the glass prevents the membrane contracting into the gel form (Byrant et al., 2001).

There are at least 5 biosynthetic pathways known for the synthesis of trehalose. For the purposes of this literature review only the first pathway discovered by Cabib in 1957 will be discussed (Figure 1.9). This pathway is the most widely distributed and has been found to be present in eurbacteria, archaea, fungi, animals and the plants. This pathway involves two enzymatic steps catalysed by trehalose-6-phosphate synthase (TPS) and trehalose-phosphatase (TPP). The TPS catalyses the transfer of glucose from uridine diphosphoglucose (UDPG) to glucose-6-phosphate, this generates trehalose-6-phosphate (T6P) and uridine diphosphate (UDP). The TPP dephosphorylates T6P to trehalose and inorganic phosphate. The TPS and TPP protein domains appear to have coevolved and have likely undergone several gene duplications and lateral gene transfer events (Avonce et al., 2006). Trehalose is degraded by the enzyme trehalase forming two molecules of D-glucose.

1.5.2.2 Proline

Many plants and microbial cells accumulate proline as a cryoprotectant (Figure 1.8 (ii)). In addition to its role as an osmolyte for osmotic adjustment, proline has been shown to enhance the stability of proteins and membranes during freezing (Rudolph and Crowe, 1985). Proline also has the ability to scavenge reactive oxygen species such as hydroxyl radicals, singlet oxygen and superoxide anions (Smirnoff and Cumbes, 1989; Alia *et al.*, 2001; Kaul *et al.*, 2008). Proline acts as a cryoprotectant by preventing ice nucleation and dehydration under freezing conditions by forming

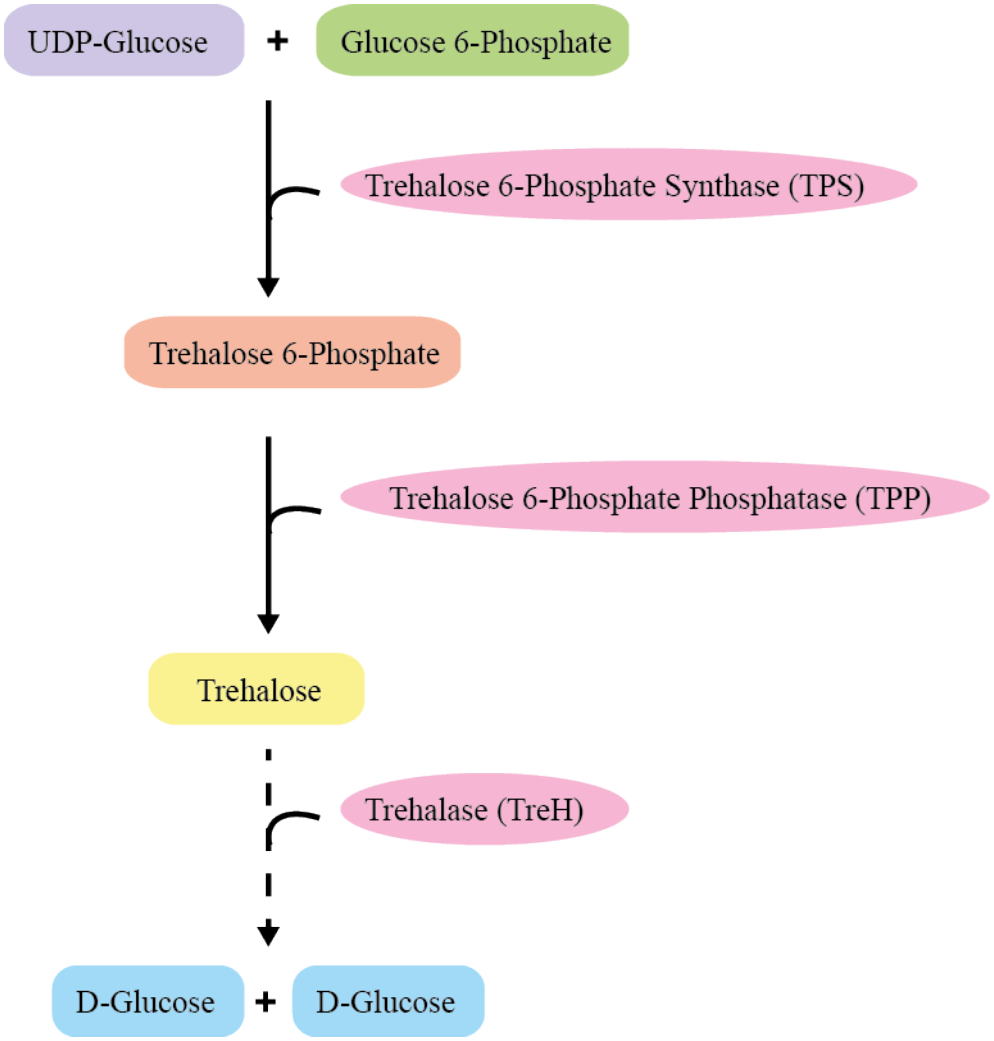


Figure 1.9. The trehalose-6-phosphate synthase (TPS)/trehalose-phosphatase (TPP) pathway for trehalose biosynthesis. Trehalose is synthesised via the enzymes TPS and TPP, and is degraded by trehalase. This image was modified from Avonce *et al.* (2006).

strong hydrogen bonds with free intracellular water without interacting with any macromolecules (Rudolph and Crowe, 1985). The action of proline as a stress protectant has been studied in detail in *S. cerevisiae*. Proline levels in *S. cerevisiae* are not normally elevated under various stress conditions. Proline accumulating *S. cerevisiae* mutants have cryoprotective activity that is nearly equal to that of glycerol or trehalose and these cells are also more tolerant than wild-type cells to desiccation, oxidative and ethanol stress (Takagi, 2008).

1.5.2.3 Glycine betaine

The quaternary ammonium compound glycine betaine (Figure 1.8 (iii)) accumulates in response to stress in many crop plants, including sugar beet (*Beta vulgaris*), barley (*Hordeum vulgare*) and wheat (*Triticum aestivum*) (Ashraf and Foolad, 2007). This compatible cryoprotectant can protect membranes, proteins and enzymes during freezing stress (Coughlan and Herber, 1982; Anchordoguy *et al.*, 1987; Zhao *et al.*, 1992; Rodes and Hanson, 1993). It has been suggested that glycine betaine can help stabilise protein tertiary structure and prevent or even reverse the effect of the disruption of the tertiary structure of proteins caused by non-compatible solutes. Glycine betaine can protect spinach thylakoids from freezing stress. It has been proposed that this protection is caused by a weak interaction between the positive quaternary ammonium cation of glycine betaine and the anionic carboxyl groups of the exposed membrane proteins (Coughlan and Herber, 1982).

1.5.3 Antifreeze proteins

Antifreeze proteins (AFPs) and antifreeze glycoproteins (AFGP) are used as part of the freeze-avoidance strategy in some fish, insects, plants and microbes. AFGP were first identified in the Antarctic teleost fishes, where they were found to lower the freezing point of the blood serum of these fishes by more than 1 ^{o}C to match or exceed the freezing temperature of the seawater (DeVries and Wohlschlag, 1969; DeVries *et al.*, 1970). This effect of lowering the freezing point of the blood serum was too great to be caused by colligative properties, demonstrating that these were a new unusual group of proteins. AFPs lower the freezing point of body fluids by interacting directly with the ice surface (Raymond and DeVries, 1977). Thus AFPs may be generally defined as proteins that have affinity for ice. When AFPs adsorb to

a growing ice front they restrict the growth of the ice front to the regions between the adsorbed protein molecules (Raymond and DeVries, 1977). The ice grows with a local curvature between the protein molecules, making it less thermodynamically favourable for more water molecules to add to the ice lattice (Figure 1.10). This results in the non-colligative, non-equilibrium lowering of the freezing point (Raymond and DeVries, 1977). The temperature difference between the colligative freezing point and the non-equilibrium freezing point is termed thermal hysteresis (TH).

In addition to the ability to depress the freezing point of body fluids, several antifreeze proteins are also capable of ice recrystallisation inhibition (RI) activity (Knight *et al.*, 1984; Tomczak *et al.*, 2003). Ice crystallisation is the growth of larger ice crystals at the expense of the smaller ones. Large ice crystals are a major cause of tissue damage in freezing sensitive organisms, therefore RI inhibtion proteins can significantly improve freezing survival. There are several organisms that have low TH activity but have high RI activity. It appears in these cases that it is more important to control ice growth and structure than to prevent freezing. This is especially prevalent in plants but is also found in fish, bacteria, fungi, algae and yeast (Ewart *et al.*, 1999; Venketesh and Dayananda, 2008).

1.5.3.1 Mechanism that Antifreeze proteins bind to the ice

The mechanisms by which AFP proteins bind to ice have been of much debate. The fish type I AFPs found in flounders and sculpins are small alanine-rich, amphipathic and α-helical in structure (Davies *et al.*, 1982; Hew *et al.*, 1985; Scott *et al.*, 1987; Chakrabartty *et al.*, 1988; Chao *et al.*, 1996). As they are the simplest of the fish AFPs, Type I AFPs have been used as a model system for studying how these proteins bind to ice.

Following the sequencing of one of the type I AFPs from winter flounder, DeVries and Lin (1977) proposed the hydrogen-bonding hypothesis to describe the binding of AFPs to ice. This peptide is predicated to form an α-helical structure which has a repeating structure of 37 amino acids dominated by threonine and aspartic acid residues, each of which are separated by 11 amino acid residues, and are predicated to

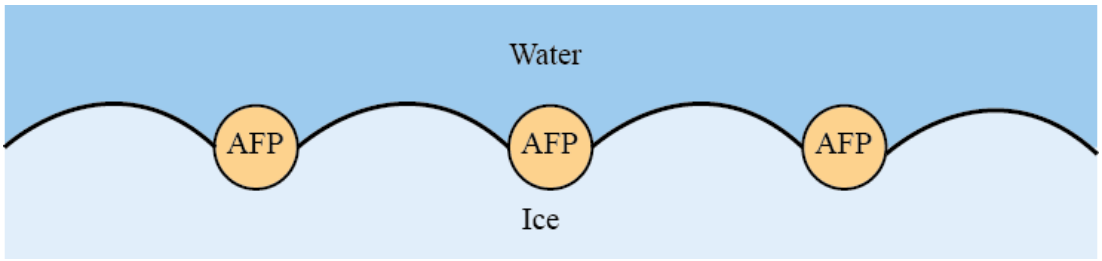


Figure 1.10. A schematic model for how AFPs at the ice/water interface induce thermal hysteresis. When AFPs adsorb to the ice it grows with a local curvature between the protein molecules making it less thermodynamically favourable for more water molecules to add to the ice lattice. This results in the lowering of the freezing point below the melting point, this difference is termed thermal hysteresis. This image was modified from Barrett (2001).

form hydrogen bonds to oxygen atoms on the primary prism plane of ice (Figure 1.11) (DeVries, 1984). The main support for the hydrogen-bonding hypothesis is the proposed distance in the α-helix of 4.5 Å between the threonine hydroxyl and the carboxyl group of the aspartates that would match with the 4.5 Å spacing between the oxygen atoms on the basal plane of a hexagonal ice crystal. This motif is repeated three times along the helix of the peptide sequence. The plane of ice that the winter flounder AFP bound to was later determined by ice etching experiments to be the pyramidal plane not the primary prism plane (Knight et al., 1991). However the hydrogen-bonding hypothesis could still adapted to the pyramidal plane (Figure 1.12) since the 16.5 Å repeats of the threonine and aspartate residues match to the surface of this plane of ice (Wen and Laursen, 1992).

Mutagenesis and amino acid replacement studies where the threonine residues were replaced with serines and valines have shown that hydrogen bonding might not be essential to ice binding (Chao et al., 1997). When the middle two threonines from the winter flounder AFP were replaced with a serine, ice-binding activity was lost, although serine can form hydrogen bonds. However, replacement with valine caused little loss in activity indicating that the methyl group of threonine may be important for ice binding, not the hydroxyl group (Chao et al., 1997). This observation and the realisation that the hydrophilic surface of the helix of Type I AFP is very variable, while the hydrophobic face that consists of regularly spaced alanines and threonine is highly conserved led to the proposal that the alanine-rich surface is in fact the ice-binding face of the helix (Baardsnes et al., 1999). Substituting the alanine residues with leucine residues further tested this hypothesis. In points where a leucine was substituted on the hydrophilic surface there was little effect on the thermal hysteresis activity. In contrast, when alanines were substituted with leucines on the hydrophobic surface activity was very low or completely lost. Based on these results Baardsnes *et al.* (1999) suggested that the ice-binding surface is the alanine-rich surface and that the methyl group of the threonines is more important for binding to ice than the hydroxyl group.

The hydrogen-bonding hypothesis was further investigated by experimentally determining the ice-binding surface of the type I AFP from shorthorn sculpin. In this instance the alanine residues were replaced with lysine. In a similar result to the

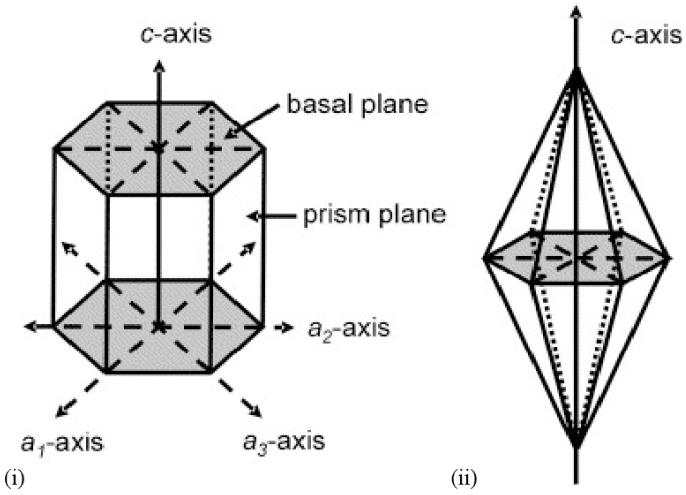


Figure 1.11. (i) A diagram illustrating the crystal axes and the main faces in a hexagonal ice crystal. (ii) A diagram of the classical hexagonal bipyrimidal ice crystal produced by most fish AFPs illustrating that the vertices of the crystal lie along the *c*-axis. The basal plane is also referred to as the pyramidal plane. The source of this image is Scotter *et al.* (2006).

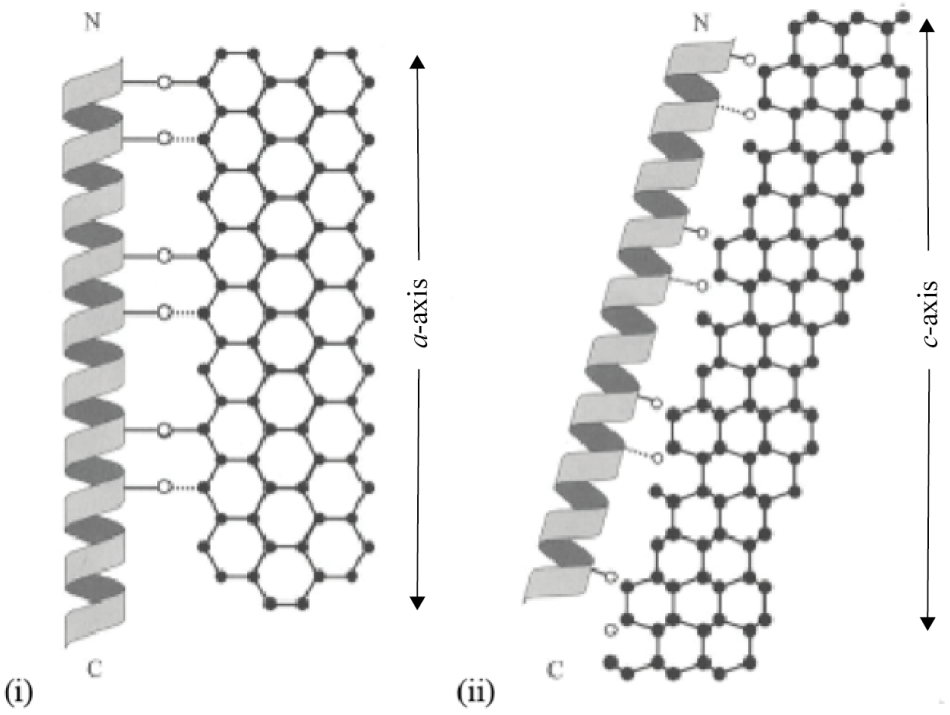


Figure 1.12. Illustrations of the hydrogen-bonding hypothesis for AFP binding to ice. (i) A representation of the original model of DeVries (1977) showing winter flounder (type I) AFP hydrogen bonding through its Thr and Asx side chains to oxygen atoms on the primary prism plane of ice. The dotted lines are from the Asx residues and indicate hydrogen bonding to oxygen atoms. (ii) The representation of the revised hypothesis proposed by (Wen and Laursen, 1992). This accommodates for the discovery of the ice plane bound by the winter flounder (Knight et al., 1991). The Thr and Asx side chains are within hydrogen-bonding range of oxygen atoms on the ice. The source of this image is Davies *et al.* (2002).

winter flounder AFP, alanines replaced on the hydrophilic surface had little effect on ice-binding activity, whereas those replaced on the hydrophobic surface had antifreeze activity eliminated (Baardsnes et al., 1999). This redefined ice-binding surface of the shorthorn sculpin AFP was modelled with the docking of the hydrophobic face to the secondary-prism plane of ice. The helical repeat has a 16.5 Å spacing between the alanine residues; this places the alanine methyl side chains within a groove on the plane of ice forming a tight complementary fit and allows van der Waals and hydrophobic interactions to occur over the length of the helix.

1.5.3.2 Fish antifreeze proteins

Fish antifreeze proteins are classified into five types based on their primary sequences and 3D structure, each type is radically different yet all bind to ice and lower the non-equilibrium freezing point below the melting point. Also, the distribution of these different protein types does not follow the evolutionary relationships of the host fish (Figure 1.13).

1.5.3.3 Fish Type I antifreeze proteins

The Type I antifreeze proteins are small (3.3- 9.6 kDa), alanine-rich (up to 60%), α-helical proteins, and that contain an 11-amino acid repeat sequence that generally commences with a threonine (Davies and Hew, 1990; Harding *et al.*, 1999). They are present in fishes from three of the fish taxonomic orders, the pleuronectiforme order includes the winter flounder and the yellowtail flounder, the scorpaeniformes order includes the sculpins and snailfish and the perciforme have the cunner (Hew *et al.*, 1985; Scott *et al.*, 1987; Low *et al.*, 1998; Evans and Fletcher, 2001; Low *et al.*, 2001; Evans and Fletcher, 2004; Wang *et al.*, 2009). The evolutionary precursor to the Type I AFPs is not known.

Until relatively recently it appeared that all fish AFPs have comparable thermal hysteresis activity, they produce ~ 1 ^{o}C of thermal hysteresis at a relatively high AFP concentrations (10-40 mg/ml). However, a hyperactive AFP was discovered in the winter flounder (Marshall et al., 2004). Homologues of this hyperactive AFP have since been found in the yellowtail flounder (*Limanda ferruginea*) and the American plaice (*Hippoglossoides platessoides*) (Gauthier et al., 2005). This hyperactive AFP

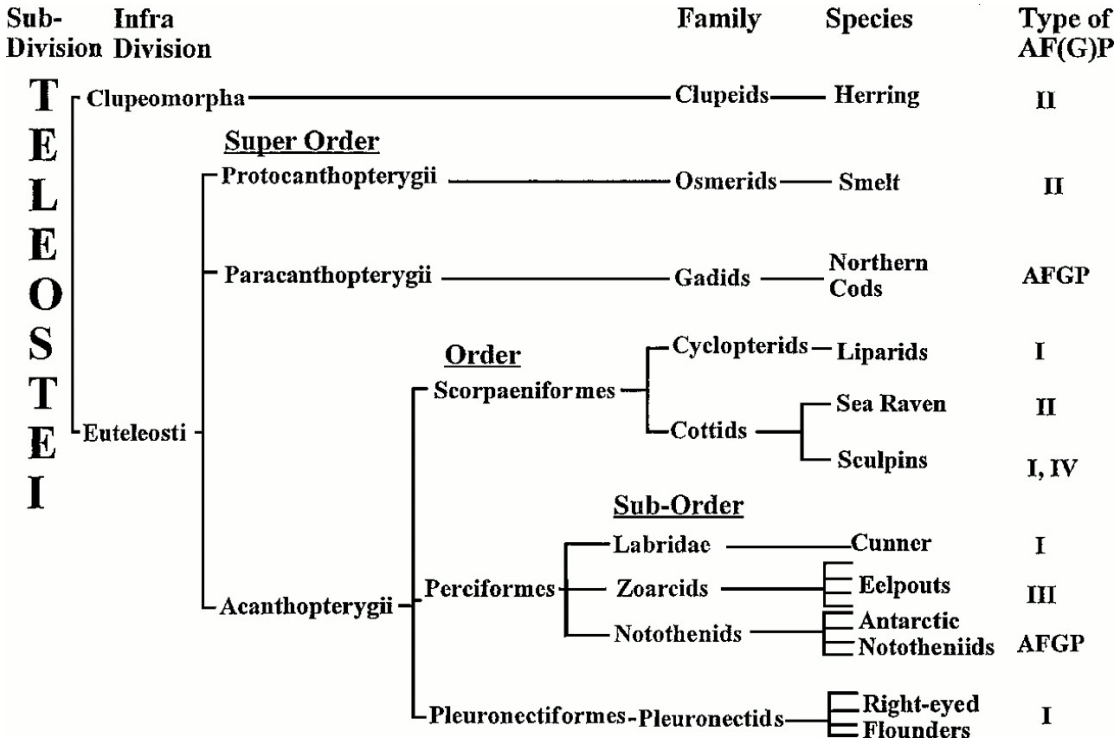


Figure 1.13. Phylogeny of AFP producing fish.

The source of this image is Fletcher *et al.* (2001).

provides ~ 1.1 ^{o}C of thermal hysteresis activity but at a concentration of only 0.1 mg/ml, suggesting that it alone may protect the fish from freezing. This protein has a mass of 17 kDa and forms a dimer making it considerably larger than the other Type I AFPs. However, they all share high alanine content and are α-helical. Due to the similarity in their structure and as the sequence clearly shows that it has a common origin to the Type I AFP, the hyperactive AFPs have been designated the name hyp-Type I (Graham et al., 2008b).

1.5.3.4 Fish Type II antifreeze proteins

The Type II antifreeze proteins are characterised as been 14-24 kDa, cysteine rich, globular proteins. They have been found in five fish species, including Atlantic herring (*Clupea harengus*), rainbow smelt (*Osmerus mordax*), Japanese smelt (*Hypomesus nipponensis*), sea raven (*Hemitripterus americanus*) and longsnout poacher (*Brachyopis rostratus*) (Ewart *et al.*, 1992; Ng and Hew, 1992; Ewart and Fletcher, 1993; Yamashita *et al.*, 2003; Nishimiya *et al.*, 2008). The Type II AFPs are homologs of the carbohydrate recognition domain (CRD) of Ca^{2+}-dependent (C-Type) lectins and all share the same characteristic fold, which includes five disulfide bridges, α-helices, β-sheets and a large proportion of coil structure (Ewart *et al.*, 1992; Sonnichsen *et al.*, 1995; Ewart *et al.*, 1996; Gronwald *et al.*, 1998). The Type II AFPs may be divided on the basis of their calcium requirement. The Atlantic herring, rainbow smelt and Japanese smelt are Ca^{2+}-independent and the sea raven and longsnout poacher are Ca^{2+}-dependent.

There have been conflicting theories for the evolution of the Type II AFPs. Liu *et al.* (2006) proposed that the Type II AFPs are derived from a ten-cysteine lectin isoform that pre-existed in the ancestor to most of the fish but has since been lost in all of the other branches. Graham *et al.* (2008) discounted the work and postulated that the acquisition of AFP activity in these relatively distantly related fish is the result of lateral gene transfer.

1.5.3.5 Fish Type III antifreeze proteins

The Type III antifreeze proteins are approximately 7 kDa, globular proteins with one flattened surface composed of 61-64 amino acids. Unlike the other fish antifreeze

proteins they are not dominated in composition by one amino acid and do not have repetitive motifs. They are highly conserved and abundant gene family with high sequence conservation (45%). The Type III AFPs have been characterised in a number of different polar fish species including, ocean pout (*Macrozoarces americanus*), wolfish (*Anarhichas lupus*), Arctic eel pout (*Lycodes polaris*) and the Antarctic eel pouts (*Austrolycicthys brachycephalus, Lycodichthys dearborni* and *Rhigophila dearborni*) (Schrag *et al.*, 1987; Hew *et al.*, 1988; Scott *et al.*, 1988; Cheng and DeVries, 1989; Wang *et al.*, 1995).

There are no obvious homologs to the Type III AFPs and their evolution is unclear. However, they do have some sequence similarity with the extreme C-terminal region of the enzyme, sialic acid synthase (SAS). A comparison of the structure between HPLC12 and SAS revealed a high similarity in the protein core and in the flat ice-binding region. The function of this AFP-like domain in SAS is not known, but as this enzyme interacts with sugars and polysaccharides, like the progenitors to other antifreeze proteins, it may possible that the Type III AFPs arose from this enzyme (Baardsnes and Davies, 2001).

1.5.3.6 Fish Type IV antifreeze proteins

The Type IV AFPs are the most recently identified AFP type. They were first isolated from the longhorn sculpin, *Myoxocephalus octodecimspinosis* (Deng et al., 1997). This 12.3 kDa protein, LS-12, contains 108 amino acids and has high glutamine content (17%). The protein contains a 20-amino-acid N-terminal signal sequence that is consistent with it being exported to the blood. Circular dichroism and conformational analysis based on sequence data indicates that the LS-12 structure forms a left-handed, antiparallel four-helix bundle with four amphipathic main helices that are oriented with their hydrophobic surfaces forming the core of the bundle and their hydrophilic surfaces directed outwards to the solvent (Deng and Laursen, 1998). LS-12 has sequence similarity to several four-helix bundle serum/haemolymph apolipoproteins, such as apolipoprotein E from guinea pig, and apolipoproteins III from the African locust, suggesting that this AFP may have arisen through recruitment and mutation of a plasma apolipoprotein (Deng et al., 1997). Recently, Gauthier *et al*. (2008) proposed that the primary role of these apolipoproteins-like

AFPs is not as antifreeze. They found that the plasma antifreeze activity concentration from the longhorn sculpin was less than 0.1 mg/ml, which is too low to protect the blood from freezing in icy seawater. They hypothesise that the Type IV AFP has the potential to develop as a functional antifreeze protein in the longhorn sculpin but may not have been selected for this purpose because of the presence of the more active Type I AFP in this fish species (Low *et al.*, 2001; Gauthier *et al.*, 2008).

1.5.3.7 Fish Type V antifreeze proteins

The fifth class of fish antifreeze proteins are the antifreeze glycoproteins (AFGPs). They characterised by having several repetitive three amino acid repeating units (Ala-Ala-Thr)$_n$, with minor sequence variation, and the disaccharide β-D-galactosyl-(1-3)-α-D-*N*-acetylgalactosamine attached to the hydroxyl oxygen of each threonine residue (DeVries *et al.*, 1971; Shier *et al.*, 1972). They are present in the Antarctic notothenioid (*Dissostichus mawsoni*) and the northern cod (*Gadus ogac*, *Microgadus tomcod*, *Eleginus gracilis* and *Boreogadus saida*) (Feeney and Yeh, 1978; Osuga and Feeney, 1978; O'Grady *et al.*, 1982). The AFGPs are 2.6-33.7 kDa in size, which corresponds to 4 to 50 repetitions of the glycosylated tripeptide unit (DeVries, 1983). The AFGPs are grouped into eight size classes based on their electrophoretic properties, AFGP1-5 are the larger and AFGP6-8 are the smaller (Feeney, 1974). The structure of the AFGPs in the presence of ice has not been described to date, but they are thought to form an amphipathic structure with a hydrophilic face containing exposed hydroxyl groups of the sugars and a hydrophobic face containing the exposed methyl groups of the amino acid and N-acetyl groups (Peltier et al., 2010).

The AFGP gene from the Antarctic notothenioid, *D. mawsoni*, derives from a gene encoding a pancreatic trypsinogen (Chen et al., 1997b). Sequence comparisons have shown that the region encoding the AFGP polyprotein has evolved by repeated duplication and rearrangement of a 9-bp segment crossing the first exon-intron boundary of the trypsinogen gene (Chen et al., 1997b). Thus the AFGPs ice-binding function is derived from the expansion of a very small pre-existing partially noncoding DNA in the original trypsinogen gene (Chen *et al.*, 1997b; Logsdon and Doolittle, 1997). A common evolutionary origin was inferred for the cod and the notothenioid AFGPs based on the structural similarity of these proteins, but the genes

encoding these proteins are completely distinct to the notothenioid AFGP (Scott *et al.*, 1986; Chen *et al.*, 1997a). They both have different signal peptide sequences, different spacer sequences that link the encoded AFGP molecules in the polyprotein, distinct codon bias for the AFGP tripeptide and have different genomic loci. This evidence strongly suggests that the near identical AFGPs derived from these unrelated fish are a rare example of independent convergent evolution (Chen et al., 1997a).

1.5.3.8 Insect and arthropod antifreeze proteins

Since their discovery in fish, AFPs have been identified in at least 50 different species of insects (Duman et al., 2004). The most extensively studied include the yellow mealworm (*Tenebrio molitor*), the spruce budworm (*Choristoneura fumiferana*), the snow flea (*Hypogastrura harveyi*) and the pyrochoroid beetle (*Dendroides canadensis*) (Graham *et al.*, 1997; Tyshenko *et al.*, 1997; Duman *et al.*, 1998; Graham and Davies, 2005). Many of the insect AFPs have low thermal hysteresis activity, but it has become apparent that some insect AFPs are considerably more potent than those described in fish making them very interesting from a biotechnological perspective (Scotter et al., 2006). The levels of AFPs produced by some insects and arthropods are under the control of developmental and environmental regulation (Graham et al., 2000).

The first observation of antifreeze activity in an organism came from observations of the yellow mealworm, *Tenebrio molitor* (Ramsay, 1964; Grimstone *et al.*, 1968). This beetle produces several isoforms of a potent antifreeze protein (TmAFP) that can have a TH activity of 5.5 ^{o}C (Graham *et al.*, 1997; Liou *et al.*, 1999). The specific activity of TmAFP is up to 100 fold greater than that of the fish antifreeze proteins (Graham et al., 1997). The TmAFP group of proteins are 8.4-13 kDa in size, cysteine and threonine rich (40%) and are characterised by the presence of varying numbers of a 12-residue repeat, Thr-Cys-Thr-X-Ser-X-X-Cys-X-X-Ala-X, where X is any amino acid (Graham *et al.*, 1997; Liou *et al.*, 1999). The structure of a mealworm isoform, Tm 2-14, has been solved using X-ray crystallography and NMR (Figure 1.14 (i)) (Liou et al., 2000). It is an extremely regular right-handed β-helix where each of the 12-residue repeats represents a single turn of the helix, stabilised by intra-turn disulphide bonds. The repeats are stacked side by side such that the structure forms a

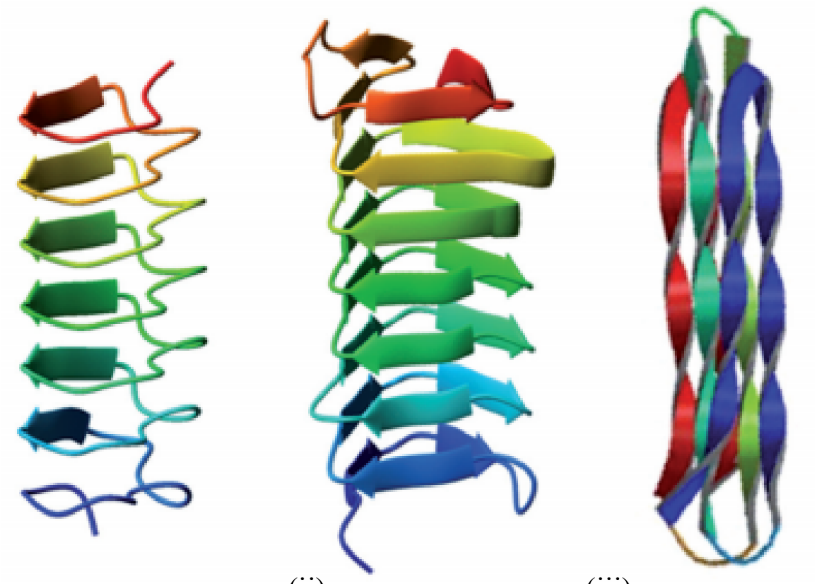


Figure 1.14. Representation of the crystal structure from a selection of insect AFPs. (i) The yellow mealworm AFP, TmAFP, has 12-amino acid repeats containing Thr-Cys-Thr motifs that are folded into a bread loaf shape formed by a right-handed β-helix, stabilised by intra-loop disulphide bonds. (ii) The spruce budworm AFP, sbwAFP, has 15-amino acid repeats containing Thr-X-Thr motifs that are folded into a left-handed β-helix with a triangular cross section, stabilised by inter-loop disulphide bonds and side chains. (iii) The snow flea AFP, sfAFP, has repetitive Gly-Gly-X motif that fold into six alternating parallel and antiparallel helical strands that are stabilised by disulphide and hydrogen bonds. The source of this image is Doucet *et al.* (2009).

regular, two-dimensional array of threonine residues on one side of the molecule. The spacing of the hydroxyl groups of the threonine residues on one side of the protein closely resembles that found between the O-atoms on the prism and the basal plane of the ice, suggesting that this is the ice-binding face of the TmAFP. The importance of threonine in ice-binding was confirmed using site-directed mutagenesis (Marshall et al., 2002).

The antifreeze proteins from the overwintering larvae of the pyrochoroid beetle, *Dendroides canadensis*, are among the most active that are known. The *D. canadensis* AFPs (DAFPs) are 7.3-16.2 kDa in size and are characterised by having 12 or 13-residue repeats with the consensus sequence, Cys-Thr-X_3-Ser-X_5-X_6-Cys-X_8-X_9-Ala-X_{11}-Thr-X_{13}, where X_3 and X_{11} are charged residues, X_5 is either threonine or serine, X_6 is an asparagine or aspartic acid, X_9 is a lysine or asparagine, and X_{13} tends towards alanine in the 13-residue repeats (Duman et al., 1998). The DAFPs activity can be enhanced in the presence of certain co-solutes. Numerous small molecules and macromolecules including glycerol have been reported to enhance antifreeze activity of the DAFP isoform, DAFP-1 (Li *et al.*, 1998a; Duman and Serianni, 2002; Wang and Duman, 2006; Amornwittawat *et al.*, 2008; Amornwittawat *et al.*, 2009). The DAFP shares 40-66% amino acid identity with the TmAFP and in both proteins the two cysteine residues within each repeat form disulphide bonds with each other (Li et al., 1998b). This in addition to the similarity in size and the repeats suggests that they may form similar three-dimensional structures and have a common evolutionary origin (Graham et al., 2007).

The spruce budworm, *Choristoneura fumiferana*, is a freeze-intolerant pest of North America. Its overwintering larvae have a tenfold elevation in glycerol levels and have a TH activity of greater than 5.0 ^{o}C (Han and Bauce, 1993). This TH activity is attributed to a novel 9 kDa, cysteine and threonine rich, hyperactive AFP (sbwAFP) (Tyshenko et al., 1997). The solution structure for sbwAFP has been resolved using both NMR and X-ray crystallography (Figure 1.13 (ii)) (Graether *et al.*, 2000; Leinala *et al.*, 2002). It forms a β-helix with a triangular cross-section and rectangular sides that form stacked parallel β-sheets with approximately 15 residues to each loop (Graether et al., 2000). This protein has no sequence similarity to the TmAFPs or the

DAFPs and it has a characteristic motif sequence, Thr-X-Thr, that is regularly spaced approximately 15-residues apart in a manner that matches the ice lattice on both prism and the basal plane (Graether et al., 2000). Ice-crystal morphology and ice-etching experiments are consistent with this model of the sbwAFP binding to both of these planes and it may explain the greater antifreeze activity of this proteins. The threonine molecules of this three reside repeat have been demonstrated to be essential for ice-binding using site-directed mutagenesis, a change of threonine to leucine resulted in a 80-90% loss of TH activity (Graether et al., 2000).

The snow flea, *Hypogastrura harveyi,* is a primitive arthropod with six legs and no wings. The crude extracts from this flea can have a TH activity of up to 6 ^{o}C. It produces two AFPs (sfAFP), the larger 15.7 kDa isoform has several fold higher activity than the smaller 6.5 kDa isoform, although it is considerably less abundant (Graham and Davies, 2005). Both of the sfAFPs are characterised by their unusual amino acid composition, glycine makes up approximately 45% of the residues and alanine is the second most abundant amino acid (15%). A model of the short isoform shows that it contains Gly-Gly-X repetitive motifs that make helical turns that are regularly disrupted by the presence of four prolines such that the glycine residues face the interior of the folded AFP (Figure 1.13 (iii)). One face of the molecule is hydrophobic, and the other is hydrophilic. Hydrogen and disulphide bonds further stabilise the structure (Lin et al., 2007). This model has since been confirmed by the resolution of the X-ray crystal structure of the sfAFP (Pentelute et al., 2008). It is hypothesised that like other AFPs the flat, hydrophobic face is the ice-binding side that will interact with the highly ordered water molecules found on the ice surface. In support of this an array of highly ordered water molecules were observed on the flat, hydrophobic face of the sfAFP crystal structure, that could bind to the ice (Pentelute et al., 2008).

1.5.3.9 Plant antifreeze proteins

Antifreeze proteins are found in a number of overwintering vascular plants, including ferns, gymnosperms and both dicotyledonous and monocotyledonous angiosperms (Urrutia *et al.*, 1992; Duman and Olsen, 1993; Griffith *et al.*, 1997; Doucet *et al.*, 2000). Plants shown to have AFP actvity include winter and spring rye, winter and

spring wheat, winter and spring canola, winter barley, spring oats, cabbage, maize, spinach, kale, carrot, peach, *Ammopipthanthus mongolicus*, *Solonum dulcamara*, *Lolium perenne* and tobacco (Marentez *et al.*, 1993; Hon *et al.*, 1994; Antikainen and Griffith, 1997; Griffith *et al.*, 1997; Atici and Nalbantoglu, 1999a,b; Yong *et al.*, 2000; Huang and Duman, 2002; Pudney *et al.*, 2003). The AFPs may also be present in different parts of the plants, including seeds, stems, crowns, bark, branches, buds, petioles, leaf blades, flowers, berries, roots, rhizomes and tubers (Urrutia *et al.*, 1992; Duman and Olsen, 1993; Doucet *et al.*, 2000). Antifreeze activity in plants is not a constitutive process, it is only present when plants have been exposed to a period of cold acclimation. Plant AFPs are significantly less active than the other identified AFPs from insects and fish and have a TH activity level of 0.2-0.6 ^{o}C. However, some of the plant AFPs can inhibit ice recrystallisation and have the ability to interact with ice nucleators, which may result in the inhibition of ice nucleation (Zamecnik and Janacek, 1992). Unlike the AFPs from the fish and insects there have been very few reports on the mechanisms that plant AFPs employ to bind to the ice. It is assumed that the plant AFPs bind to the ice using the same mechanisms found in the fish and insects but more sequencing and structural studies are needed to confirm this.

A plant AFP that has been studied and that has its structure solved is the perennial ryegrass, *Lolium perenne* (LpAFP) (Figure 1.15) (Sidebottom et al., 2000). This 11.8 kDa LpAFP is heat stable, hydrophilic, rich in asparagine, valine, serine and threonine, and is characterised by the presence of a semi-conserved, 7-residue repeating motif, X-X-Asn-X-Val-X-Gly, over its entire length (Sidebottom et al., 2000). Wheat (*Triticum aestivum*), hair grass (*Deschampsia antarctica*) and barley (*Horeum vulgare*) all have AFPs that have a homologous domain to the LpAFP. The LpAFP like the other plant AFPs has a very low TH activity level of only 0.5 ^{o}C but has higher ice-recrystallisation inhibition activity than other AFPs (Sidebottom et al., 2000). This property allows the grass to control the growth of the ice crystals rather than preventing them from growing, this may be the critical factor for their survival during freezing. This low TH activity but high ice-recrystallisation inhibition is also seen in the shrub, *Forsythia suspensa*, the legume, *Ammopiptanthus monogolicus* and the carrot AFPs (Worrall *et al.*, 1998; Doucet *et al.*, 2000; Wang and Wei, 2003). The LpAFP was modelled as a β-roll with two ice-binding sites, one on either side of the AFP that are complementary but bind imperfectly to the prism side of the prism face

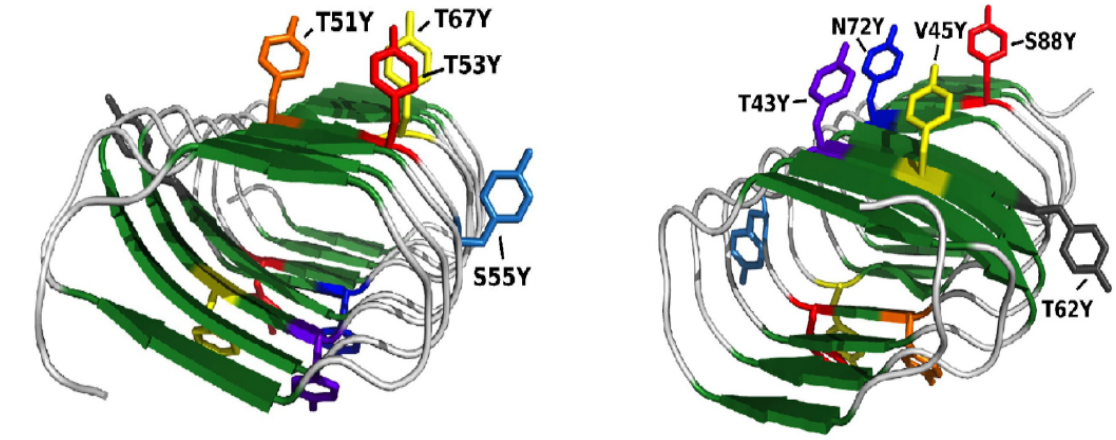


Figure 1.15. Theoretical model for the AFP from the plant *Lolium perenne*, LpAFP. The "a-side" of the protein is shown in (i) and the "b-side" is shown in (ii). The sites that were mutated in testing the location of the ice-binding site of this model are shown in the diagrams by stick representations. This image is modified from Middleton *et al.* (2009).

of the ice (Kuiper et al., 2001). This model has since been validated using site-directed mutagenesis, and it also revealed that there is only one ice-binding site present on the LpAFP (Middleton et al., 2009). A solution structure will be required to further elucidate the mechanism that this AFP uses to bind to the ice.

There have been seven AFPs ranging in molecular mass from 11 to 35 kDa identified in the apoplastic fluid of the winter rye during cold acclimation (Antikainen and Griffith, 1997). They are present in the leaves, have a TH activity of 0.3 ^{o}C and have been shown to have a direct role in freeze tolerance (Marentez et al., 1993). The interesting feature of these plant AFPs is that most of them have homology to pathogenesis-related (PR) proteins (Hon et al., 1995). The N-terminal sequences of individual rye AFPs are nearly identical to β-1,3-glucanases, chitinases or thaumatins (Griffith *et al.*, 1992; Hon *et al.*, 1995). PR proteins isolated from other plants exhibit antifungal activity and as disease resistance is induced by low temperatures in grasses, it is possible that the winter rye AFPs have a dual role in freezing tolerance and disease resistance (Hon et al., 1995). This theory is further supported by the purification of chitinase from cold-acclimated rye, which showed antifreeze activity while a chitinase purified from tobacco had no antifreeze activity (Hon et al., 1995). It is unknown what distinguishes a pathogen-induced PR protein from the cold-induced PR related AFP. It is possible that two types of proteins may have evolved the ability to interact with different molecules associated with their distinct functional roles.

An antifreeze protein has been purified from the carrot, *Daucus carota* (Worrall et al., 1998). It is a 36 kDa, glycosylated, glycoprotein that has sequence similarity to the polygalacturonase inhibitor proteins (PGIPs) but is unable to inhibit polygalacturonase (Worrall *et al.*, 1998; Zhang *et al.*, 2006). The PGIPs are a group of proteins that can specifically inhibit fungal polygalacturonase activity (Collmer and Keen, 1986). They are part of a large family of proteins that are known as the leucine-rich-repeat (LRR) proteins (Kobe and Deisenhofer, 1995). The DcAFP consensus sequence is similar to the LRR motif, it has 10-30 repeats of a approximately 24 amino acid peptide containing regularly spaced leucine residues (Worrall et al., 1998). The DcAFP may have evolved from a LRR region that acquired the ability to bind ice and lost its primary function (Worrall *et al.*, 1998; Meyer *et al.*, 1999). A theoretical

structural model has been proposed for the DcAFP, it includes a 10-loop β-helix containing the leucine residues stacked internally to form a hydrophobic core and has conserved asparagine residues that create the ice-binding sites and are complementary to the prism plane of the ice. This was confirmed using site-directed mutagenesis and it was also found that TH activity is increased with an increase in the length of the asparagine-rich binding site (Zhang et al., 2004). The DcAFP differs from the fish AFPs in that the ice-binding site is strongly hydrophilic, and it is not known how this protein interacts with ice (Griffith and Yaish, 2004).

1.5.3.10 Algal AFPs

Antifreeze activity has been found in the sea diatoms and *Chlamydomonas*-like algae (Graether *et al.*, 2000; Raymond and Knight, 2003; Raymond, 2009).

A secreted 25 kDa AFP was isolated from the diatom, *Navicula glaciei* (Janech et al., 2006). Four isoforms of a homologous AFP were isolated from the *Fragilariopis cylindrus* (Krell et al., 2008). Both of these proteins are likely homologs of the AFPs from the snow mould, *Thalassiosira pseudonama*, and as a group they represent a new class of AFPs that are distinct from the other identified AFPs. These proteins contain a high proportion of non-polar (40%) and polar (20%) amino acids. Secondary structure predictions reveal that these proteins have two α-helices and several β-sheets, the latter separated in part by short, coiled regions, which may provide the basis for a β-helical tertiary structure (Krell et al., 2008).

Diatom AFPs are strong recrystallisation inhibitors and a semi-pure AFP solution from *N. glaciei,* was found to reduce the haemolysis of frozen human red blood cells. This indicates that these proteins protect the cell membrane from injury by surrounding ice (Janech et al., 2006). Several isoforms of a different type of AFP were found in an Antarctic intertidal green alga, the *Chlamydomonas*-like strain, CCMP681 (Raymond, 2009). This protein is very effective in ice recrystallisation inhibition and at retaining brine in the ice, helping to preserve a liquid environment in the ice (Raymond, 2009). The four isoforms present in CCMP681 have six highly conserved Thr-Phe-Thr and Thr-Trp-Thr motifs. This Thr-X-Thr motif is present in the ice-binding site of the spruce budworm, suggesting that it may also form the ice-

binding site of CCMP681 AFP and that it may bind to the ice in a similar way (Tyshenko *et al.*, 1997; Graether *et al.*, 2000).

1.5.3.11 Fungal AFPs

Fungal AFPs have been isolated from the basidomycetes, *Coprinus psychromorbidus* and *Typhula ishikariensis* and more recently in the ascomycete, *Antarctomyces psychrotrophicus* (Hoshino *et al.*, 2003; Xiao *et al.*, 2010). The AFPs from *C. psychromorbidus* and *T. ishikariensis* are approximately 22 and 23 kDa, respectively. Recombinant AFP from *T. ishikariensis,* TisAFP, was produced and compared to the 28 kDa AFP from *A. psychrotrophicus,* AnpAFP (Hoshino *et al.*, 2003; Xiao *et al.*, 2010). The AnpAFP has a maximum TH activity of 0.42 ^{o}C, while that of TisAFP is up to 1.9 ^{o}C. The AnpAFP TH activity is comparable to the level seen in fish and this protein also shows ice recrystallisation inhibition activity (Davies and Hew, 1990). The TisAFP is more similar in ice crystal growth and in TH activity to the insect and bacterial AFPs (Gilbert *et al.*, 2005; Scotter *et al.*, 2006). The amino acid sequences of both fungal AFPs contain high level of threonine residues, as these are implicated in ice binding in several other AFPs it has been suggested that it is also important in AnpAFP and TisAFP ice-binding (Wen and Laursen, 1992; Chao *et al.*, 1994; Jia *et al.*, 1996; Haymet *et al.*, 1998; Zhang and Laursen, 1998). Although there is similarity in the N-terminal residues between AnpAFP and TisAFP the differences in activity suggest that these two AFPs might have evolved independently (Xiao et al., 2010).

1.5.3.12 Yeast AFPs

A number of psychrophilic yeasts have been isolated from alpine regions, polar ice and glaciers (Nakagawa *et al.*, 2006; de Garcia *et al.*, 2007; Shivaji *et al.*, 2008; Turchetti *et al.*, 2008). A yeast AFP has recently been purified from an isolate, *Leucosporidium* sp. AY30 (Lee et al., 2010). This is a 26 kDa, glycosylated secretory AFP. This protein is the first AFP to be isolated from yeast and has both ice recrystallisation inhibition activity and TH, with an activity level of 0.7 ^{o}C. This protein shows high sequence similarity with the AFPs from fungi, algae and bacteria, it may be possible that, like the fish Type II AFPs, it may have arose through horizontal gene transfer (Graham *et al.*, 2008a; Lee *et al.*, 2010).

1.5.3.13 Bacterial AFPs

Antifreeze proteins have been cloned and characterised from several bacterial genera: *Flavobacteriaceae*, *Colwellia*, *Marinomonas primoryensis*, *Moraxella* sp. and *Pseudomonas putida* (Yamashita *et al.*, 2002; Gilbert *et al.*, 2005; Raymond *et al.*, 2007; Garnham *et al.*, 2008; Raymond *et al.*, 2008). In general, bacteria show modest levels of TH activity, but like some of the plant AFPs, they have high ice recrystallisation inhibition activity. This indicates that the bacterial AFPs may decrease cellular damage caused by ice recrystallisation rather than preventing freezing completely (Gilbert *et al.*, 2004; Gilbert *et al.*, 2005). In sea ice or glaciers, ice recrystallisation inhibition activity may help conserve the boundaries between the ice grains where these organisms may be located (Raymond et al., 2008).

An AFP from the Antarctic bacterium, *M. primoryensis* has been studied. This bacterium produces a very large (approximately 1.5 MDa) Ca^{2+}-dependent AFP (MpAFP) that contains two highly repetitive segments that are divided over five distinct regions (Gilbert *et al.*, 2005; Garnham *et al.*, 2008). The antifreeze activity of MpAFP is located in region IV (MpAFP_RIV). This 34-kDa region that has thirteen 19-residue repeats. In contrast to most bacterial AFPs it is a hyperactive AFP, with a TH of 2 ^{o}C. Recently, a X-ray crystal structure has been determined for the MpAFP_RIV (Figure 1.16) (Garnham et al., 2011). It is the largest AFP structure determined to date, it folds as a Ca^{2+}-bound parallel β-helix with an extensive array of ice-like surface waters that are anchored by hydrogen bonds directly to the polypeptide backbone and adjacent side chains (Garnham et al., 2011). This structure gives a more detailed molecular explanation on the mechanism that AFP binding may occur. The relative hydrophobicity of an AFPs ice-binding site can order its water molecules via the hydrophobic effect into an ice-like lattice that is anchored onto the protein by hydrogen bonds. These “anchored waters” then allow an AFP to bind ice by matching to a specific plane(s) of the ice crystal (Garnham et al., 2011).

1.5.4 Ice nucleator proteins

Ice nucleation describes the initiation of the phase transition of water from a liquid to ice. Ice nucleation may arise by two mechanisms. Homogenous nucleation occur where the water molecules spontaneously aggregate without the help of another structure. The probability of this occurring increases as the temperature decreases and

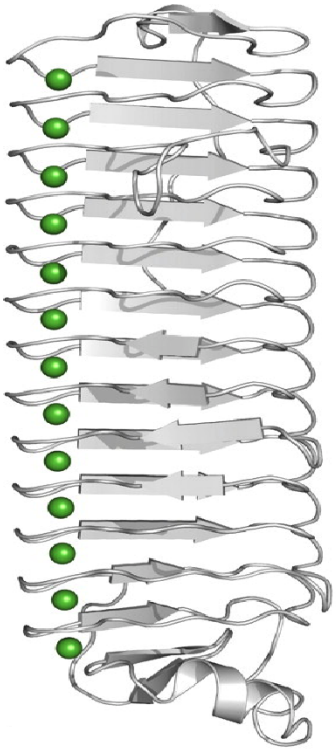


Figure 1.16 A representation of the structure of region IV of the AFP from the Antarctic bacterium, *M. primoryensis* known as MpAFP-RIV. Calcium ions are shown as the green spheres. The source of this illustration is Garnham *et al.* (2011) .

with the length of chilling. In biological systems, ice nucleation is believed to occur by heterogeneous nucleation at high subzero temperatures. This occurs when a structure other than water forms the template upon which ice will begin to form (Knight, 1967). A structure than can cause ice nucleation is termed an ice nucleator. Biological ice nucleators are typically proteins and these have been identified in organisms from across all phyla. In addition to proteinaceous nucleators, crystalloid inorganic compounds such as calcium phosphate, calcium carbonate, potassium phosphate or uric acid may function as nucleators. The insect *Eurosta solidaginis* uses calcium phosphate crystals to induce ice nucleation (Mugnano et al., 1996). The ice nucleators have been categorised according to their activity, Type I ice nucleators nucleate between -2 and 5 ^{o}C, Type II nucleate between -5 and -7 ^{o}C and Type III nucleate between -7 and -10 ^{o}C (Yankofsky et al., 1981). The actual activity of an ice nucleator depends on a number of factors, including the size of the aggregates that the nucleator forms and whether other substances such as glycerol are present. It is possible that glycerol can increase membrane fluidity and that this would allow larger aggregates to form.

The most active and best studied biological ice nucleators are INPs from the ice nucleating bacteria. They were first discovered when the potent ice nucleators on decaying leaves were found to be caused by the gram negative bacterium, *Pseudomonas syringae* (Schnell and Vali, 1972; Maki *et al.*, 1974). The susceptibility of plants infected with *P. syringae* to frost was shown to be caused by ice nucleation at high temperatures, rather than by a reduced frost resistance caused by the infection itself (Arny et al., 1976). The INPs produced by *P. syringae* can allow ice to form at -2 ^{o}C, classifying them as Type I nucleators. It appears that in *P. syringae* the primary role of the INPs is adaptive: the INPs induce frost damage in plants, thereby giving the bacteria access to a supply of nutrients from the plant (Buttner and Amy, 1989).

INPs have also been reported in freeze-tolerant insects such as the hornet (*Vespula maculata*), the cranefly (*Tipula trivittata*) and the pyrochoroid beetle (*Dendroides canadensis*) (Duman *et al.*, 1984; Duman *et al.*, 1991; Wu *et al.*, 1991). In insects the role of the INPs appears to differ greatly from those of bacteria. In freezing-avoiding insects the INPs are surprisingly thought to increase the survival in sub-zero temperatures. This is possible as they can induce freezing slowly forming small ice

crystals at temperatures above which sudden spontaneous freezing would occur, allowing time for survival mechanisms to be employed. In the freezing-tolerant insects it would seem that ice nucleators would be unnecessary, as they can survive spontaneous freezing at low temperatures, but they may still be present. The 70 kDa INP purified from *D. canadensis* has a primary function of enhancing the activity of the AFP, DcAFP, and its suggested that the INPs are produced to ensure that ice is formed only in the compartments of the body where it can be tolerated (Zachariassen and Hammel, 1976; Wu *et al.*, 1991).

1.5.5 Gene and protein induction in response to low temperature stress

There are several other genes and proteins that may be associated with the freezing response in addition to the cryoprotectants and ice-binding proteins. Carrasco *et al.* (2011) have constructed a cross-species compendium of differentially expressed proteins/gene products in response to low temperature based on 39 primary research papers, 20 of which were proteomics based, 18 microarray based, and one that dealt with cDNA library screening in yeast. These authors identified 2,030 differentially regulated gene products and proteins, of which 1,353 were up-regulated while 549 were down-regulated in response to low temperature. Despite this large number of differentially regulated gene products, the authors could only identify 58 proteins/gene products with consistent responses across species, 50 that were up-regulated and eight that were down-regulated and there were proteins/gene products whose response to low temperature was inconsistent (varied across species, up-regulated or down-regulated depending on the species). It is possible that the relatively low number of gene products with consistent responses may be caused by the diversity of taxa in the compilation. The numbers of species represented in the database were as follows: 9 bacteria, 1 yeast, 11 plants, 1 nematode (*Plectus murrayi),* 7 insects and 5 vertebrates. Carrasco *et al*. also point out that these ‘‘inconsistently regulated’’ proteins/gene products may represent species-specific responses, treatment specific responses (long term winter acclimatization versus short term laboratory cold acclimation), or temperature specific responses (freezing temperatures versus sub-zero non-freezing temperatures). Table 1.1 gives a summary of the proteins that are “consistently up-regulated” in response to cold stress in the dataset of Carrasco *et al*. (2011), their putative functions and the taxa in which they have been found to be differentially expressed. The consistently up-regulated genes give an

Table 1.1. Proteins which are "consistently up-regulated" in response to cold stress, their putative functions and the taxa in which they have been found to be differentially expressed (based on data in Table 4 and Supplemental Material 1 from Carrasco *et al*., (2011)).

Putative Function	Protein	Taxa
Signalling	Phosphoinositide-specific phospholipase	Yeast, Plantae
	Protein phosphatase 2C	Plantae
	Putative protein kinase DC2	Insecta
Transcription	bZIP DNA-binding protein	Plantae
	RNA polymerase	Bacteria, Yeast
	Transcription elongation factor	Bacteria
	Transcription termination factor Rho	Bacteria
	Cold shock protein (CspA-family homologue)	Bacteria
Metabolic, Catabolic and Biosynthetic pathways	Anthranilate synthase	Bacteria, Yeast
	Aspartate carbamoyltransferase	Bacteria
	Cathepsin	Nematoda*, Insecta
	Citrate synthase	Bacteria, Plantae
	DNA gyrase subunit A	Bacteria
	Electron transfer flavoprotein α-subunit	Bacteria
	Enolase	Bacteria, Plantae
	Fructose-bisphosphate aldolase	Plantae
	Nucleoside diphosphate kinase	Plantae
	Lysophospholipase	Bacteria, Plantae
	Phosphoglycerate kinase	Bacteria, Plantae
	Phosphoglycerate mutase	Bacteria
	Putative 2,3-biphosphoglycerate-independent	Plantae
	Phosphoglycerate mutase	Bacteria
	Sucrose synthase	Plantae
	Threonine synthase	Bacteria, Yeast
	Trehalase precursor	Insecta
	Trehalose-6-phosphate synthase	Plantae, Insecta
	Xyloglucan endo-1, 4-beta D-glucanase precursor	Plantae
Detoxification	Aldo/keto reductase	Bacteria
	Ferritin	Bacteria, Plantae, Amphibia, Reptilia
	Glyoxalase	Plantae
Antioxidant Activity	Superoxide dismutase	Bacteria, Plantae, Nematoda*, Insecta, Mammalia
	Peroxiredoxin	Plantae, Nematoda*, Reptilia
	Thioredoxin	Bacteria, Plantae, Amphibia
Protein folding	FKBP-type peptidyl-prolyl cis-trans isomerase	Bacteria
	Peptidyl-prolyl cis–trans isomerase	Bacteria, Plantae, Insecta
Molecular chaperone	Small heat shock protein	Nematoda*, Insecta
Cytoskeletal remodelling	Actin depolymerization factor	Plantae, Insecta
Structural protein	Outer membrane lipoprotein	Bacteria, Plantae
Miscellaneous	Leucine rich repeat (LRR) protein	Nematoda*, Insecta
	Thaumatin like protein	Plantae
	Dehydrins	Plantae

*The genes for *Plectus murrayi* in this database were differentially expressed in response to desiccation stress, not cold stress.

indication of the broadly conserved physiological responses to low temperature. These include up-regulation of the pathways involved in transcription, metabolism, biosynthesis and transport. Proteins involved in detoxification, cellular repair, protein folding and cytoskeletal remodelling also form an important part of the physiological response to low temperature.

1.5.5.1 Cold-shock proteins

Most of the research on cold shock proteins (Csps) has concentrated on *Escherichia coli*. In *E. coli* a downshift in temperature causes a transient inhibition of most protein synthesis, resulting in a growth lag called the acclimation phase. It is during this acclimation phase that the Csps are induced (Jones et al., 1987). When the acclimation phase is over the Csp protein synthesis returns to normal levels, general protein synthesis returns and the cell resumes growth although at a reduced rate (Thieringer et al., 1998). In certain bacteria such as *Bacillus subtilis* and *Lactobacillus lactis* this growth lag does not occur, these bacteria may be constitutively cold adapted to low temperature growth even when growing at their optimal temperature (Phadtare and Severinov, 2010). The main Csps of *E. coli* and their roles are listed in Table 1.2. They are heavily involved in cellular transcription and translation in a cold state and the Csps, RbfA and CsdA are even capable of binding to ribosomes, converting them to a cold-adapted state that can translate non-cold-shock mRNAs (Jones and Inouye, 1996). There are only two well-characterised Csps in mammals, Cirp (cold-inducible RNA binding protein) and Rbm3 (RNA binding motif protein 3). It is thought that they are also involved in modulation of transcription and translation. Both are highly similar and belong to the glycine-rich RNA-binding family (Al-Fageeh and Smales, 2006). Although the exact function(s) of these cold-inducible proteins is unknown, it is thought that they modulate translation and function as RNA chaperones that facilitate translation upon the perception of cold stress (Al-Fageeh and Smales, 2009).

1.5.5.2 Cold-regulated genes

Work on *Arabidopsis thaliana* has revealed that plants produce a superfamily of cold-regulated (COR) genes that can confer cold and freezing tolerance in plants (Thomashow, 1999). The molecular analysis of genes of this superfamily shows that

Table 1.2. A selection of the cold-shock genes and gene products induced in response to low temperatures in by *Escherichia coli*. These genes have a tendency to be involved in transcription and translation at low temperatures.

Gene product	Function/Properties	Reference
AceE	Pyruvate dehydrogenase	(Jones and Inouye, 1994)
AceF	Pyruvate dehydrogenase	(Jones and Inouye, 1994)
ClpB	Protein chaperone	(Phadtare and Inouye, 1999)
CspA	RNA chaperones	(Goldstein et al., 1990)
CspB	RNA chaperones	(Jones and Inouye, 1994)
CspG	RNA chaperones	(Nakashima et al., 1996a)
CspI	RNA chaperones	(Wang et al., 1999)
DeaD	RNA helicase	(Toone et al., 1991)
DnaA	RNA binding and replication/Transcriptional regulator	(Atlung and Hansen, 1999)
GyrA	DNA-binding	(Jones et al., 1992)
H-NS	Transcription repressor/DNA supercoiling	(Dersch et al., 1994)
HscB	DnaJ homologue	(Lelivelt and Kawula, 1995)
Hsc-66	DnaK homologue (Hsp70-type protein chaperone)	(Lelivelt and Kawula, 1995)
HUβ	DNA supercoiling	(Giangrossi et al., 2002)
IF1	Translation initiation factor/ RNA-binding	(Gualerzi et al., 2003)
IF2	Translation initiation factor/Protein chaperone	(Caldas et al., 2000)
IF3	Translation initiation factor/RNA binding	(Gualerzi et al., 2003)
NusA	Transcription factor	(Jones and Inouye, 1994)
OtsA	Trehalose phosphate synthase	(Kandror et al., 2002)
OtsB	Trehalose phosphatase	(Kandror et al., 2002)
PNPase	Polynucleotide phosphorylase	(Yamanaka and Inouye, 2001)
pY	Translation (A-site) inhibitor	(Nakashima et al., 1996b)
RbfA	Ribosome-binding factor	(Dammel and Noller, 1995)
RecA	Homologous recombination	(Jones and Inouye, 1994)
Trigger Factor	Multiple stress protein	(Kandror and Goldberg, 1997)

they contain a regulatory *cis* element, known as CRT (C-repeat) or DRE (dehydration -responsive element), that controls gene expression during cold and desiccation stress (Baker *et al.*, 1994; Yamaguchi-Shinozaki and Shinozaki, 1994). This element has a consensus sequence, CCGAC that is located in the promoter of many cold and desiccation responsive genes as well as the COR genes. The transcription factor CBF1 induces the expression of the COR genes *cor6.6*, *cor15a*, *cor47*, and *cor78*. Overexpression of this transcription factor improves the freezing tolerance of non-acclimated *A. thaliana* plants and chilling sensitive tomatoes (Jaglo-Ottosen *et al.*, 1998; Hsieh *et al.*, 2002). *A. thaliana* transformation with another transcription factor, the DREB1A gene under the control of a strong constitutive promoter resulted in increased expression of the downstream genes, *rd29A*, *rd17*, *cor6.6*, *cor15a*, *erd10* and *kin1* with a significant increase in tolerance to water, salinity and freezing stress (Kasuga et al., 1999). Overexpression of the transcription factor CBF3 also increases freezing tolerance (Al-Fageeh and Smales, 2006) and it causes elevated levels of proline and soluble sugars, cryoprotectants commonly associated with freezing survival (Gilmour et al., 2000). The components of the *A. thaliana* cold-response pathway were also found in *Brassica napus* and other plant species (Jaglo et al., 2001). Constitutive overexpression of the *A. thaliana* CBF genes in *B. napus* plants induced the expression of genes that were orthologs of those associated with CBF expression in *A. thaliana* and increased freezing tolerance (Jaglo et al., 2001). The plant hormone abscisic acid (ABA) accumulates in response to a number environmental stresses including cold. Gene expression analysis shows that ABA is involved in the induction of the expression of different COR genes in *A. thaliana* via the cis CRT/DRE elements and the CBF1-3 transcription factors (Knight et al., 2004).

LEA (late embryogenesis abundant) proteins are hydrophilic proteins which are synthesized during the desiccation and maturation of seeds, and are also induced in response to drought, high salinity or low temperature in the vegetative tissues of a number of plant species (Tunnacliffe and Wise, 2007). Subsequently LEA group 3 proteins have been isolated from anhydrobiotic nematodes, rotifers, tardigrades, and the anhydrobiotic insect *Polypedilum vanderplanki* (Burnell and Tunnacliffe, 2011). One of the LEA group II proteins is the dehydrins (DHN), which so far appear to be plant-specific. Dehydrin had 10 entries in the cold stress in the dataset of Carrasco *et al*. (2011) across five different plant species and three different types of treatments.

Dehydrins are multifunctional proteins and among the many functions ascribed to them are the prevention of protein denaturation, stabilization of phospholipid membrane, stabilization of the cytoskeleton and binding to nucleic acids (Hara, 2010).

In vertebrates a group of early response genes and signal transduction pathways are expressed to counteract all stresses. These genes and pathways are likely to be upstream regulators of genes involved in stress. The c-Jun NH_2-terminal kinase (JNK) and the p38 kinase pathway are often activated in many vertebrate species, including frogs, mammals, turtles and mammals (Cowan and Storey, 2003). The JNK transcription factors are responsible for initiating many transcriptional responses during stress. Both JNK and p38 have been shown to be activated in a organ specific manner in response to freezing and thawing in the freezing-tolerant wood frog, *R. sylvatica* (Greenway and Storey, 2000). In animals, a hormonal stress response may also occur. The endocrine responses are mainly coordinated by catecholamines and glucocorticoids. Cold exposure in tilapia, *Oreochromis aureus*, causes an accumulation of catecholamines and cortisol (Chen et al., 2002). During stress, catecholamines stimulate cardiac output leading to an increase in ventilation rate, branchial blood flow, and oxygen transport (Kassahn et al., 2009). Cortisol is important in coordinating the physiological and transcriptional responses to stress. This occurs by stimulating gluconeogenesis in the liver and lipolysis increasing plasma free fatty acid levels (Kassahn et al., 2009).

1.5.5.3 Heat shock proteins

Heat shock proteins (Hsps) function as molecular chaperones during periods of stress by binding to other proteins, preventing the negative effects of misfolding of proteins and assisting the return of proteins to their native conformation when favourable conditions return. Generally, heat shock proteins are associated with high temperatures but several heat shock proteins have been found to be significantly up-regulated in response to cold stress, desiccation and anoxia (Parsell and Lindquist, 1993; Wu, 1995; Feder and Hofmann, 1999). The level of the heat shock response that occurs in response to cold depends not only on the magnitude and duration of the stress, but also on acclimation, the seasons and previous exposure to stress (Hofmann, 2005; Kassahn *et al.*, 2009).

There are five conserved families of Hsps, Hsp100, Hsp90, Hsp70, Hsp60 and small Hsp (sHsp). The expression of Hsps is generally under the regulation of the transcription factor, heat shock factor 1 (HSF1). HSF1 is normally bound to the Hsps but when protein damage occurs during stress it may migrate to the nucleus and induce the transcription of genes with a heat shock element, such as Hsps (Kim *et al.*, 1997; Kline and Morimoto, 1997). HSF1 requires phosphorylation for activity, which is controlled by the mitogen-activated protein kinases.

In response to low temperatures the Hsps are thought to be induced but the number of examples of this are limited. In several polar marine fish species, the *Hsp70* and the *hsc71* genes are constitutively expressed (Place and Hofmann, 2004). In the insects Hsps are developmentally regulated and up-regulation occurs during diapause. This is believed to be a major factor in contributing to cold-hardiness in overwintering insects. The flesh fly, *Sacrophaga crassipalpis,* up-regulates Hsp23, Hsp70 and a small Hsp during its overwintering pupal diapause (Yocum *et al.*, 1998; Rinehart *et al.*, 2000). Silencing of Hsp23 and Hsp70 in these flies do not alter the decision whether to enter diapause, but has a profound effect on the pupa's tolerance to low temperatures (Rinehart et al., 2007). In plants the sHsps are found to be the most prevalent. Similarly to other Hsps they are expressed in response to a wide range of stresses that include low temperature (Sabehat *et al.*, 1998; Wang *et al.*, 2003). Two tomato sHsps, *tom66* and, the Hsp21 gene, *tom111,* were induced by low temperature in pre-heated fruits (Sabehat et al., 1998). In plants Hsps are particularly interesting as they have potential to be used to create stress tolerant transgenic plants. A transgenic carrot plant which constitutively expressed the carrot Hsp17.7 gene was found to have more cold tolerance than the controls (Malik et al., 1999). However, constitutive expression of a Hsp needs to be done with caution, since expression of Hsps during nonstress conditions can lead to deleterious effects in some cases, including reduced or complete stop of growth (Malik et al., 1999).

1.5.5.4 Alteration of cell membrane lipid composition

When the cell membrane is below a critical temperature, known as the transition temperature, the membrane will become reorganised from the efficient, fluid and flexible liquid crystalline phase to the rigid solid gel phase (Figure 1.17). If

membranes are at low temperatures for a prolonged period of time, then a transition to the gel phase will spread along the whole membrane, causing loss of elasticity that in turn leads to loss in functionality of membrane proteins and membrane rupturing resulting in leakage of water, ions and metabolites from the cell. This phase transition is considered one of the most important causes of cold injuries that occurs under non-freezing conditions (Drobnis et al., 1993).

Many organisms have developed mechanisms to compensate for the phase transition by increasing the concentration of membrane phospholipids containing unsaturated fatty acids. Phospholipids with unsaturated fatty acids have lower melting points and are more flexible than those that contain saturated fatty acids. This remodelling of cell membrane lipids in response to changes in temperatures has been termed homeoviscous adaptation (HVA), and it was first demonstrated in *E. coli* (Sinensky, 1974). Fatty acid desaturase enzymes play an essential role in HVA by increasing the ratio of unsaturated to saturated fatty acids in the cell membranes and their expression has been shown to be increased under low temperatures in bacteria, fish, insects and plants (Tiku *et al.*, 1996; Sakamoto and Bryant, 1997; Vega *et al.*, 2004; Kayukawa *et al.*, 2007).

In plants, some isoforms of the enzyme glycerol-3-phosphate acyltransferase (GPAT) are also produced in response to chilling (Murata et al., 1992). The GPAT enzyme is associated with changes in fatty acid composition of the plastid membranes (Murata et al., 1992). Transgenic plants that overexpressed a gene for GPAT from *E. coli* showed lower levels of saturated fatty acids and have tolerance to chilling (Wolter et al., 1992). An increase in phospholipids and free sterols, along with a decrease in steryl glucosides, acylated steryl glycosides, and glucocerebrosides is also seen in plants during cold acclimation (Miguel *et al.*, 1993; Thieringer *et al.*, 1998).

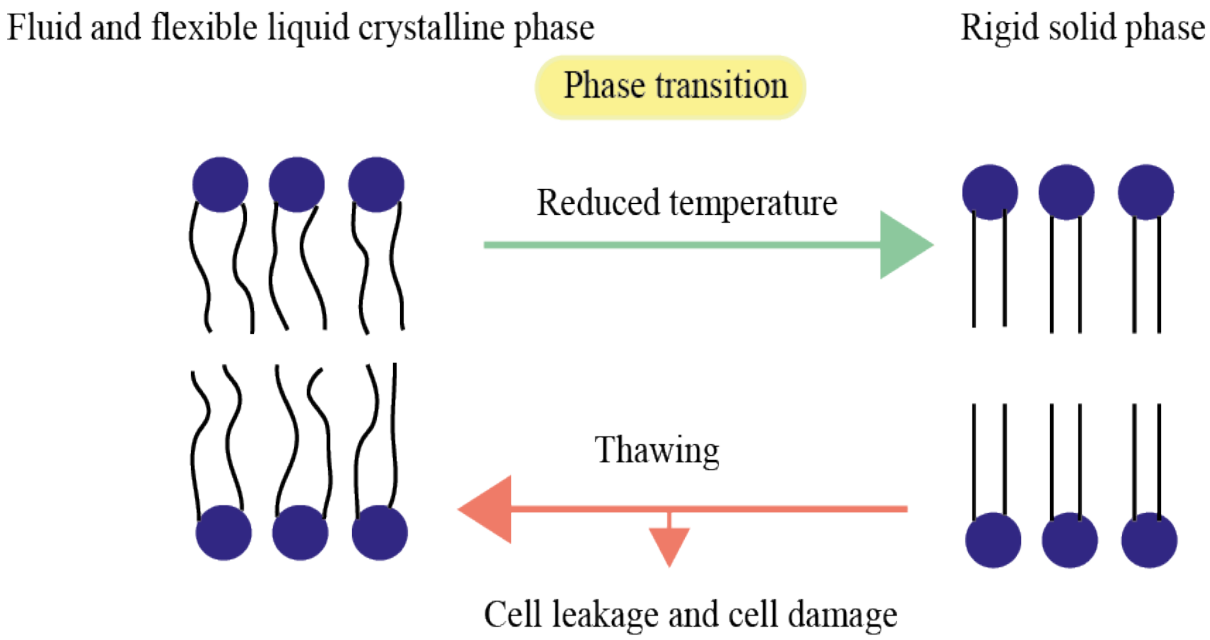


Figure 1.17. The phase transition in cell membranes during freezing. When the temperature is reduced the cell membrane changes from a fluid, flexible crystalline phase to a rigid, solid phase. This causes cellular damage, thawing causes further cell leakage and damage.

1.6 Relationship between freezing and desiccation survival

Desiccation tolerance is the ability to survive periods of loss of water. The most extreme version of desiccation tolerance is anhydrobiosis, a term used to describe the condition whereby desiccated organisms enter a state of suspended animation in order to survive. Anhydrobiotes can lose between 95-99% of their water during this process and upon rehydration they resume normal biological activity (Crowe and Crowe, 1992; Danks, 2000). There are three invertebrate phyla that are capable of anhydrobiosis at all life stages, the tardigrades, nematodes and rotifers (Tunnacliffe and Lapinski, 2003). Desiccation generally causes oxidative stress and damage to proteins, lipids, nucleic acids and the phospholipid bilayer. Anhydrobiotes survive these stresses through the synthesis of disaccharides such as trehalose, the down-regulation of metabolism, the expression of proteins such as LEAs, antioxidants in the form glutathione, up-regulation of general stress proteins such as Hsps and the partitioning of amphiphilic solutes into cell membranes (Burnell and Tunnacliffe, 2011).

As ice forms extracellularly, the major stress perceived by cells during freezing is a reduction in water as it flows out of the cells increasing the concentration of intracellular solutes. Since freezing stress can result in cellular desiccation, it seems logical that the protective mechanisms that occur during freezing may be related to those that protect against desiccation stress. A connection between desiccation and cold tolerance has been identified in several species, including frogs, nematodes and insects (Wharton and Barclay, 1993; Storey and Storey, 1996; Clark and Worland, 2008).

1.7 Freezing survival in nematodes

Nematodes are a group of organisms that are often exposed to periods of environmental stresses. The occurrence and duration of a change in optimal conditions can be unpredictable, so several nematodes have evolved a number of physiological and biochemical adaptations that allow their survival. As low temperatures occur over much of the Earth, free-living and parasitic nematodes may be exposed to subzero temperatures that may involve the risk of freezing for short or extended periods of time. In nematodes the body contents are frozen by inoculative

freezing via the openings in the body wall, primarily the excretory pore (Wharton and Ferns, 1995). There has been very little investigation into the mechanisms that different nematodes employ to survive freezing. The Antarctic nematode, *Panagrolaimus davidi,* has been the main model for studying nematode cold tolerance (Wharton, 2003).

Certain nematodes from the Antarctic appear to be specifically adapted to living in low temperatures. *Plectus antarcticus* lays eggs and completes its life cycle at lower temperatures than nematodes from temperate regions while *Scottnema lindsayae* has a high development rate and better survival at 10 ^{o}C as opposed to 15 ^{o}C (Wharton and Ferns, 1995). In contrast, *P. davidi* and the temperate nematodes used in this study grow best at 20-25 ^{o}C (Wharton, 1997; Shannon *et al.*, 2005). There are three main strategies of cold tolerance that occur in temperate nematodes, freeze avoidance, where they supercool; freeze tolerance, where they can withstand ice formation; and cryoprotective dehydration, where they dehydrate because the vapour pressure of supercooled water within their bodies is higher than the surrounding ice (Wharton, 2003).

The model species *P. davidi* is a cold-tolerant nematode that can survive temperatures down to -80 ^{o}C (Wharton and Brown, 1991). It is capable of withstanding 82% of its water freezing, this is a considerably higher ice content than the majority of freezing-tolerant animals (Storey and Storey, 1996). Complete freezing may occur within 0.21 seconds and it is the only animal that has been shown to survive intracellular freezing (Wharton and Ferns, 1995). Although this nematode is freezing tolerant it seems that at high subzero temperatures and slow freezing rates, *P. davidi* does not freeze, the supercooled water within the nematode is lost to the surrounding ice, preventing freezing and allowing cryoprotective dehydration (Wharton, 2003). There is little evidence of this occurring in other nematodes, but recently a typical free-living nematode *Panagrellus redivivus* was shown to survive by cryoprotective dehydration in addition to freezing tolerance (Hayashi and Wharton, 2011).

A period of acclimation may enhance nematode survival to cold stress. Freezing tolerance is improved in the larvae of the plant-parasitic nematode *Meloidogyne hapla* after acclimation at low temperatures (Forge and MacGuidwin, 1990). The

supercooling ability of eggs of the animal-parasitic nematode *Nematodirus battus* increased after acclimation (Forge and MacGuidwin, 1990). The nematodes, *P. redivivus* and *Steinernema carpocapsae* also show improved cold tolerance when placed a low temperatures for a few days before freezing (Jagdale and Grewal, 2003; Hayashi and Wharton, 2011).

Similarly to other cold tolerant organisms some nematodes produce cryoprotectants in response to cold. The principal cryoprotectants produced by *P. davidi* are trehalose and glycerol (Wharton et al., 2000). The entomopathogenic nematodes, *Steinernema feltiae, Steinernema carpocapsae* and *Steinernema riobrave* all have increased levels of trehalose in response to cold acclimation, however this also occurs with heat stress indicating that trehalose accumulation is a general response of nematodes to thermal stress (Jagdale and Grewal, 2003). *P. redivivus* produces glycerol as an end point of anaerobic catabolism but there does not appear to be any correlation between glycerol concentration and supercooling points in this nematode (Hayashi and Wharton, 2011). Nematodes from the genus *Panagrolaimus* have increased trehalose when preconditioned to desiccation, as there is a strong correlation between freezing and desiccation in several organisms it is likely that trehalose is also elevated in response to cold (Shannon et al., 2005).

It has been suggested that nematodes may also produce ice-binding proteins. In *P. davidi* there are no INPs and there is little to no TH activity in protein extracts but it does appear to produce a recrystallisation inhibition protein (RIP) (Wharton, 2003; Wharton *et al.*, 2005). This RIP is heat stable, affected more by acid than alkaline pH, is not calcium dependent and is not affected by reagents that target carbohydrate residues or sulphydryl linkages. It has not been isolated as it is only present at low concentrations (Wharton et al., 2005). The nematodes *P. rigidus*, *Rhabditophanes* sp., *S. carpocapsae* and *P. redivivus* all show recrystallisation inhibition activity so these likely also contain a RIP (Smith et al., 2008). A molecular analysis of the cold adapted nematode *Plectus murrayi* found a Type II AFP, which showed high similarity to that of the Atlantic herring, *Clupea harengus* (Adhikari et al., 2009). Therefore some nematodes may produce more potent ice-binding proteins in combination with RIPs to survive freezing stress.

1.8 Objectives of this project

The studies on cryobiosis in nematodes to date have focused on those from the Antarctic McMurdo dry valleys. *P. davidi* has been extensively used as a model for freezing tolerance (Wharton and Brown, 1991; Wharton and Ferns, 1995; Wharton, 1997; Wharton *et al.*, 2000; Wharton *et al.*, 2003) but very little work has focused on the freezing tolerance of other *Panagrolaimus* strains and species. Nematodes from the genus *Panagrolaimus* have been found be anhydrobiotic (Shannon et al., 2005). Preliminary work at NUI Maynooth (Shannon, 2007) showed that strains and species of the genus *Panagrolaimus* also had freezing tolerance. Thus the main objective of this project was to compare the relationship between biogeography, phylogeny, cryobiotic and anhydrobiotic phenotypes of these nematodes to get a better understanding of the molecular mechanisms underlying these phenotypes. As the project developed, the specific objectives became focused on the following:

1. To determine the cryobiotic and anhydrobiotic survival of fifteen *Panagrolaimus* species and strains from polar, subpolar and temperate regions. To compare these survival phenotypes to their ability to prevent ice growth and phylogenetic relationships.
2. To determine cryobiotic and anhydrobiotic survival of ten new *Panagrolaimus* species and strains from tropical regions and to compare these phenotypes to their phylogenetic relationships.
3. To estimate the divergence times for *Panagrolaimus* sp. using the molecular clock approach to provide insight on the dispersal of *Panagrolaimus* sp. and the evolution of cryobiosis and anhydrobiosis.
4. To implement an affinity purification method and to use this method to attempt to purify and identify an ice-binding protein from *P. superbus*.

As an ice-binding protein was not successfully identified an additional aim was added to the project:

5. To determine the genes up-regulated in *P. superbus* in response to a period grown at reduced non-freezing temperatures using the next generation sequencing method RNA-seq.

Chapter II Materials and Methods

2.1 Materials

2.1.1 Chemicals

Chemicals were obtained from Sigma-Aldrich Co. Ltd. (Gillingham, UK), Thermo Fisher Scientific Ltd. (Massachusetts, US), Novagen – Division of Merck/EMD (Wisconsin, US), BioRad Laboratories Ltd. (Hemel Hempstead, UK), Invitrogen Ltd. (Paisley, UK), Promega UK Ltd. (Southampton, UK), Amersham Biosciences-Division of GE Healthcare (Buckinghamshire, UK), Bioline Ltd. (London, UK), Millipore (Carrigtwohill, IRE), Fermentas (Maryland, US), Cargille Laboratories (New Jersey, US) and Pierce – Division of Thermo Fisher Scientific Ltd. (Cramlington, UK). Enzymes and enzyme inhibitors were purchased from New England Biolabs (NEB) (Beverly, MA, US) or Roche (Clarecastle, IRE), Promega or Novagen. Oligonucleotide primers were purchased from Eurofins MWG Operon (Edersberg, DE). Cell culture plates were purchased from Greiner Bio-One GmbH (Frickenhausen, DE). Sterile plasticware was purchased from Sartorius AG (Goettingen, DE). Molecular biology kits were purchased from Qiagen (Cologne, Germany) or Novagen.

2.1.2 Bacterial strains and plasmids

Escherichia coli TOP10 chemically competent cells were purchased from Invitrogen, *E. coli* OP50 were obtained from the *Caenorhabditis elegans* Genetics Centre (Minnesota, US). Plasmid pCR2.1 TOPO was from Invitrogen and pJET 1.2/blunt was from Fermentas.

2.1.3 Source of *Panagrolaimus* species and strains

The geographical location and source of the *Panagrolaimus* sp. strains used in this study are listed in Table 2.1.

Table 2.1. The source and geographical location of species and stains used in this study.

Species and Strain	Geographical Location	Obtained from
Panagrolaimus davidi	McMurdo Sound, Antarctica	Prof. Dee Denver
Panagrolaimus hygrophilus (PS1732)	Iceberg Lake, Mt. Whitney California, USA	John DeModena
Panagrolaimus paetzoldi	Paulina saltmarsh, Westerscheldt Estuary, Netherlands	Dr. Ilse De Mesel
Panagrolaimus rigidus (AF36)	Fayette County Pennsylvania, USA	CGC
Panagrolaimus superbus (DF5050)	Surtsey Island, Iceland	Prof. Bjorn Sohlenius
Panagrolaimus sp. AS01	Leixlip, Co. Kildare, Ireland	Dr. Adam Shannon
Panagrolaimus sp. AS03	Maynooth, Co. Kildare, Ireland	Dr. Adam Shannon
Panagrolaimus sp. PS443	Byurakan, Armenia	CGC
Panagrolaimus sp. PS1159	North Carolina, USA	CGC
Panagrolaimus sp PS1579	California, USA	CGC
Panagrolaimus sp. JB115	Forest Falls, San Bernadino County, California, USA	Prof. Dee Denver
Panagrolaimus sp. JB051	Diourbel, Senegal	Prof. Steven Nadler
Panagrolaimus sp. SN103	Eagle Lake Field Station, Lassen County, California	Prof. Steven Nadler
Panagrolaimus sp. PS5056	Cambodia	Dr. John DeModena
Panagrolaimus sp. Galileo desert sp.	California, USA	Dr. John DeModena
Panagrolaimus sp. JU765	Yangshuo, Guangxi, China	Dr. Marie-Anne Félix
Panagrolaimus sp. JU1361	Periyar Natural Preserve, Kerala, India	Dr. Marie-Anne Félix
Panagrolaimus sp. JU1365	Between Periyar and Madurai, Tamil Nadu, India	Dr. Marie-Anne Félix
Panagrolaimus sp. JU1366	Between Madurai and Tanjore, Tamil Nadu, India	Dr. Marie-Anne Félix
Panagrolaimus sp. JU1367	Between Madurai and Tanjore, Tamil Nadu, India	Dr. Marie-Anne Félix
Panagrolaimus sp. JU1369	Cuddatore, Tamil Nadu, India	Dr. Marie-Anne Félix
Panagrolaimus sp. JU1371	Government Place, Pondicherry, India	Dr. Marie-Anne Félix
Panagrolaimus sp. JU1387	La Réunion, Indian Ocean	Dr. Marie-Anne Félix
Panagrolaimus sp. JU1645	Between Matinho and Caibros on Santo Antao Island, Cape Verde	Dr. Marie-Anne Félix
Panagrolaimus sp. JU1646	Ribeira (Valley) Cumba below Renque de Purga on Santiago Island, Cape Verde	Dr. Marie-Anne Félix

2.2 Methods

2.2.1 Nematode culturing methods

2.2.1.1 Culturing of *Panagrolaimus* species and strains

Nematode Growth medium (NGM) plates were prepared by dissolving 3 g NaCl, 2.5 g peptone, 17 g agar in 970 ml of H_2O. The solution was autoclaved and once cooled 1 ml cholesterol in ethanol (5 mg/ml), 25 ml 1M potassium phosphate buffer pH6, 1ml 1M $CaCl_2$, 1 ml 1M $MgSO_4$ and Streptomycin to a final concentration of 10 μg/ml was added. The media was poured into 9 cm plates and left to set. Upon setting 40 μl of *E. coli* HB101 culture was spread on the plates and left to grow overnight at 37 ^oC. A 1 x 1 cm chunk from an established NGM nematode culture was transferred using a sterile scalpel onto the plates. The plates were cultured in the dark at 20 ^oC until the nematodes reached a large mixed population (7-14 days depending on strain). The nematodes were regularly subcultured (every 7- 14 days) using 1 x 1 cm agar squares. To culture larger quantities of nematodes 15 cm NGM plates were used.

2.2.1.2 Mass culturing nematodes

Liquid cultures were prepared by inoculating 250 ml of S Medium (1 L S Basal, 10 ml 1M Potassium citrate pH 6.0, 10 ml Trace metal solution, 3 ml 1M $CaCl_2$ and 3 ml $MgS0_4$) with a concentrated *E. coli* HB101 pellet made from 2-3 litres of an overnight culture. Four large plates of nematodes were washed with 5 ml of S Medium and added to the S Medium/*E. coli* mixture. The flask was placed at 20 ^oC and vigorously shaken so that the culture was well oxygenated. A drop of the culture was checked under the microscope to monitor the nematode growth. When the food supply became depleted (the solution is no longer visibly cloudy) more concentrated *E. coli* HB101 was resuspended in S Medium and added to the nematode culture. The nematodes were harvested when large numbers were seen in each drop (approximately 10 days).

2.2.1.3 Harvesting nematodes

Nematodes were rinsed off the NGM plates using sterile H_20. The nematodes were left to gently agitate on a shaker for 20 min in order to digest any bacteria present in the nematode gut. The liquid was poured off the plates into sterile 50 ml Falcon tubes

and left to settle at 4 ^{o}C for 20 min. The supernatant was removed and fresh sterile H_20 was added. This process was repeated a total of three times in order to obtain bacteria free nematodes. When harvesting liquid cultures the nematodes were also poured into 50 ml Falcon tubes and cleaned using the same method. When the nematodes were used for protein extraction 100mM Ammonium bicarbonate was used to rinse and wash the cultures. The supernatant was removed and the pellet was snap frozen in liquid nitrogen prior to storage at -80 ^{o}C. If the nematodes were required for gDNA/RNA extraction, the pellet was resuspended in either Nematode Lysis buffer (20 mM Tris pH 7.5, 50 mM EDTA, 200 mM NaCl, 0.5% SDS (w/v)) or Trizol Reagent® (Invitrogen) before freezing. For stress experiments the nematodes were resuspended in a suitable volume of H_20. The number of nematodes per ml was estimated by counting the number of nematodes in a 10 μl aliquot (10 replicates). The volume of the nematodes suspended in H_20 was adjusted to give the required nematode density for each experiment (approximately 2,000 nematodes/ml). Mixed age populations of nematodes were used for all experiments.

2.2.1.4 Cleaning nematode stocks

Harvested nematodes were brought to a final volume of 30 ml with S Basal buffer. The nematodes were treated with a freshly prepared solution comprising 8 ml 1:3 NaOCl solution and 2 ml 1M NaOH. This solution was shaken vigorously for 4 min to disintegrate the adult nematodes and to allow the release of their eggs. Released eggs were pelleted by centrifugation (2,000 x *g* for 2 min). The supernatant was removed and fresh sterile S Basal was added. This process was repeated a total of three times to obtain sterile NaOCl/NaOH/bacteria free eggs. An aliquot of the eggs were examined under the microscope and resuspended in S Basal to a final volume of approximately 4,000 eggs per ml. Eggs were either added directly to a fresh liquid culture or 50 μl was added onto a fresh NGM plate to start a new sterile culture.

2.2.1.5 Freezing nematode stocks

Plates of freshly starved nematodes (NGM plate where the bacteria is cleared) were rinsed with 0.6 ml S Buffer (129 ml 0.05 M K_2HPO_4, 871 ml 0.05 M KH_2PO_4, 5.85 g NaCl). The liquid was poured off into a sterile 50 ml Falcon tube. An equal volume of sterile S Buffer containing 30% Glycerine (v/v) was added and mixed well. A 1 ml

aliquot of the mixture was added to a 1.8 ml cryovial and labelled. The cryovial was placed in a styrofoam box left at -80 ^{o}C overnight. The next day the cryovial was transferred to a liquid nitrogen tank for long-term storage.

2.2.1.6 Recovery of frozen nematode stocks

A cryovial of frozen nematodes was removed from the liquid nitrogen and allowed to thaw at room temperature. The contents were poured onto a NGM plate with *E. coli* HB101 lawn. The plates were observed under a microscope to check for recovery (nematodes start wriggling within a few minutes). After 2-3 days the nematodes were transferred onto new plates, then subcultured as required (every 7-14 days).

2.2.1.7 Desiccation stress experiments

Nematodes were harvested (Section 2.2.1.3) and adjusted to a concentration of approximately 2,000 nematodes per ml. One ml of the worm suspension was vacuum filtered onto a Supor®-450 filter membrane disc (Pall), transferred to 3 cm Petri dishes without lids and placed in a desiccation chamber set at 98% RH (using a saturated solution of potassium dichromate) for either 0, 24, 48, 72 and 96 h. After this preconditioning period the nematodes were transferred to a desiccation chamber containing freshly activated silica gel for 48 h. Following this drying period the nematodes were rehydrated in distilled water and allowed to recover for 24 h at 20 ^{o}C with shaking at 50 RPM. Percentage survival was assessed by microscopic observation of movement. Three biological replicates were prepared for each preconditioning time point. Statistical significant differences was determined using ANOVA testing with the Graphpad (PRISM) statistical software.

2.2.1.8 Freezing stress experiments

To determine the optimal acclimation regime nematodes were acclimated at varying temperatures for different lengths time before transfer to -80 ^{o}C for 24 h. NGM plates containing mixed populations of nematodes were acclimated using the following regimes; plates at 4 ^{o}C for 3 days, 4 ^{o}C for 10 days, 10 ^{o}C for 3 days, and 10 ^{o}C for 10 days. The nematodes were harvested (Section 2.2.1.3) and adjusted to a concentration of approximately 2,000 nematodes per ml. The nematodes were aliquotted into a 1.5 ml tube, transferred to a polystyrene box and cooled at approximately 0.05 ^{o}C per

second in a -80 °C freezer. After 24 h the nematodes were thawed rapidly at 30 °C, placed in a 3 ml Petri dish and allowed to recover for 24 h at 20 °C with shaking at 50 RPM. Percentage survival was assessed by microscopic observation of movement. Four biological replicates were prepared for each treatment and for the control worms (worms growing at 20 °C prior to transfer to -80 °C). Statistically significant differences was determined using ANOVA testing with the Graphpad (PRISM) statistical software.

2.2.2 Molecular methods

2.2.2.1 Genomic DNA extraction

Nematodes were cultured and harvested as described in Section 2.2.1.1 and 2.2.1.3 respectively. Pellets were resuspended in 500 µl nematode lysis buffer and snap frozen in liquid nitrogen. Pellets were stored at –80 °C until required. The tubes were defrosted at room temperature and ground into a fine powder (≥3 times) using an autoclaved pestle and mortar and liquid nitrogen. The supernatant was then pippetted into 1.5 ml Eppendorf tubes and proteinase K (10 mg/ml in H_2O) was added to a final concentration of 2 mg/ml. The mixture was then incubated at 56 °C for 1 h (with regular inversion of the tubes) or until no nematode carcasses could be identified on examination under a microscope. The solution was cooled to room temperature, RNase A was added to a final concentration of 1.2 mg/ml and the tube was incubated for 15 min. The solution was then extracted with 2.5 vol of phenol:chloroform:isoamyl alcohol (24:24:1) at room temperature for 10 min. The tubes were spun in a cooled microcentrifuge (Eppendorf) at 4 °C at 12,000 x *g* for 10 min. The aqueous layer was transferred to a new tube and a second phenol:chloroform:isoamyl alcohol extraction was performed. The solution was spun as previously and the supernatant transferred to a new tube. A final 10 min chloroform:isoamyl alcohol (24:1) extraction was carried out to remove any residual phenol and the sample was centrifuged as previously. The aqueous layer was then transferred to a new tube and 1/30th vol of 3 M sodium acetate and 2.5 vol of ice-cold 100% ethanol were added. The solution was stored at -20 °C for 30 min. The DNA was pelleted by centrifugation at 12000 x *g* for 20 min at 4 °C. The supernatant was removed and the pellet washed with 1 ml of 75% EtOH and spun for 5 min at 7,500 x *g*. The supernatant was removed and the pellet air-dried. The pellet was then

resuspended in 25 µl of sterile H_20 and the DNA concentration and integrity were determined as described in Section 2.2.2.4 and 2.2.2.5.

2.2.2.2 RNA extraction

Nematodes were cultured and harvested as described in Section 2.2.1.1 and 2.2.1.3. Pelleted nematodes (~100 µl) were resuspended in 500 µl of Trizol® reagent (Invitrogen) and snap frozen in liquid nitrogen and were stored at –80 ^{o}C until required. All areas used for RNA preparation, as well as the pipettes, were thoroughly cleaned with EtOH and RNase Zap solution (Invitrogen) before use. Sterile filtered tips (Molecular Bioproducts) and certified DNase and RNase free Eppendorf tubes were used for all RNA work.

Tubes containing frozen nematodes were defrosted, as required, at room temperature and the nematodes ground into a fine powder (≥3 times) using a pestle and mortar (baked O/N at 200 ^{o}C) and liquid nitrogen. The homogenised extract was incubated at room temperature (RT) for 10 min. Two hundred microlitres of chloroform was then added, shaken vigorously for 10 sec and incubated at RT for 15 min. The tubes were then spun at 12,000 x *g* for 15 min at 4 ^{o}C in a microcentrifuge (Eppendorf). The aqueous phase was removed to a new tube and the RNA precipitated with 500 µl of isopropyl alcohol. The samples were incubated at RT for 15 min and centrifuged for 10min as previously described. The supernatant was removed and the pellet resuspended in 1 ml 75% EtOH and centrifuged at 7,500 x *g* for 5 min. The supernatant was then removed and the pellet air-dried. The pellet was resuspended in DEPC treated H_2O. All RNA was treated with RQI DNase (Promega) according to the manufacturer's protocol. The sample was phenol:chloroform extracted and ethanol precipitated as previously described to remove the salts left by RQI DNase reaction buffer. The RNA concentration and integrity was determined by the Qubit® 2.0 Fluorometer and the Agilent Bioanalyzer as in Section 2.2.2.4.

2.2.2.3 cDNA synthesis

cDNA for qPCR experiments was synthesized using a Transcriptor First Strand cDNA synthase kit (Roche) according to the manufacturer's instructions. cDNA was prepared using 1 µg of RNA and Oligo dT primers.

2.2.2.4 Determination of RNA/DNA concentration and quality

DNA/RNA concentration and purity were determined by using a Qubit® RNA/DNA assay kit with the Qubit® 2.0 Fluorometer. RNA/DNA was also visually compared to standards of known concentration on ethidium bromide stained gels. RNA quality was determined based on the RNA integrity number algorithm calculated using the Agilent Bioanalyzer. RNA was prepared for the Bioanaylzer with the RNA 6000 Nano kit according to the manufacturer's instructions.

2.2.2.5 Agrose gel electrophoresis

Electrophoresis was carried out using 0.7-1% agarose gels (depending on the degree of separation required) in 1X Tris Acetate EDTA buffer (TAE): 40 mM Tris, 20 mM acetic acid and 2 mM EDTA, pH 8.1. Agarose powder was dissolved by heating in 1X TAE buffer. When RNA was to be electrophoresed all buffers were made using DEPC treated H_2O. Upon cooling to 60 ^oC ethidium bromide was added (10 mg/ml). The solution was then poured into a casting tray and allowed to solidify. Samples were mixed with loading buffer (5mg/ml bromophenol blue, 5 mg/ml xylene cyanol, 50% glycerol) at a ratio of 5:1. Six µl of 1 kb or 100 bp bench top DNA ladders (Promega) and HyperLadder I (Bioline) was loaded on each gel. Gels were typically run at 100V using BioRad electrophoresis equipment.

2.2.2.6 DNA/RNA visualisation

The ethidium bromide stained DNA or RNA gels were visualized under UV light using a UV transilluminator at 365 nm. Gels were photographed using an Eagle-Eye® gel documentation system (Stratagene), or an AlphaDigiDoc® gel documentation system (Alpha Innotech).

2.2.2.7 Polymerase Chain Reaction (PCR) DNA amplification

All PCR reactions were carried out in 25 µl volumes. General PCR reactions were carried out using Promega's GoTaq, however Platinum® Taq from Invitrogen was utilised for reactions requiring high fidelity. The primers and their sequences used in this study are shown in Table 2.2. The PCR reaction conditions used can be seen in Table 2.3. The different cycling parameters required are shown in Table 2.4.

Table 2.2. PCR primer sequences used in this study.

Amplification Experiment	Primer name	Sequence 5' to 3'
Cloning plasmid inserts	M13F	GTAAAACGACGGCCAC
	M13R	CAGGAAACAGCTATGAC
	T7F	TAATACGACTCACTATAGGG
	T7R	GCTAGTTATTGCTCAGCGG
	pDNRlib_F	GCATAACTTCGTATAGCATAC
	pDNRlib_R	AAACAGCTATGACCATGTTC
	pJET1.2F	CGACTCACTATAGGGAGAGCGGC
	pJET1.2R	AAGAACATCGATTTTCCATGGCAG
18S rDNA amplification	18S_StartF	TAAACACGAAACCGCGTA
	18S_InternalR	ATCTGATCGCCTTCGATCCT
	18S_InternalF	GTGAAATTCGTGGACCCTTG
	18S_EndR	TACGGCCACCTTGTTACGAC
	2F(18s)	GAAACCGCGTATGGCTCATTA
	3R(18s)	CAGATCCAACTACGAGCT
D3 rDNA amplification	D3A	GACCCGTCTTGAAACACGGA
	D3B	TCGGAAGGAACCAGCTACTA
ITS rDNA amplification	rDNA1	TTGATTACGTCCCTGCCCTTT
	rDNA2	TTTCACTCGCCGTTACTAAGG
	Pan5.8F	GCTTGCTGCGTTACTTACC
	Pan5.8R	GGTAAGTAACGCAGCAAGC
	SSU18P	TGATCCWKCYGCAGGTTCAC
	ITS2F	GATGAAGAACGCAAAATGCGTTA
	TW81	GTTTCCGTAGGTGAACCT
	AB28	ATATGCTTAAGTTCAGCG
	ITS1F	GGCCGGATCATTACAAGAAA
	ITS2R	TCAGCGGGTAATCACACTTG
	5.8SF	CCATTGGCACATTTGTTTGA
	5.8SR	AAAAACCATCGGCGCAAT
Quantitative PCR	T6PS_F	GGTTGCAAGACAAGATGCAA
	T6PS_R	GTTTGTCCGGTTGGATCAAT
	LEA5_F	GGAGCTGCAAAGGTTAAAGC
	LEA5_R	ATGGCATCTTGTTGTTCACG
	HSP90_F	GGGAGGATCATCTTGCTGTT
	HSP90_R	GCTTGATGGAATTCTTGGTC
	Pyruvkin_F	GGGAGATCTTGGCATTGAAA
	Pyruvkin_R	TCGAGTTGGACGAGGCTTAT
	ThaumatinT6F	TGTGGAACTGGATTCTGTGG
	ThaumatinT6R	GATTTCCTGCTCCATCAAGG
	LipaseT6F	GCAGCTGCCGCATATAGTAA
	LipaseT6R	AACGGCAGTAAAACCAGAGC
	LectinT6F	AAGCGGAGGTATCTGGATTG
	LectinT6R	TTCTCCACCAGGGTTATTGG
	GSTT6F	GCTTTGGCAGATTTTCTTCG
	GSTT6R	GTTTTACGGCTGGAAGGAAG
	CatalaseT6F	TTTGGACCTCGTGGTACTCC
	CatalaseT6R	TCCTTGACGGGTCTTCAGAT
	SAMT5F	ACAATGTCAATGCCACAAGC
	SAMT5R	GTGCATTCTCTGCCATCAAC
	AnnexinT5F	TGCATGGAAGCTTAATCGTG
	AnnexinT5R	AAAAACGAGGCGAAGTTGAG
	RSK5F	AAAAAGATACCGCGCATACC
	RSK5R	AGAGCTCACCACCGGATAAA

Table 2.3. Components required for PCR amplifications using GoTaq (Promega) and Platinum Taq (Invitrogen).

Reagent	Volume per reaction
GoTaq®	
5X Green or Colourless reaction buffer	5 µl
PCR Nucleotide Mix (10 mM each)	1 µl
$MgCl_2$	2.5 µl
Primer forward (10 pmol)	1 µl
Primer reverse (10 pmol)	1 µl
GoTaq Polymerase	0.2 µl
Template DNA	100 ng-0.5 µg/50 µl
PCR grade H_20	To total of 25 µl
Platinum® Taq	
10X High Fidelity buffer	2.5 µl
PCR Nucleotide Mix (10 mM each)	.5 µl
$MgCl_2$ (50 mM)	1 µl
Primer forward (10 pmol)	1 µl
Primer reverse (10 pmol)	1 µl
Platinum Taq	0.25 µl
Template DNA	100 ng-0.5 µg/50 µl
PCR grade H_20	To total of 25 µl

Table 2.4. PCR cycling conditions required for the GoTaq and Platinum Taq polymerases.

PCR Step	GoTaq Temperature/Time	Platinum Taq Temperature/Time
Denaturation/Activation	95°C/2 min	94°C/30 sec
Denaturation	95°C/0.5-1 min	94°C/30 sec
Annealing	-5°C below T_m/0. 5-1 min	-5°C below T_m/30 sec
Extension	72°C/1 min	68°C/1 min/kb
Final Extension	72°C/10 min	68°C/10 min

Denaturing, annealing and extension steps were repeated for 25-40 cycles. Annealing temperatures were estimated as ~5 ^{o}C below the melting temperature (T_m) of the primers used. The extension times used were 1 min/kb of DNA to be synthesised. Reactions were carried out using an Eppendorf PCR Thermal Cycler, G-Storm GS1 Thermal Cycler, or Piko® Thermal Cycler.

2.2.2.8 Purification of DNA from agrose gels

The band of interest was cut from the ethidium bromide stained agarose gel and purified using a QIAquick Gel Extraction Kit (Qiagen) according to manufacturer's instructions.

2.2.2.9 Purification of PCR amplified DNA

PCR amplicons were purified using the QIAquick PCR purification kit from Qiagen according to the manufacturer's instructions.

2.2.2.10 Cloning methods

PCR products were cloned into the pCR2.1® TOPO vector (Invitrogen) or the pJET1.2/blunt vector (Fermentas) and transformed into *E. coli* TOP10® cells according to manufacturer's instructions.

2.2.2.11 *E. coli* growth conditions

E. coli was grown on Luria-Bertani (LB) agar plates and LB broth at 37 ^{o}C with shaking at 180 RPM. A concentration of 50 µg/ml ampicillin was used to select for the pCR2.1 TOPO and pJET1.2/blunt vectors.

2.2.2.12 Storage of *E. coli* strains

Recombinant *E. coli* strains containing the various plasmids were stored at -80 ^{o}C. One ml of *E. coli* culture was mixed with 1 ml sterile glycerol:LB broth (1:1) and frozen in 1 ml aliquots in cryovials.

2.2.2.13 Plasmid purification

Plasmids were purified from overnight 5 ml cultures of *E. coli* inoculated with a single transformed colony using a Qiagen Miniprep kit (Qiagen) according to the manufacturer's instructions.

2.2.2.14 DNA sequencing

The PCR products and purified plasmids were sent for DNA sequencing to Agowa (www.agowa.de). The DNA sequences obtained were vector screened using the VecScreen program at NCBI (http://www.ncbi.nlm.nih.gov/Vecscreen). The sequences were then analysed on the NCBI website using the BLAST program (www.ncbi.nlm.nih.gov/BLAST) to search the GenBank database.

2.2.2.15 Quantitative PCR

RNA was isolated as described in Section 2.2.2.2 and cDNA synthesis was performed using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche) as described in Section 2.2.2.3. Real time qPCR reactions were performed using a Roche Lightcycler 480 thermocycler and SYBR green master mix (Roche). Each reaction had a total volume of 10 µl and was comprised of: 5 µl SYBR green master mix, 1.875 µl PCR grade H_2O, 0.5 µl forward primer (5 pmol), 0.5 µl reverse primer (5 pmol), 0.125 µl Lightcycler Uracil-DNA Glycosylase (Roche) and 2 µl cDNA. The PCR program used consisted of 40 °C for 10 min, 95 °C 10 min followed by 45 cycles of 95 °C for 10 sec, 58 °C for 20 sec, 72 °C for 10 sec. Melting curve analysis was carried out for each experiment. The *P. superbus* RNA polymerase II and 60S ribosomal subunit genes were used as reference genes. Negative controls containing no cDNA were also performed. Each reaction was replicated a minimum of 4 times on each plate. By comparison of the cycle threshold values of the control (housekeeping genes) to the sample, the relative concentration of the template was determined using the second derivative maximum method for relative quantification as described in the Roche Lightcycler manual. Statistically significant differences between the control and treatments were confirmed by comparing the normalised mean crossing points using ANOVA (P<0.001) using Graphpad (PRISM) statistical software.

2.2.3 Phylogenetic methods

2.2.3.1 Database searching

DNA sequences used for phylogenetic analysis but not sequenced in this project were obtained from GenBank on the National Centre for Biotechnology Information (NCBI) website located at www.ncbi.nlm.nih.gov. The species origin of nematode sequences was identified by searching against the GenBank database using BLASTN. Peptide fragments from LC-MS/MS analysis were searched against GenBank using the MASCOT online search software available at www.matrixscience.com. Peptides from *P. superbus* were searched against the *P. superbus* transcriptome using the Agilent Spectrum Mill software.

2.2.3.2 Gene sequence assembly

DNA regions from a particular gene that was sequenced in parts, sequenced in the forward and reverse direction or internally sequenced were assembled using the Cap3 online assembler (http://pbil.univ-lyon1.fr/cap3.php). The alignment of the assembly was manually checked for errors that may have been in the consensus sequence.

2.2.3.3 Multiple sequence alignment

Multiple sequence alignments were created using the MUSCLE multiple sequence alignment tool within the MEGA 5 software suite (www.megasoftware.net). Different alignment file formats were required for each of the phylogenetic analysis programs, FASTA for maximum likelihood and NEXUS format for the Bayesian analysis. MEGA 5 output in FASTA format was converted to NEXUS format using the online Readseq conversion tool (http://www.ebi.ac.uk/cgi-bin/readseq.cgi). All alignments were performed using default gap penalties and checked by eye for obvious errors.

2.2.3.4 Structural multiple sequence alignment

Ribosomal RNAs are highly structured, with large parts of the molecules exhibiting very strong conservation of their base pairing patterns. Therefore, alignment accuracy can be improved by incorporating information about rRNA secondary structure conservation. Structural multiple sequence alignment was performed using the program RNAsalsa (Stocsits et al., 2009b). The source code for this program was

downloaded from www.bioinf.uni-leipzig.de/Software/RNAsalsa/. The method implemented by RNAsalsa takes into account a known accurate RNA secondary structure and primary sequence information and uses well-established folding algorithms to generate a consensus structural alignment. This is a command line program. The inputs for this program are a multiple sequence alignment and a constraint file. In this study, *S. cerevisiae* was used as a constraint for 18S alignments, these constraint and sequence files are provided with program. The multiple sequence alignment was performed using MUSCLE. This alignment also included *S. cerevisiae*. The alignment and constraint file were copied into the RNAsalsa folder. Within the terminal console, the directory was changed to the RNAsalsa directory and the following command was used to execute the program:

```
RNAsalsa -i <input_alignment> -c <constraint_file>
```

The –i flag tells RNAsalsa the name of the multiple sequence alignment to be structurally aligned and –c flag gives the name of the constraint file. There are several outputs including individual structural predications for all the sequences. The output file SALSA_structaln.aln contains the structural alignment, which was used for subsequent phylogenetic analysis.

2.2.3.5 Phylogenetic reconstruction methods

2.2.3.5.1 Maximum likelihood

The ML method for creating phylogenetic trees was performed using a FASTA nucleotide multiple sequence alignment file format. A test to find the best evolutionary model for the DNA multiple alignment dataset was performed using MEGA 5. The phylogenetic analysis was then performed on this alignment with the calculated best model of nucleotide substitution and with 1,000 bootstrap replicates. All other values were left as default. The resulting tree was exported in Newick file format and viewed using the FigTree phylogenetic tree viewer, freely available at http://tree.bio.ed.ac.uk/software/Figuretree/. The tree was rooted to the most basal nematode in the tree, i.e. *Panagrellus redivivus*.

2.2.3.5.2 Bayesian

The Bayesian method for creating phylogenetic trees was performed using Mr Bayes (http://mrbayes.sourceforge.net/index.php) with nucleotide multiple sequence alignment created using MUSCLE within MEGA 5. In order to implement this Bayesian method the FASTA formatted alignment converted to NEXUS (.nxs) format using the online Readseq tool. This .nxs file was then opened using a text editor and checked for the correct file format (as in the following example):

```
#NEXUS
BEGIN DATA;
dimensions ntax=11 nchar=908;
format datatype=DNA interleave=yes gap= -;

matrix
Sequence 1 CACGAAACCGCGTATGGCTCATTATAACAGCTAAAATTTACTTGATTTTG
Sequence 2 CACGAAACCGCGTATGGCTCATTATAACAGCTATAATTTACTTGATTTTG
Sequence 3 CACGAAACCGCGTATGGCTCATTATAACAGCTAAAATTTACTTGATATTG
;
end;
```

This .nxs file was saved and the following commands were entered line by line into the Mr. Bayes program:

```
execute <filename.nxs>
lset nst=6
lset rates=gamma
mcmcp samplefreq=100
mcmc ngen=<number of generations>
sump burnin=<X>
sumt burnin=<X>
```

Briefly the NEXUS alignment file is executed, any output files will have the same name with the relevant extension. The lset nst=6 sets the model to the general time reversible model (GTR) of nucleotide evolution (Waddell and Steel, 1997); lset rates=gamma sets the model to the general gamma model of variation; mcmcp samplefreq=100 tells the program to save the current tree to file every 100 generations; mcmc ngen sets the number of generations. Generally, enough

generations were applied to ensure that the trees have converged and give a split frequency of less than 0.01.

The burnin value is the most important value within the analysis as it instructs the program to discard all the trees that have not converged. The value number selected for the burnin was 25% of the samples. For example, with 100,000 generations and a sample frequency of 100 the burnin value used was 250.

The resulting consensus tree was then outputted to the specified filename with a .con extension. This file was viewed using the FigTree phylogenetic tree viewer. The tree was rooted to the most basal nematode in the tree i.e. *Panagrellus redivivus*.

2.2.4 Molecular clock methods

2.2.4.1 Molecular clock analysis

The divergence time between different strains and species of nematodes was estimated using the program Phylobayes (Lartillot et al., 2009). This analysis was performed on the high performance computing facility at NUI Maynooth using an SGI Altix 8200 ICE cluster (Sioc) with 43 compute nodes. The following script was written by Dr. Davide Pisani, NUI Maynooth to run the Phylobayes molecular clock analysis:

```
#!/bin/bash
#PBS -N <filename>
#PBS -l ncpus=1
#PBS -l Qlist=Short
#PBS -l walltime=<time>
#PBS -r n
#PBS -j oe
#PBS -m bea
#PBS -M <email address >
cd $PBS_O_WORKDIR
source /etc/profile.d/modules.sh
module load gsl-1.12
module load phylobayes
```

```
pb -d <alignmentfile> -T <treefile> -r <outgroup file> -rp <root age
and standard deviation> -bd -sb -cal <calibration file> -cir(or –
ugam) -gtr -cat -f <output filename>
```

The initial part of the script (first 13 lines) is required to be included in a script where a job is submitted to Sioc. It specifies the queue that the job is to be submitted to (long or short) and the amount of time required after which the job will be terminated. In this analysis the jobs were submitted to the short queue for 23.30 h (max possible is 24 h). The module Phylobayes is also loaded.

At the start of the script Phylobayes is called (pb). The program is given multiple sequence alignment (-d) (NEXUS or a generalised PHYLIP format), a tree file (-T)(NEWICK format), an outgroup file (-r) (text file), the specified mean root age followed by the standard deviation (-rp) (text file) and calibration file (text file) containing the species names and dates use to calibrate the tree (-cal). The bounds of the analysis (-bd) may be set to be either soft or hard. In this analysis soft bounds were used (-sb), this has a default of 0.05 (5%) but this may be adjusted. The model for the rate of evolution can also be adjusted. In this analysis both the CIR (-cir) model and the UGAMMA (-ugam) were used. Dirichlet process combined with a GTR matrix of exchange rates was used in all analyses (-gtr -cat). The name for the output files is specified by the –f prior. The script was saved as a shell script file (.sh) and uploaded to Sioc. The necessary tree, alignment, calibration and outgroups files were also uploaded to the same folder as the script on Sioc. The script was sent to the queue using the command:

```
qsub <filename.sh>
```

2.2.4.2 Molecular clock post-analysis

This script described in Section 2.2.4.1 ran one “chain” for several thousand cycles generating a tree for each. When the run was complete the Phylobayes implemented post-analysis program Readdiv was executed. This calculates the posterior mean chronogram and allows for the unconvereged trees to be removed (similar to burnin in Mr. Bayes). Initially the number of trees generated was calculated using the unix command:

```
grep –c ";" *treelist
```

Approximately one third of the trees were discarded using the Readdiv program:

```
readdiv -x <number of trees to remove> 1 <name of file>
```

The posterior mean chronogram was then outputted to same file name with a .con extension. This file was viewed using the FigTree phylogenetic tree viewer. The tree was rooted to the outgroups. The dates at each node were outputted to a .dates file. Several chains were ran for each individual analysis and the average dates for each node calculated. Significant differences between analyses with different parameters were estimated using the Graphpad (PRISM) statistical software.

2.2.5 Protein methods

2.2.5.1 Small scale protein extraction from nematodes

Nematodes were cultured and harvested as described in Section 2.2.1.1 and 2.2.1.3. Approximately 6-10 plates of a healthy mixed stage culture provided sufficient nematodes numbers for a small-scale protein extraction (e.g. for use in ice-shaping activity assay, Section 2.2.5.8). A 500 µl pellet of compact worms was collected and frozen at – 80 ^{o}C. The nematode pellets were defrosted at room temperature and ground into a fine powder (≥3 times) using an autoclaved pestle and mortar and liquid nitrogen. The supernatant was pipetted into a 1.5 ml tube and an equal volume of 100 mM ammonium bicarbonate was added. Protease inhibitor cocktail tablets (Roche Complete Mini) were added at a concentration of 1 tablet per 10 ml of extraction solution. The extraction solution was mixed with 0.25 g of glass bead (<106 microns, Sigma). The samples were homogenised in beadbeater (Eppendorf) 3 times for 30 sec at high speed with cooling on ice for 1 min between each round of beating. The samples were then centrifuged at 8,000 RPM for 5 min at 4 ^{o}C and the supernatant placed in a 50 ml tube. The protein concentration was determined using the BCA protein assay kit (Section 2.2.5.2).

2.2.5.2 Large scale protein extraction from nematodes

Nematodes were cultured and harvested as described in Section 2.2.1.2 and 2.2.1.3. When approximately 40 ml of compact nematodes were collected (up to 3 months of

constant nematode growth) they were defrosted at room temperature and ground into a fine powder (≥3 times) using a large autoclaved pestle and mortar and liquid nitrogen. The supernatant was pipetted into a 100 ml sterile beaker and an equal volume of 100 mM ammonium bicarbonate was added. Protease inhibitor cocktail tablets (Roche Complete Mini) were added at a concentration of 1 tablet per 10 ml of extraction solution. The extraction solution was transferred into 1.5 ml tubes and 0.25 g of glass bead were added (<106 microns, Sigma). The samples were homogenised in beadbeater (Eppendorf) 3 times for 30 seconds at high speed with cooling on ice for 1 min between each round of beating. The samples were then centrifuged at 8,000 RPM for 5 min at 4 ^{o}C and the supernatant placed in a sterile 100 ml beaker. The protein concentration was determined using the BCA protein assay kit (Section 2.2.5.2) and the extract diluted to 1.5 mg/ml. Extract was aliquoted in 50 ml tubes and placed at 4 ^{o}C until use.

2.2.5.3 Determination of protein concentration

Concentrations of mixed protein extracts were determined using the BCA (bicinchoninic acid) protein assay kit (Novagen) according to the manufacturer's instructions. Dilutions of BSA from 0.1–1.5 mg/ml were used to produce a standard curve for each assay.

2.2.5.4 SDS-PAGE

Glycine-SDS-PAGE (Laemmili, 1970) and Tricine-SDS-PAGE (Schagger, 1987) were carried out in a Biorad Mini protean 3 gel tank according to the manufacturer's instructions. A 12% gel with a 5% stacking (see Table 2.5) was used unless otherwise stated. Protein samples were diluted 1:4 in 4X SDS sample buffer (Novagen) and heated to 95 ^{o}C for 10 min and cooled on ice for 5 min prior to loading. Approximately 40 μg of crude protein extract was loaded on each gel where possible using a Hamilton syringe. The markers used were BioRad *Precision Plus All Blue* or *Kaleidoscope Standards* and NEB *Broad Range Marker*.

2.2.5.5 Protein staining methods

2.2.5.5.1 Coomassie brilliant blue protein staining

Gels were stained in 10% (v/v) acetic acid, 45% (v/v) methanol, and 0.2% (w/v) Coomassie brilliant blue for 1 h followed by destaining in 10% (v/v) acetic acid and 30% (v/v) methanol.

Table 2.5. SDS-PAGE gel components.

	Stacking gel	Separating gel
Tris-glycine SDS-PAGE		
30% (w/v) Acrylamide/bis (ml)	1.3	11.3
1.5M TrisHCl pH 8.8 (ml)	-	7.0
0.5M TrisHCl pH 6.8 (ml)	2.5	-
Deionised water (ml)	6.1	9.3
10% (w/v) SDS (µl)	100	280
10% (w/v) APS (µl)	200	100
TEMED (µl)	20	23
Tris-tricine SDS-PAGE		
30% (w/v) Acrylamide/bis (ml)	0.8	5.0
3M TrisHCl, pH8.5	1.55	5.0
50% (w/v) Glycerol(ml)	-	3.0
Deionised water (ml)	3.9	2.0
30% (w/v) APS (µl)	14	14
TEMED (µl)	14	14

2.2.5.5.2 Colloidal coomassie protein staining

Gels were incubated in 50% (v/v) ethanol, 3% (v/v) phosphoric acid, 47% (v/v) deionised water for a minimum of 3 h, and washed for 20 min with deionised water (3 times). They were then incubated in 34% methanol, 3% phosphoric acid, 17% (w/v) ammonium sulphate and 63% water for 1 h. After 1 h a few grains of Serva blue G were added to the gel. Once all the protein bands had developed the gel was washed with deionised water.

2.2.5.6 Recombinant protein expression

A recombinant globular Type III AFP from eel pout was acquired from Prof. Peter Davies (Queen University, Kingston, ON, Canada). This was used a positive control for the ice-binding purification (Section 2.2.5.8). The protein was produced by the Davies group in the pET20b vector transformed into *E. coli* BL21 (DE3) cells. The cells were grown overnight on LB plates supplemented with ampicillin to a final concentration of 30 μg/ml. Single colonies were used to inoculate 3 ml cultures of LB ampicillin (30 μg/ml) and incubated at 37 ^{o}C with shaking at 250 RPM until an OD600 reading of approximately 0.5 was reached. This culture was used to inoculate 100 ml cultures of LB ampicillin (30 μg/ml) and grown at 37 ^{o}C with shaking at 250 RPM until an OD600 reading of 0.8 was reached. IPTG (isopropyl- β-D-1-thiogalactopyranoside) was added to a final concentration of 1 mM and the cells were induced at 37 ^{o}C with shaking at 250 RPM for 3 h.

2.2.5.7 Recombinant protein extraction

Three hour IPTG-induced *E. coli* cells were collected by centrifugation at 4000 x *g* for 30 min. The pellets were stored at -20 ^{o}C until required as defrosting aided in cell lysis. Pellets were defrosted on ice and proteins were extracted using Bugbuster protein extraction reagent (Novagen) according to the manufacturer's instructions. All cell lysis extracts were treated with Benzonase® Nuclease (Novagen) to remove contaminating nucleic acids according to the manufacturer's instructions.

2.2.5.8 Ice-shaping activity assay

A nanolitre osmometer (Otago Osmometers, Dunedin, NZ) was used to determine the presence of ice-binding proteins and to measure the freezing and melting temperatures of ice crystals grown in solutions of *Panagrolaimus sp.* proteins extractions. The nanolitre osmometer is a device that consists of a controller box, a cooling stage and a sample holder (Figure 2.1). The cooling stage is connected to a refrigerated water bath and placed onto a microscope containing a camera attachment. The water bath is set to -10 ^{o}C to cool the cooling stage prior to use. When the stage is cool the osmometer is set to $\geq$ 0 ^{o}C for loading the samples. A drop of oil (Cargille laboratories, Cedar Grove, NJ, USA) was placed into the wells of a sample holder and the holder was placed on the stage. A protein sample was extracted as described in Section 2.2.5.1, a 10 µl sample was required at a concentration of 1.5-2 mg/ml per assay. The protein sample was placed into the centre and under the drop of oil in the sample holder well using an elongated microcapillary tube with a mouth pipette (Sigma). Loading of the samples into the sample holder wells was done using the microscope at 10X magnification. The temperature of the protein sample was lowered to $\leq$ 0 ^{o}C before it was frozen quickly and slowly melted until only a single ice crystal remained. The morphology of the ice crystal was noted. The cryostage temperature was lowered further until the ice crystal began to grow spontaneously and then the cryostage was slowly warmed to arrest the growing process. The temperature at which the ice crystal growth arrested and remained stable was taken as the freezing temperature. The sample was then warmed further, and the temperature at which the ice crystal started to lose its shape and melt was taken as the melting temperature. Thermal hysteresis was calculated as the difference between the melting and the freezing temperatures of the ice crystal.

2.2.5.9 Ice-binding purification

A modified version of a brass ice-binding cold finger (Kuiper et al., 2003) was constructed by Mr Joe O' Sullivan (Biology Dept., NUIM) as depicted in Figure 2.2. The cold finger was constructed from brass tubing and silver solder. This was connected with insulated plastic tubing to a refrigerated water bath containing a 3:1 mixture of water and ethylene glycol. The protein sample was extraction as described in Section 2.2.5.2 and placed in 100 ml beaker and insulated with polystyrene foam.

Figure 2.1 The Otago nanolitre osmometer. It consists of a controller box and a cooling stage that contains a slot to place a sample holder. The cooling stage is connected to refrigerated water bath with a plastic tube containing coolant and place on a microscope with a camera attachment. The source of this image is www.otago-osmometers.com/nanolitreosmometers.html.

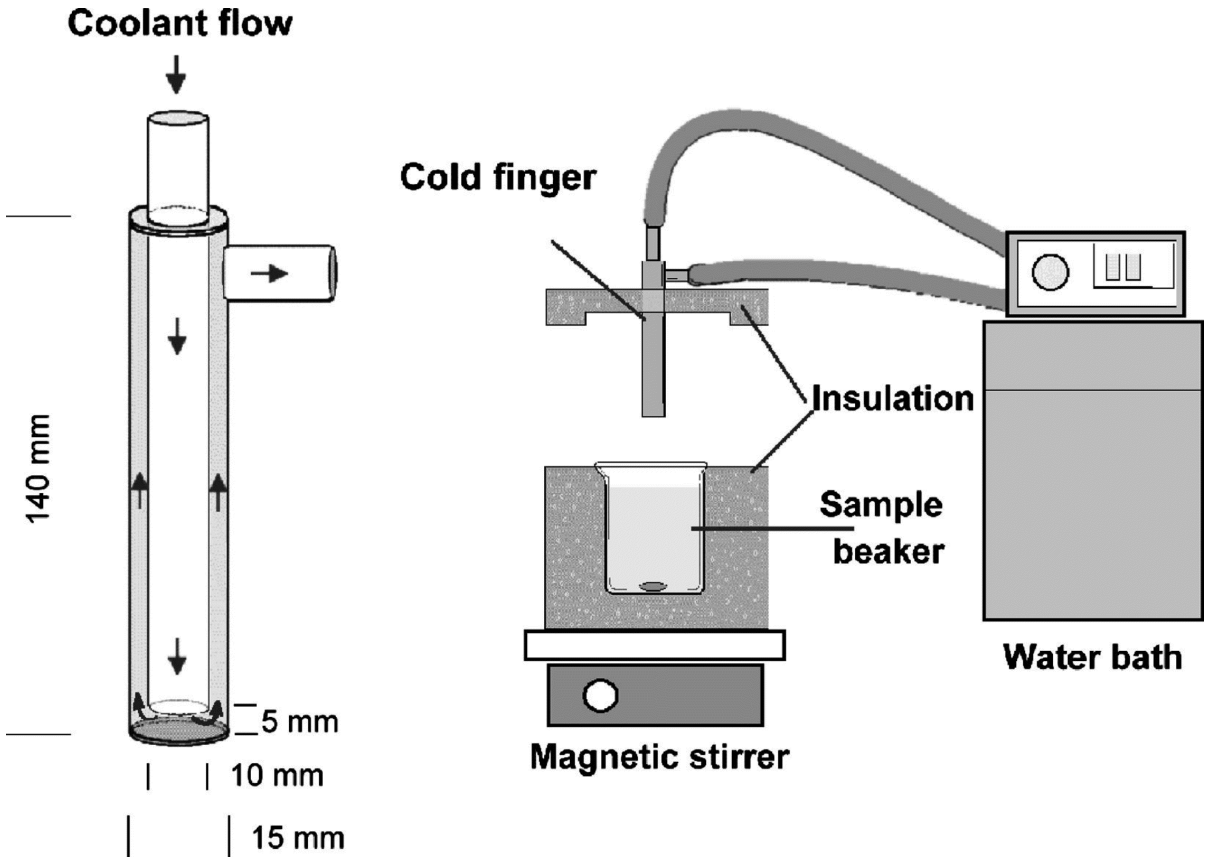


Figure 2.2 The cold finger apparatus. The cold finger is connected to a refrigerated water bath with insulated tubing. The cold finger is lowered into a beaker containing the protein sample that is placed on a magnetic stirrer for mixing of the sample. The source of this diagram is Kuiper *et al*. (2003).

The insulated beaker was placed on a magnetic stirrer to allow gentle mixing of the protein sample while the ice grows. The starting protein sample was diluted with 100 mM ammonium bicarbonate to give an approximate salt concentration of ≤100 mM and protein concentration of ≤1 mg/ml in a total volume of 80-100 ml. The cold finger was seeded for ice growth by cooling to –1.5 ^{o}C followed by immersion in chilled distilled water containing pieces of ice that will nucleate freezing, and form a thin layer of ice around the finger within 10 min. The nucleated cold finger was lowered into a pre-chilled starting protein sample with clearance at the bottom to allow the magnetic stir bar to rotate. The temperature of the cold finger was gradually lowered from -1.5 ^{o}C to -2.5 ^{o}C over 48 h to freeze approximately 50% of the original solution. After the run the cold finger was removed from the remaining "liquid" fraction and the "ice" fraction was washed with distilled water. The temperature was increased to 1 ^{o}C to detach the ice fraction from the cold finger into a clean beaker. This ice fraction was melted and used as the starting solution for a second round of purification. When the desired number of rounds of purification was completed the ice fraction was freeze-dried and resuspended in 10-20 μl of H_20. Starting, liquid and ice fraction protein concentration was determined and they were ran on an SDS-PAGE (2.2.2.4) gel prior to selection of bands for LC-MS/MS analysis (Section 2.2.5.10).

2.2.5.10 LC-MS/MS analysis

2.2.5.10.1 In-gel digestion for LC-MS/MS analysis

Ice-binding purified protein extracts from starting, liquid and ice fractions were separated using SDS-PAGE (Section 2.2.5.4). Selected protein bands were in-gel digested with trypsin (Shevchenko et al., 2006). All tubes were rinsed with acetonitrile to prevent keratin contamination. Briefly, the destained gel was rinsed with MilliQ H_2O for one hour. Specific bands were excised using a sterile scalpel and cut into 1 x 1 mm cubes. 100 μl of 100 mM ammonium bicarbonate/acetonitrile (1:1, vol/vol) was added and the gel slice was incubated at room temperature with occasional vortexing for 30 min to destain. The liquid was removed and 500 μl neat acetonitrile was added and the sample incubated at room temperature with vortexing until the gel piece had shrunk and become white in colour. The acetonitrile was removed and 50 μl of trypsin buffer (13 ng/μl in 10 mM ammonium bicarbonate

containing 10% (vol/vol) acetonitrile) was added. The solution was incubated for 120 min at 4 ^{o}C and the trypsin volume topped up to keep the samples completely saturated. 20 μl of 10 mM ammonium bicarbonate buffer was then added and the tubes incubated at 37 ^{o}C overnight.

2.2.5.10.2 Extraction of peptide digestion products for LC-MS/MS analysis

100 μl of extraction buffer (1:2 (v/v) 5% formic acid/acetonitrile) was added to the trypsin-digested gel pieces (Section 2.2.5.10.1) and these were incubated for 15 min at 37 ^{o}C in a shaker**.** The extraction buffer was added such that the approximate ratio between volumes of the digest and the extraction buffers was 1:2. The sample was briefly centrifuged, the supernatant removed and dried in a vacuum centrifuge.

2.2.5.10.3 LC-MS/MS analysis

Dried digested products (Section 2.2.5.10.1 and 2.2.5.10.2) were redissolved by adding 20 μl of 0.1% (v/v) trifluoroacetic acid, vortexing and incubating for 5 min in a sonication bath. The sample was centrifuged for 15 min at 10,000 RPM to withdraw an aliquot for analysis. The aliquot was added to a spin filter and centrifuged for 1 min to remove any remaining particulates.

The mass spectrometric analysis of peptides was carried out in the Proteomics Suite in NUI Maynooth with a Model 6340 Ion Trap LC/MS apparatus from Agilent Technologies (Santa Clara, CA). Separation of peptides was performed with a nanoflow Agilent 1200 series system, equipped with a Zorbax 300SB C18 5 mm, 4mm, 40 nl precolumn and a Zorbax 300SB C18 5 mm, 43 mm x 75 mm analytical reversed phase column using HPLC-Chip technology. The mobile phases utilised were A: 0.1% formic acid, B: 50% acetonitrile and 0.1% formic acid. Samples (5 μl) were loaded into the enrichment column at a capillary flow rate set to 4ml/min with a mix of A and B at a ratio 19:1 (v/v). Tryptic peptides were eluted with a linear gradient of 10-90% solvent B over 15 min with a constant nano pump flow rate of 0.60 μl/min. A 1 min post time of solvent A was used to remove sample carry over. The capillary voltage was set to 2000 V and the flow and the temperature of the drying gas was 4 μl/min and 300 ^{o}C, respectively.

A *P. superbus* transcriptome was sequenced from a mixture of cDNA obtained from unstressed nematodes (control) and those stressed by oxidative, heat, desiccation or cold. The transcriptome was prepared from the hybrid assembly by CAP3 from Newbler (version 2.5 pre-release) and MIRA (version 302r1) assemblies (Georgina O'Mahony, NUIM). This CAP3 assembly was then further improved by putting it through the CLoBB assembly program and the CLoBB improved by putting it through the Phrap assembly program. The Phrap assembly was processed through the prot4EST pipeline by Dr. Stephen Bridgett (University of Edinburgh). Proteins were identified from the protein bands of interest using the *P. superbus* transcriptome prot4EST database using Agilent Spectrum mill software and the Mascot MS/MS Ion search software (Matrix Science).

2.2.6 RNA-seq methods

2.2.6.1 RNA extraction and library preparation

An overview of the RNA-seq methods is shown in Figure 2.3. Nematodes were grown as described in Section 2.2.1.1 until a mixed population stage was reached. The plates of nematodes were acclimated at the selected temperature for the desired amount of time. Following this period of acclimation the nematodes were harvested and placed in Trizol Reagent® (Invitrogen) before freezing, as described in Section 2.2.1.3. RNA was prepared as described in Section 2.2.2.2. RNA quality was analysed using Qubit® 2.0 Fluorometer and the Agilent Bioanalyzer according to manufacturers instructions (Section 2.2.2.4).

The cDNA library construction for each treatment was performed by TrinSeq (TCD, Dublin) according to the Illumina TrueSeq RNA sample preparation low throughput protocol. The libraries were quantified using a Qubit dsDNA HS (high sensitivity) assay kit with the Qubit® 2.0 Fluorometer and the Agilent qPCR NGS library quantification assay. Following quantification samples were pooled for sequencing.

2.2.6.2 RNA-sequencing

Paired-end RNA-sequencing was performed on an Illumina Genome Analyzer II (GAII) over seven lanes of a flow-cell by TrinSeq. The resulting data was uploaded onto the Darwin cluster on the high performance computing facility at NUIM.

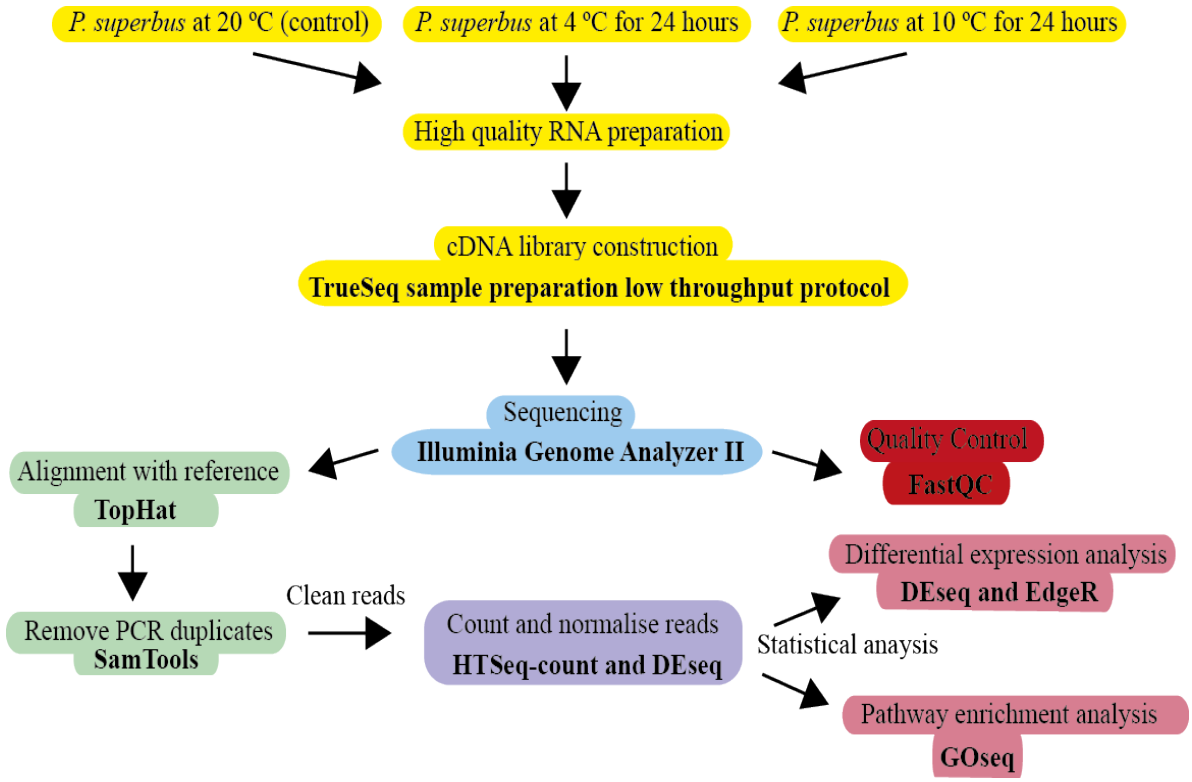


Figure 2.3. Overview of RNA-seq method. Software/protocols used in RNA preparation and sequence analysis.

2.2.6.3 Quality control

Sequencing reads from the Illumina sequencer were exported in FASTQ format with the corresponding Phred quality scores. The program FastQC was used to perform quality control (www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). This program performs a series of analyses that produce the following graphs and tables to give an overall assessment of the data: basic statistics, per base sequence quality, per sequence quality scores, per base sequence content, per base sequence content, per base GC content, per sequence GC content, per base N content, sequence length distribution, duplicate sequences, overrepresented sequences and overrepresented kmers.

2.2.6.4 Alignment of sequencing reads

The reference transcriptome was indexed using Bowtie (available at http://bowtie-bio.sourceforge.net/index.shtml). The reads from each sample were aligned to this indexed reference transcriptome. An example of the command used to build the transcriptome index is:

```
bowtie-build mytranscriptome.fa myindexedtranscriptome_1.0
```

This command calls Bowtie to index the specified FASTA formatted transcriptome and to output the indexes for the transcriptome in the specified filename. This creates six new files that constitute the index. These files were moved to the indexes folder in Bowtie for mapping. The following is an example of the command that was used to align the sequence reads to this indexed version of the transcriptome:

```
bowtie -S -m 1 myindexedtranscriptome_1.0 -1 s_1_1_sequence.txt -2
s_1_2_sequence.txt s_1_bowtie.alignments.txt
```

Briefly the flags in the script do the following: bowtie, calls the Bowtie software; -S specifies that the ouput is in SAM format; -m 1 specifies that it should suppress output for any reads that have more than one reportable alignment; myindexedtranscriptome_1.0 is the name of indexed transcriptome to which to align; -1 s_1_1_sequence.txt is the name of the one pair FASTQ file of reads of one pair from the sequencer; -2 s_1_2_sequence.txt is the name of the second pair FASTQ file

of reads from the sequencer; s_1_bowtie.alignments.txt is the name for the output file of the alignments of the reads to the transcriptome.

2.2.6.5 Removal of PCR duplicates

Removal of PCR duplicates (identical sequence reads) was preformed using SAM tools (Li et al., 2009). SAM Tools provide various utilities for manipulating alignments in the SAM format. The alignments were viewed and converted to binary (BAM) format using the command:

```
samtools view –b –t <index filename> -o <output.bam> <input.sam>
```

This tells Sam tools to view the index and the input alignment file in SAM format and convert it to BAM format. The BAM alignment files are sorted with the command:

```
samtools sort –o <sorted filename > <input.bam>
```

This command tells Sam tools to sort the input BAM format alignment files and to output them in the specified filename also in BAM format. The following command is used to remove PCR duplicates:

```
samtools rmdup –o <sorted and duplicates removed filename> <sorted
filename>
```

This removes all duplicates from the inputted sorted alignment file and outputs the alignment file with all duplicates removed into the newly specified filename.

2.2.6.6 Counting mapped reads

The number of reads that mapped to a transcriptome identified gene sequence were counted with HTSeq-count, part of the python package HTSeq available at www.huber.embl.de/users/anders/HTSeq/doc/overview.html. This program requires a SAM alignment input format, so BAM files were converted back to SAM format with SAM tools. This program will take the reads and count the number of times they mapped to a specific to a given gene annotation in the transcriptome (gff file). The following command was used to count the number of reads per gene:

```
HTSeq-count [options] –m union –t exon –s no <sam_file> <gff_file> -o
<output.sam>
```

HTSeq-count is used to call the software; -m union sets a mode that allows the program to handle reads that overlap more than one genomic feature; -t exon is the default and tells the program what the type of feature the ids in the gff file are; -s no is whether the data is from a strand-specific assay, the default is yes so this flag was specified as no. These commands were followed by the alignment file in SAM format and the gff file containing the list of the reference transcriptome gene annotations. The –o command if the output filename in SAM format.

2.2.6.7 Normalisation of counts

A single text file was constructed containing all counts for each biological replicate for each transcriptome gene id. The counts for different treatments and replicates were normalised to adjust for varying lane sequencing depths using the R package DEseq. Before normalisation the data must be formatted into the correct structure. The tab-delimited file of the count data with column headers in the first line was read in with:

```
countsTable <- read.delim("examplefile", header=TRUE,
stringsAsFactors=TRUE);
```

The table structure should look like the following example if the head(countsTable) command is used:

	gene	TR1	TR2	T1R3	C1R1	C1R2	C1R3
1	Gene_00001	0	0	2	0	0	1
2	Gene_00002	20	8	12	5	19	26
3	Gene_00003	3	0	2	0	0	0
4	Gene_00004	75	84	241	149	271	257
5	Gene_00005	10	16	4	0	4	10
6	Gene_00006	129	126	451	223	243	149

In this example the first column is the gene ID, the actually gene IDs for each row of counts is in the second column so this column is specified as containing the gene ID for each count and the first column is removed by:

```
rownames(countsTable)<- countsTable$gene;
countsTable<- countsTable[, -1];
```

This table now has six named columns that refer to each of the conditions and the count data for each gene. It should look like this if the head command is used:

```
> head(countTable);
                 TR1     TR2     T1R3    CR1     CR2     CR3
      Gene_00001 0       0       2       0       0       1
      Gene_00002 20      8       12      5       19      26
      Gene_00003 3       0       2       0       0       0
      Gene_00004 75      84      241     149     271     257
      Gene_00005 10      16      4       0       4       10
      Gene_00006 129     126     451     223     243     149
```

In this example the first three labelled are three replicates of a treatment (T1) and the last three are three replicates of a control (C1). Samples that are replicates of the same treatment or control are designated the same value “T” or “C”. The information that some of the samples are replicates is applied, and assigned a vector name (conds):

```
conds<-c(“T”, “T”, “T”, “C”, “C”, “C”);
```

When this is complete it is possible to form the *CountDataSet* class that is the central data structure in the DEseq package. This is done using the following command:

```
cds<- newCountDataSet(countsTable, conds);
```

The data may now be normalised by estimating the size factors from the count data. The is done using the function estimateSizeFactors and to apply them to the count data set is:

```
cds<- estimateSizeFactors (cds);
```

2.2.6.8 Multidimensional scaling plot

The R package edgeR (Robinson et al., 2010) was used to make a multidimensional scaling plot (MDS) from the counts data for each condition. This is an initial overall estimation of differences in the count data between each condition and each replicate before any statistical testing. The package edgeR is called by the command:

```
library(edgeR);
```

The plot is drawn using the command:

```
d<-DGEList(counts=d, group=conds);
plotMDS.dge(d, main="MDS Plot for Data");
```

2.2.6.9 Estimating variance and resulting plots

The core assumption of the model employed by DEseq is that the mean is a good predicator of the variance. For each condition a function that allows the variance to be predicted from the mean must be estimated. This computation is performed by the following command:

```
cds<- estimateVarianceFunctions (cds);
```

A squared coefficient of variation (SCV) plot to visualise whether the package estimated the variance functions accurately so that it fits the data well can be drawn with the command:

```
scvPlot(cds);
```

Another diagnostic plot to check the fit of the variance function can be drawn with the command:

```
diagForT<-varianceFitDiagnostics(cds, "T");
```

2.2.6.10 Heatmap plot

A plot was drawn using DEseq to visualise how each of the replicates of each condition clusters together. It applies as a variance stabilising function (VST) that transforms count data such that its variance becomes independent of the mean. In this example the CountDataSet is given a new name (cds2), initially the same commands are used as Section 2.2.6.7 then the following commands may be used:

```
cds2<- newCountDataSet (countsTable, conds);
cds2<-estimateSizefactors(cds3);
```

```
cds<-estimateVarinaceFunctions(cds3);
```

Then the function getVarianceStabilizedData is called:

```
vsd<-getVarianceStabilizedData (cds);
```

A distance calculation then may be applied and the distance matrix visualised as a heatmap using the following commands:

```
heatmap(as.matrix(dis), symm=True)
```

2.2.6.11 Calling differential expression

Having estimated the variance-mean dependence (Section 2.2.6.9), the number of differentially expressed genes may be calculated. To compare two conditions to see whether there is any differential gene expression the function nbinomTest is called. The nbinom test uses a negative binomial model i.e. the number of counts for a given gene in sample j come from a negative binomial distribution with the mean $s_j\mu_p$ and the variance $s_j\mu_p + s_j^2 v(\mu_p)$ where s_j is the relative size of library j, μ_p is the mean value for condition p and $v(\mu_p)$ is the fitted variance for the mean μ_p. The null hypothesis for the nbionom test is that the experimental condition has no influence on the expression of the gene under consideration ($\mu_{p1} = \mu_{p2}$) (Anders and Huber, 2010).

In an example were the treatment is T and the control is C the following command is used:

```
res<- nbinomTest(cds, "N", "T");
```

This returns a dataframe with the fold-change between the treatment and the control and all of the supporting statistical data. The nbinom test uses a negative binomial model for statistical testing (Anders and Huber, 2010). The fold-change for each gene may be visualised using the command:

```
plot(
+ res$baseMean,
+ res$log2FoldChange,
+ log="x", pch=20, cex=.1,
```

```
+ col=ifelse( res$padj <.1, "red", "black"));
```

The significantly up/downregulated genes are seen in red, while the non-significant are black. This dataframe may be saved as a tab delimited file and imported into an Excel file for further analysis.

2.2.6.12 Pathway analysis

The biochemical pathways that are significantly differentially expressed were determined using the R package GOseq (Young et al., 2010). The GOseq package can be loaded into R using the command:

```
library(GOseq);
```

The Kyoto Encyclopaedia of Genes and Genomes (KEGG) database, the *C. elegans* database and the database of genomic features were also loaded by the following commands:

```
library(org.Ce.eg.db);
library(KEGG.db);
library(GenomicFeatures);
```

In order to do this analysis the gene IDs must be in a form that R can recognise that can also be recognised by KEGG database. To do this all of the *P. superbus* transcriptome was analysed by BLAST against the *C. elegans* genome (version 190 downloaded from Ensembl at www.ensembl.org). Those with hits to *C. elegans* were used for the pathway analysis and the *P. superbus* gene IDs were changed to *C. elegans* Ensembl gene IDs. Separate tab delimited files with a list of all the genes expressed in a dataset (with *C. elegans* IDs), with a list of only the differentially expressed genes ($P<0.01$ and 2 fold-change) and with the statistical analysis for all of the genes expressed were created. The data were then ready for analysis by GOseq. The table of all the genes and the table of only the significantly up-regulated genes were read into GOseq using the commands:

```
de.genes<-scan("T1VC1_sig.CeWS190"), what character ());
all.genes<-scan("T1VC1_all.CeWS190.txt"), what character());
```

In this example the file containing the list of the significantly up-regulated genes that have been converted into *C. elegans* IDs between the treatment (T1) and control (C1) is read in first and this is followed by the file containing all of the differentially expressed genes that were found. The table containing of the statistics relating to these genes was read in using the command:

```
deseq.results<-read.table("allstats_resultsT1VC1.CeWS190.txt",
header=TRUE, as.is=c
(TRUE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE,
FALSE));
```

The vectors containing all the genes to be assayed (all.genes) and the significantly differentially expressed genes (de.genes) can now be used to construct a named vector suitable for use with GOseq using the following command:

```
genes=as.integer(all.genes%in% de.genes)
names(genes) = all.genes
pwf = nullp(genes, "ce6", "ensGene");
```

The mappings of the *C. elegans* Ensembl IDs to the KEGG pathways are stored in the package org.Ce.eg.db. In order to test for KEGG pathway over representation in the differentially expressed genes this information needs to be extracted and put into a format suitable for GOseq. The mappings between the ENSEMBL gene ID and KEGG pathway must be constructed. This was done with the following code:

```
en2eg=as.list(org.Ce.egENSEMBL2EG)
eg2kegg=as.list(org.ce.egPATH)
grepKEGG=function(id, mapkeys){
     unique (unlist(mapkeys[id], use.names=FALSE))
     };
```

The probability weighting function (PWF) is applied that gives the probability that a gene will be differentially expressed based on its length alone. This is produced as follows:

```
kegg= GOseq(pwf, gene2cat=keg);
```

The identification of the pathways that are overrepresented and the corresponding P-value for each pathway can be calculated and exported into a text file with the commands:

```
KEGGpathsIDs2names<-as.list(KEGGPATHID2NAME);
Enriched.KEGG=KEGG[KEGG$over_represented_pvalue<0.05,];
write.table(enriched.KEGG, "KEGG_enriched.txt", sep="\t");
getname = function(ID) {
+      as.list(KEGGPATHID2NAME)[[ID]]
+ };
names = as.matrix(lapply(enriched.KEGG$category, getname));
```

The bioinformatic analysis of the RNA-Seq data was carried out under the guidance of Dr. Chris Creevey （Animal and Grassland Research and Innovation Centre, Teagasc, Dunsany, Co. Meath).

Chapter III An investigation of the phylogenetic relationships and the cryotolerance and desiccation tolerance phenotypes of species and strains of *Panagrolaimus* from polar, subpolar and temperate regions.

3.1 Introduction

The genus *Panagrolaimus* contains many anhydrobiotic species, which occupy niches ranging from polar and subpolar to semi-arid soils and terrestrial mosses. In this chapter, I present data on the freezing and desiccation tolerance of fifteen isolates of *Panagrolaimus* representing three species and eleven unassigned strains. The geographic origins of the *Panagrolaimus* species used in this study are presented in Table 2.1. One of the *Panagrolaimus* species in this study (*P. superbus*) was isolated from a gull's nest on Surtsey Island, Iceland 18 years after the volcanic formation of the island. Other isolates are from high altitude locations such as Iceberg Lake in Mount Whitney, California (PS1732) and Byurkan, Aremenia (PS443), from continental locations (*P. paetzoldi*), polar (*P. davidi*) soils or from temperate soils, roof-moss (AS03) and roof-gutter (AS01) samples.

In this chapter, the survival ability of *Panagrolaimus* sp. nematodes following exposure to freezing and desiccation stress is assessed and compared to their phylogenetic relationships. The effect of acclimation and preconditioning on stress survival is also investigated. Tissue extracts from selected strains are assessed for their ability to prevent the growth of ice. Sequences of regions from the SSU and LSU were used for species identification and their resulting multiple sequence alignments were used to construct phylogenetic trees. The anhydrobiotic phenotypes of six of these strains have been described previously by Shannon *et al.* (2005) and the anhydrobiotic and cryobiotic phenotypes of the Antarctic nematode has been described by Wharton *et al.* (1991; 1993).

Low temperatures and freezing conditions affect the lives of all organisms. The conditions may impair biochemical reactions and cause changes in membrane fluidity and protein conformation. Growing ice crystals can cause significant mechanical

damage to the subcellular structures and extracellular ice growth can cause water to flow out of the cells leading to dehydration (Margesin et al., 2007). Certain species have adapted to survive in these conditions for long periods of time. In animals this is achieved by either freeze-avoidance or freeze tolerance (Storey and Storey, 1996). Existing morphological features such as waterproof cuticles may be used to avoid freezing. Freeze-avoiding animals can also supercool, maintaining the liquid state below the equilibrium freezing point. Supercooling can be enhanced by antifreeze proteins (Davies et al., 2002) or cryoprotectant production (Storey, 1997). There is some evidence that nematodes can employ a freeze-avoidance strategy. The unhatched second stage larva of *Globodera rostochiensis* retains its egg shell as a protective sheath to resist freezing and allow supercooling to -25 ^{o}C (Perry and Wharton, 1985). The nematode *P. davidi* can also survive by freeze avoidance, its eggshell prevents inoculative freezing and adults can also supercool when free of surface water (Wharton and Brown, 1991).

When temperatures drop slowly and to below a certain point, animals lose their body water and dehydrate instead of freezing. This is known as cryoprotective dehydration (Holmstrup and Westh, 1994; Holmstrup *et al.*, 2002). *P. davidi* may survive freezing by cryoprotective dehydration (Wharton et al., 2003). The success of cryoprotective dehydration as a survival strategy in nematodes depends on the nucleation temperature, the cooling rate and the freezing rate of the surrounding medium. It is already known that some *Panagrolaimus* species and strains are capable of surviving extreme desiccation (Wharton, 2003; Shannon *et al.*, 2005). This is achieved by entering a state of suspended animation known as anhydrobiosis. Anhydrobiosis is an almost complete loss of body water and no metabolic activity is detectable. The *Panagrolaimus sp.* and other anhydrobiotic organisms can stay in this state for a prolonged period of time, after which upon rehydration they are able to resume normal metabolic activity. Anhydrobiotic nematodes produce sugars, in particular trehalose that protect the lipid bilayer from injury during desiccation (Crowe and Crowe, 1992). As both freezing and desiccation stress cause cellular dehydration they may employ similar molecular adaptations. In this study the ability of *Panagrolaimus* strains to survive desiccation stress will be assessed and compared to freezing survival.

A period of cold acclimation has been shown to improve freezing tolerance of many organisms including nematodes. Cold acclimation allows an individual time to conduct the necessary biochemical and metabolic adjustments that are required to survive the freezing temperatures that may follow. This may involve the accumulation of cryoprotectants. In insects, cryoprotectants glycerol and sorbitol are synthesised from glycogen and glucose-6-phosphate over a period of weeks when exposed to low temperatures (Joanisse and Storey, 1995). In freezing tolerant organisms, cell membranes are protected by increasing the concentration of phospholipids with unsaturated fatty acids and decreasing in saturated fatty acids. This remodelling of the cell membrane lipids is seen in plants during cold acclimation (Miguel *et al.*, 1993; Thieringer *et al.*, 1998). Cold-acclimation has also been shown to up-regulate expression of a number of genes, including antifreeze proteins and heat shock proteins (Antikainen and Griffith, 1997; Hofmann, 2005). A period of cold acclimation has also been shown to enhance nematode survival to freezing such as *P. davidi*, *P. redivivus* and *S. carpocapsae* (Wharton *et al.*, 2000; Jagdale and Grewal, 2003; Hayashi and Wharton, 2011). In this study, in addition to their ability to survive freezing when directly exposed to -80 ^{o}C, the affect of acclimation *Panagrolaimus* sp. freezing survival will be investigated.

In this study, the 18S gene, the D3 expansion region of the 28S gene and the IT1 and ITS2 regions were sequenced from 15 *Panagrolaimus* strains and species to confirm the identity and to construct a phylogeny from their concatenated multiple sequence alignment. The alignment for the 18S gene was improved by performing a secondary structure alignment. The SSU is a non-coding RNA molecule and its function is dependent on the 3D-structures rather than the primary sequence. Including this structural information in sequence alignment and tree construction can reduce the effects of multiple substitutions and insertion deletion events that commonly cause errors in phylogenetics (Letsch et al., 2010). In this analysis the program RNAsalsa was used to structurally align the sequences (Stocsits et al., 2009a) as outlined in Section 2.3.2.4. The method implemented by RNAsalsa accurately takes into account a known accurate secondary structure, primary sequence information and use well-established folding algorithms to generate a consensus alignment. The phylogenetic methods used in this study to infer the phylogeny of this group of nematodes were maximum likelihood (ML) (Edwards, 1972) and Bayesian (Yang and Rannala, 1997).

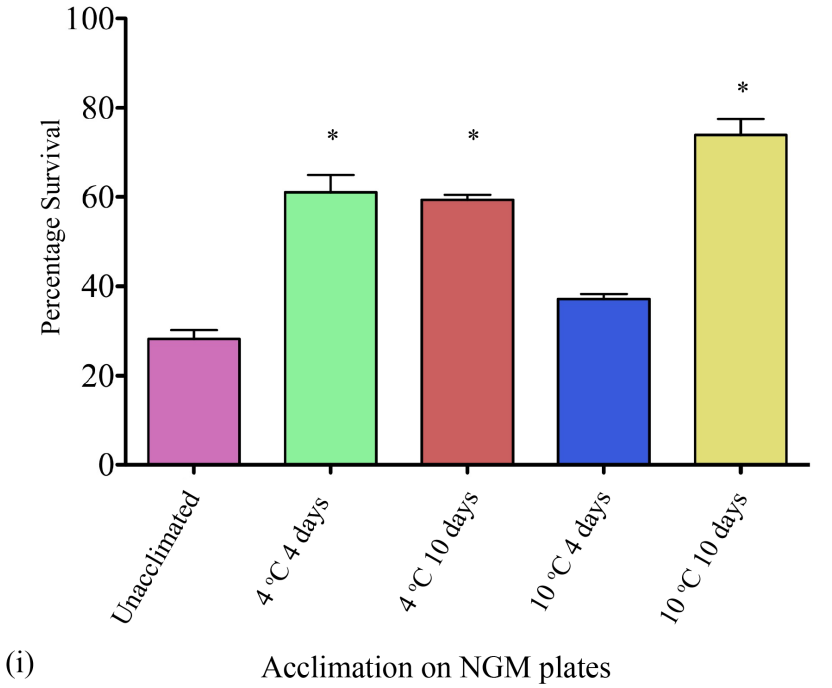
Maximum likelihood calculates a set of bootstrap values and Bayesian calculates posterior probabilities to describe the accuracy of each branch in the phylogenetic trees. For the purposes of this study posterior probability values will be considered equivalent to bootstrap values, although they are usually not considered to be as stringent as bootstrap values (Simmons et al., 2004).

3.2 Results

3.2.1 Effect of cold acclimation temperature and time on freezing survival of *Panagrolaimus superbus*

The effect of acclimation at different temperatures over varying periods of time was investigated using the nematode *P. superbus* (Figure 3.1). The nematodes were left to grow on NGM plates that contained bacteria and the NGM plates were then incubated at either 4 ^{o}C or 10 ^{o}C for 4 or 10 days (Figure 3.1 (i)). Acclimation at 4 ^{o}C for 4 days, 4 ^{o}C for 10 days and 10 ^{o}C for 10 days resulted in a significant increase in freezing survival (ANOVA $P<0.05$ when compared the unacclimated control). A *post-hoc* Tukey's Test revealed that the optimal acclimation condition is 10 ^{o}C for 10 days ($P<0.001$). This regime was used for all subsequent acclimation experiments.

Shannon *et al.* (2007) had investigated the effect of cold acclimation in *Panagrolaimus sp.* The protocol used by Shannon was as follows: the nematodes were washed off the NGM plates and adjusted to 2000 per ml. They were placed in 1 ml aliquots in a 1.5 ml Eppendorf tube and put in a waterbath at 4 ^{o}C or 10 ^{o}C for 6 h. In addition to exposing the nematodes to lower temperatures, this regime may also be subjecting them to caloric restriction, as the nematodes were deprived of their bacterial food during acclimation. The effect of this acclimation on freezing survival was investigated (Figure 3.1 (ii)). The results obtained show that incubation at 4 ^{o}C for 6 h produced a significant increase in survival compared to the unacclimated control but the survival values obtained following acclimation at 10 ^{o}C for 6 h was not significantly different from those of the unacclimated controls. The survival value was for 4 ^{o}C in water was still much lower than when the nematodes were acclimated at 10 ^{o}C on NGM plates for 10 days (> 20%).



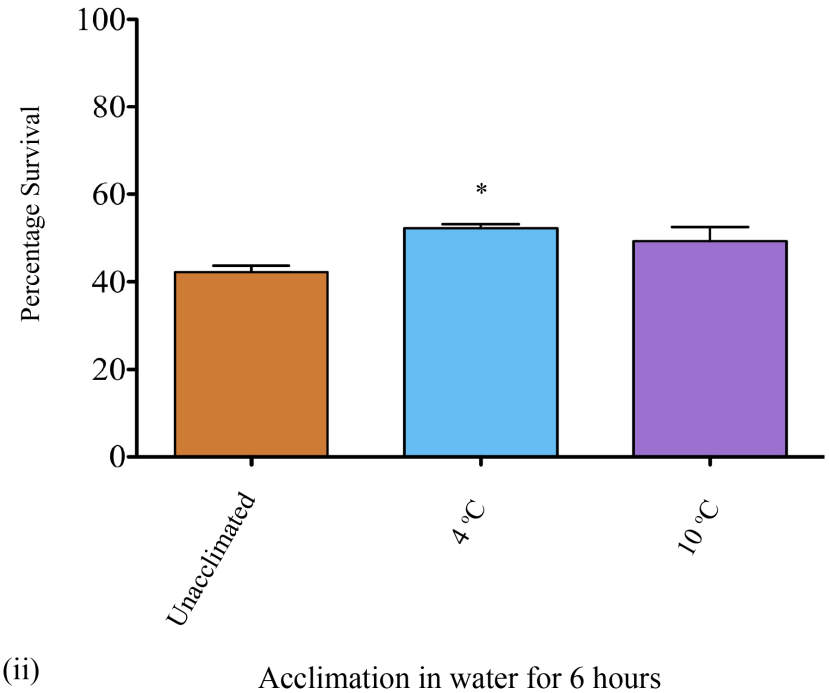


Figure 3.1. The effect of acclimation temperature and time on freezing survival *in P. superbus*. The nematodes were either acclimated on NGM plates containing *E. coli* for different times and temperatures (i) or in water at different temperatures for 6 h (ii). Survival values are the mean +/- SEM of 4 replicates. (*= $P<0.05$ in ANOVA).

3.2.2 Effect of freezing on survival of *Panagrolaimus* sp.

The effect of the freezing treatment (with or without acclimation) on the survival of fifteen strains of *Panagrolaimus* was investigated and the results are presented in Figure 3.2. The source and geographic location of these strains is presented in Table 2.1. Nematodes were cultured and harvested as described in Section 2.2.1.1 and 2.2.1.3. The rate of freezing to -80 ^{o}C was determined to be -0.05 ^{o}C per second using a low temperature probe. Statistical analysis using the ANOVA (P<0.05) and Bonferroni *post-hoc* tests suggested that significant differences existed between survival levels among the strains. Twelve of the fifteen strains (*P. superbus,* JB051, SN103, PS1579, PS1159, AS03, PS443, *P. davidi*, AF36, Galileo desert sp., JB115, AS01) have a level of initial freezing tolerance that varies greatly between strains. Acclimation at 10 ^{o}C for 10 days significantly improved freezing survival for eleven out of fifteen of the strains (*P. superbus,* PS1579, PS1159, AS03, *P. davidi*, AF36, Galileo desert sp., JB115, AS01, PS5056 and *P. paetzoldi*) for ANOVA at P<0.01. The strains SN103, JB051 and PS443 show high initial freezing tolerance but a non-significant increase in freezing survival following acclimation. The strain PS1732 is freezing sensitive and thus is not affected by acclimation.

3.2.3 Effect of desiccation on survival of *Panagrolaimus* sp.

Fifteen strains of *Panagrolaimus* species and strains were preconditioned at 98% relative humidity (RH) for various time periods (0-96 h) followed by exposure to dried silica gel for 48 h (Figure 3.3). Nematode survival following rehydration was then estimated by microscopic observation. Statistical analysis using ANOVA test revealed significant differences between survival levels among strains and the Bonferonni *post-hoc* test showed that they could separated into three groups. *P. superbus* has high desiccation survival (>70%) when exposed directly to 0% relative humidity (RH)(silica gel) and may be called a “fast dehydration strategist”. The majority of the strains have a range of desiccation survival from 1% to 51% survival that significantly increases when they are preconditioned at 98% RH over varying lengths of time. These strains may called the “slow dehydration strategists” and comprises of SN103, *P .davidi*, AF36, JB051, AS03, PS1159, PS443, JB115, Galileo desert sp., PS1579, AS01 and PS5056. The strains PS1732 and *P. paetzoldi* showed no desiccation tolerance and thus could not be improved by a preconditioning step. These strains are “desiccation sensitive”.

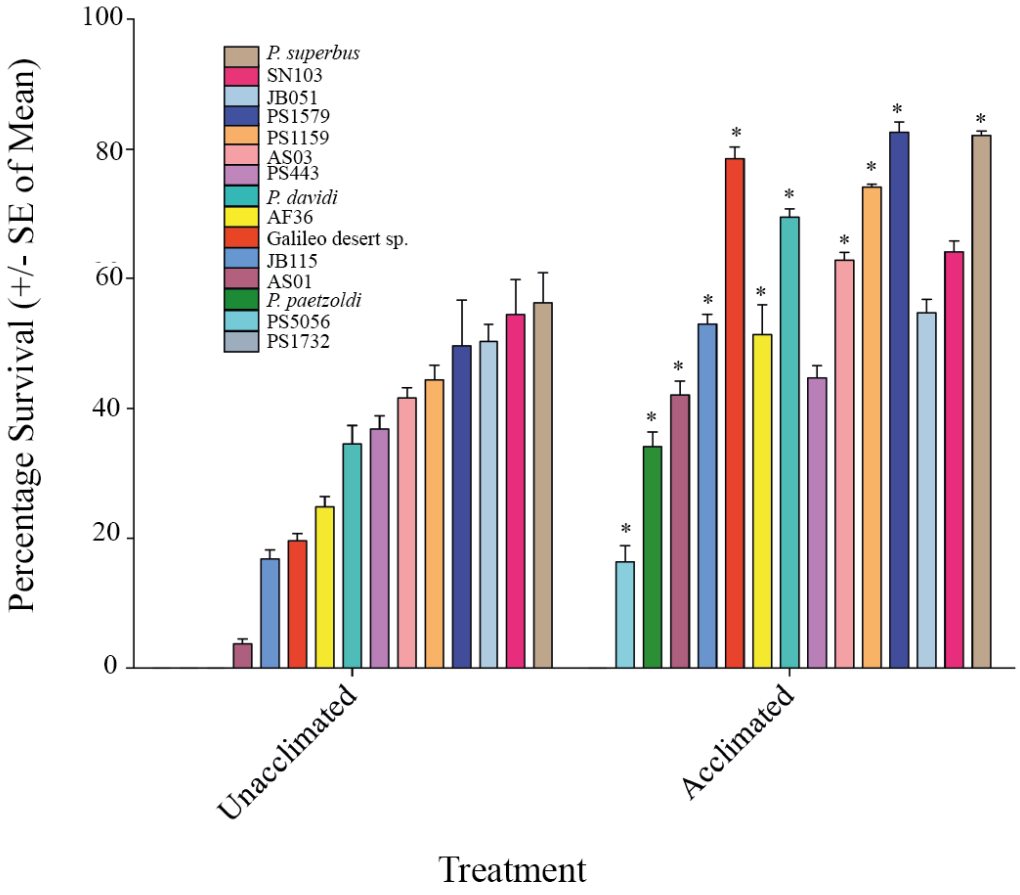


Figure 3.2. The effect of cold acclimation on the freezing survival of 15 strains of *Panagrolaimus*. Unacclimated: Direct exposure to -80 ^{o}C for 24 h. Acclimated: Acclimation at 10 ^{o}C for 10 days followed by exposure to -80 ^{o}C for 24 h. Survival values are the mean of three replicates +/- the standard error. (*=P<0.01 in Bonferonni *post-hoc* test). The strains PS5056 and *P. paetzoldi* showed 0% freezing survival in the unacclimated treatment. The strain PS1732 showed 0% freezing survival in the unacclimated and acclimated treatments.

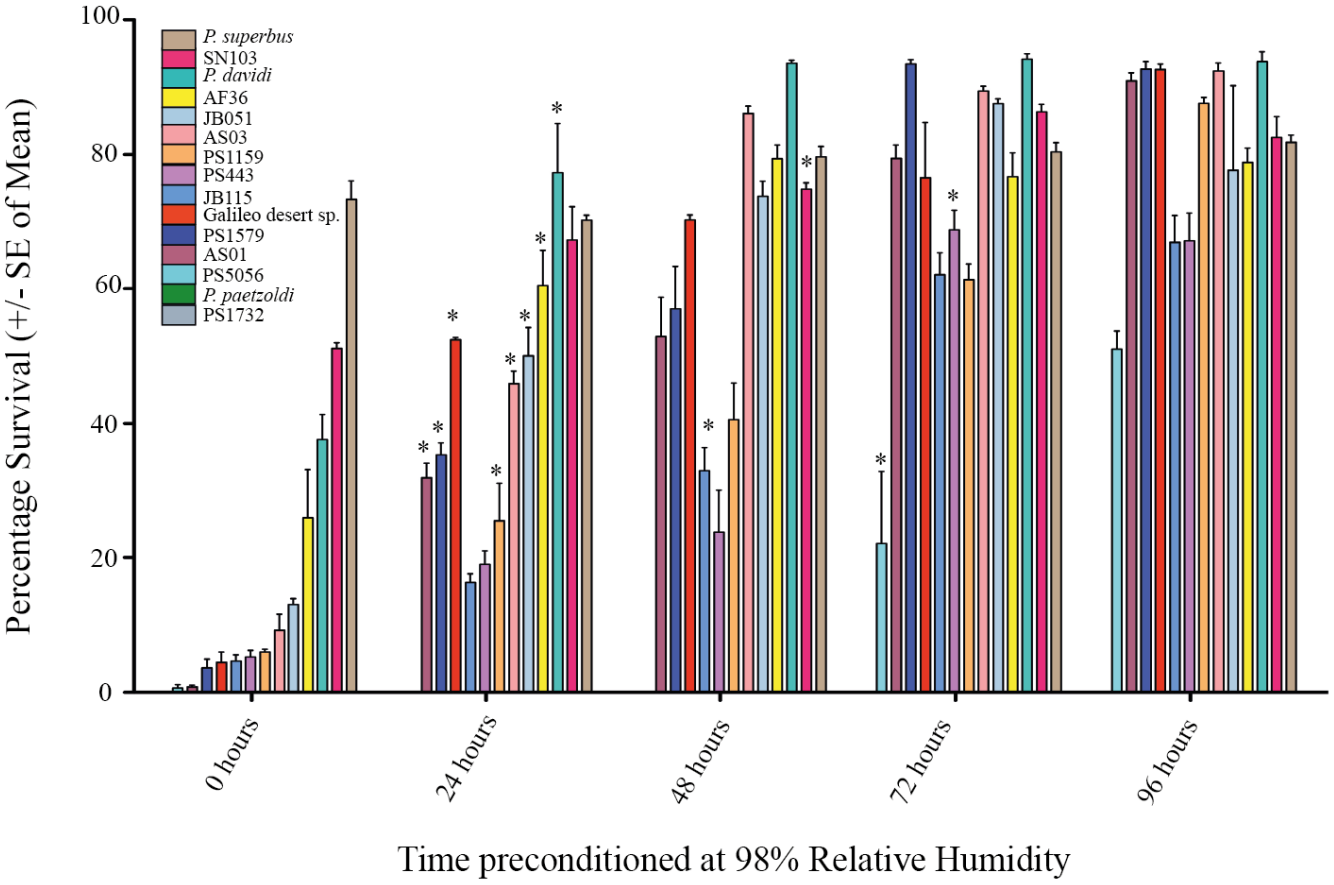


Figure 3.3. The effect of preconditioning time at 98% RH on the anhydrobiotic survival of 15 strains of *Panagrolaimus*. Treated nematodes were preconditioned at 98% RH at 20 ^{o}C for 0-96 h. The nematodes were then desiccated for 48 h over freshly activated silica gel and hydrated in H_20 for 24 h before survival was determined. Survival values are the mean of three replicates +/- the standard error (*=P<0.01 in Bonferonni *post-hoc* test). AS01 had 0% desiccation survival after 0 h preconditioning treatment. PS5056 had 0% desiccation survival after 0, 24 and 48 h preconditioning treatments. The strains PS1732 and *P. paetzoldi* had 0% desiccation survival after all treatments.

Regression analysis revealed that for the slow dehydration strategists there is a significant relationship between the time preconditioned and survival with the exception of PS5056. The minimum time of preconditioning at 98% RH required for a significant increase in survival (ANOVA $P<0.01$) was 24 h for the strains and species *P. davidi*, AF36, JB051, AS03, PS1159, Galileo desert sp., PS1579 and AS01. A minimum of 48 h was required for SN103 and JB115 and a minimum of 72 h for PS443 and PS5056. *P. superbus* has very high desiccation survival values following direct exposure to 0% RH that is not further improved by a significant level by preconditioning.

3.2.4 The relationship between freezing and desiccation tolerance in *Panagrolaimus*

Regression analysis suggests that a correlation exists between the maximal survival levels observed for freezing and desiccation in acclimated strains of *Panagrolaimus* (Figure 3.4). The maximal freezing value is the percentage survival following cold acclimation and at 10 ^{o}C for 10 days the maximal desiccation value is the percentage anhydrobiotic survival following 96 h at 98% RH. The R-squared value of 0.6291 indicates a good fit of the points to the fitted line in the regression plot. The majority of the strains and species form a group (in red box). In this group all nematodes with over 40% freezing survival also have over 60% desiccation survival.

3.2.5 Ice-shaping activity

Whole tissue protein extracts from a selection of *Panagrolaimus* strains and species were tested for thermal hysteresis activity their ability to shape ice crystals. Protein was extracted from the nematodes and adjusted to a final concentration of 1.5- 2 mg/ml (described in Section 2.2.5.1). The sample was tested using a nanolitre osmometer and visualised under a microscope that was connected to a camera. The *Panagrolaimus* sp. had either no or very low thermal hysteresis activity (0.03 ^{o}C) but did show shaping. The nematodes *P. superbus*, PS1159 and *P. davidi* (Figure 3.5 (i), (ii) and (iii)) produced a very clear hexagonal ice crystal shape. The Galileo sp. and AS01 (Figure 3.5 (iv) and (v)) tissue extracts also shaped the ice but the edges of the hexagonal appeared slightly more rounded. This distinct faceted morphology is characteristic of extracts containing ice-binding proteins. The protein extract from the

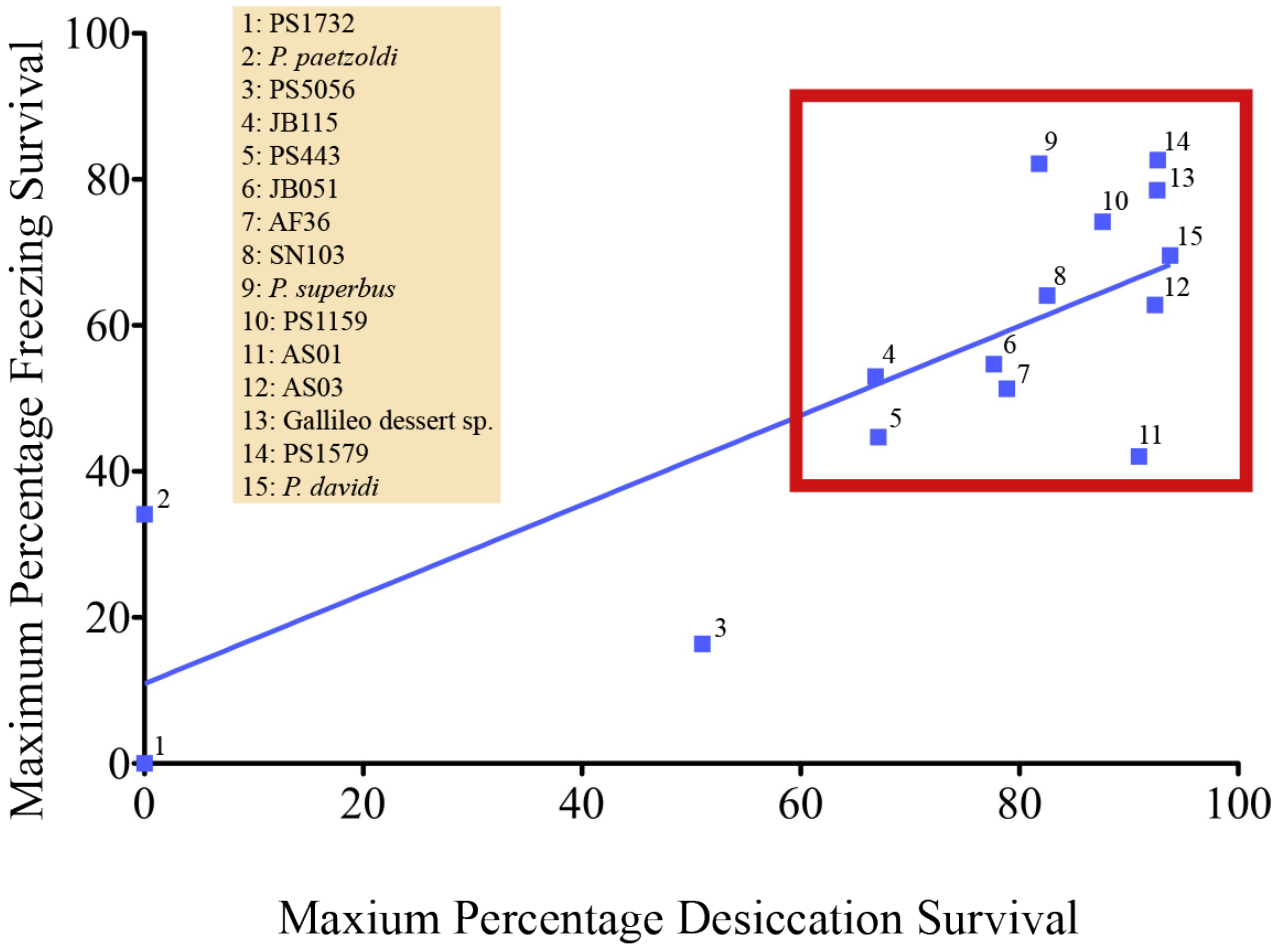


Figure 3.4. Regression analysis of maximal freeze and desiccation survival levels in fifteen strains and species of *Panagrolaimus*. The X-axis shows the maximum percentage desiccation survival for each *Panagrolaimus* species and strain. The Y-axis shows the maximal percentage survival for freezing survival. The R-squared value was 0.6291. The numbered points indicate the identity of each species or strain.

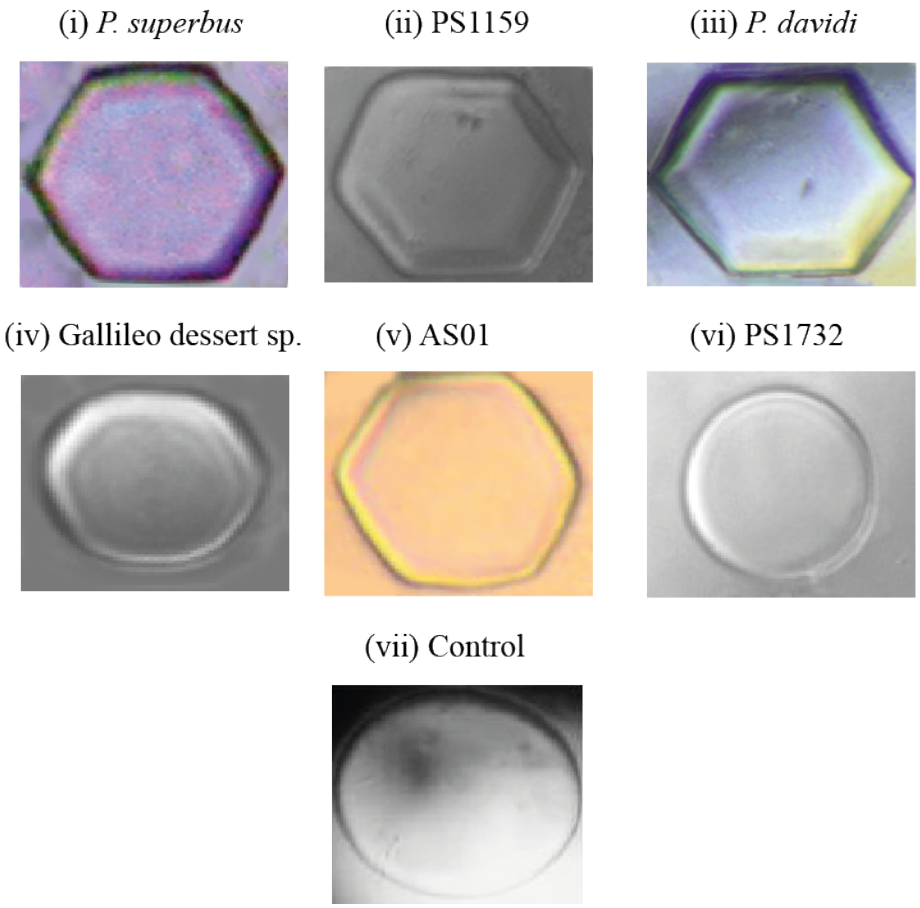


Figure 3.5. The morphology of ice crystals formed in the presence of whole tissue protein extracts of *Panagrolaimus* species and strains. The crystal formed in the presence *P. superbus*, PS1159, *P. davidi*, Galileo desert sp. and AS01 (i) – (v) have a distinct faceted morphology characteristic of ice-binding proteins. The crystals formed in the presence of PS1732 (vi), a freeze-sensitive strain, and the controls (100 mM Ammonium bicarbonate), are round, indicating an uninhibited growth of the ice crystal.

freeze sensitive strain PS1732 (Figure 3.5 (vi)) did not shape the ice and has the same round shape as the control (Figure 3.5 (vii)).

3.2.6 Ribosomal RNA-sequencing

The 18S gene, D3 expansion region and the ITS ribosomal regions were sequenced from the *Panagrolaimus* species and strains and checked for sequence identity by BLASTN searches against the GenBank database. This was undertaken to confirm the identity of *P. superbus*, SN103, JB051, PS1579, PS1159, AS03, PS443, *P. davidi*, AF36, JB115, AS01, *P. paetzoldi* and PS1732 in this collection were derived from the expected nematodes. The nematodes Galileo desert sp. and PS5056 were morphologically determined to be *Panagrolaimus* sp. by Dr. John DeModena (Personal communication), sequences confirmed the genus identification by BLASTN of the rRNA 18S and D3 against the GenBank database. The entire 18S rRNA gene from the LSU was successfully amplified and sequenced for the strains and species used in this study. This gene is over 1500bp in size so was sequenced in two overlapping parts. *Panagrolaimus* specific primer sets were designed for each of the parts such that when sequenced the region could be assembled using the CAP3 assembly program (Table 2.2). The assembled 18S rRNA gene was approximately 1700 bp in each *Panagrolaimus sp*. All the 18S rRNA gene sequences had blast hits to *Panagrolaimus* nematodes in the GenBank database. *P. davidi* had a 100% match with its corresponding full-length sequence contained in the database, confirming the identity of this nematode. There are only partial 18S sequences available in GenBank for *P. superbus*, SN103, JB051, PS1579, PS1159, PS443, AF36, JB115, *P. paetzoldi* and PS1732. These nematodes with the exception of two (*P. paetzoldi* and PS1732) had 98-100% match to the partial sequences in the database, confirming that the identity of these nematodes in this collection is correct. *P. paetzoldi* has similarity to a full-length sequence in the database (87%), if the identity of this species was correct the identity match would be higher. The 18S gene from the nematodes AS03, AS01, Galileo desert sp. and PS5056 were not previously sequenced so there are no sequences available in the database. The high similarity of these new strains to other *Panagrolaimus* nematodes confirms the correct genus classification of these new strains. The PS1732 sequence had no similarity to that found in GenBank.

The D3 expansion region of the LSU was amplified effectively using standard PCR conditions for all of the strains and species. The desired PCR product was of approximately 300 bp in length. The D3 sequences were also checked for the correct identity against the GenBank database. The strains and species SN103, PS1579, AS03, PS443, *P. davidi* and AS01 have 100% matches with the corresponding sequences in the database. The nematodes *P. superbus*, JB051, PS1159, JB115 and *P. paetzoldi* had 98-99% identity to those in GenBank. The high identity matches of these nematodes for the 18S gene and the D3 region with the corresponding sequences in GenBank gives confidence that the identity of these nematodes is correct. AF36 is only 93% similar to the sequence in the database, however as the 18S sequence is considerably longer in size and had 98% identical to the sequence in the database this nematode is presumed to be correct. The D3 region for PS5056 and Galileo desert sp. sequences are not available in GenBank. They have similarity of over 90% to other *Panagrolaimus* nematodes further confirming that they were classified into the correct genus. The D3 region for PS1732 sequence had no similarity to that found in GenBank.

The ITS region that includes the 5.8S region was amplified. Initially two sets of general nematode ITS primers were used, TW81 and AB28 or rDNA1 and rDNA2 (Vrain *et al.*, 1992; Joyce *et al.*, 1994). Neither set of primers were reliable for the *Panagrolaimus* strains and species, so specific primers were designed. Existing *Panagrolaimus* sequences were aligned to locate conserved regions from which more robust primers could be designed (ITS1F and ITS2R). This amplified region was approximately 700 bp when sequenced. Previously annotated ITS regions were aligned against the sequences to determine the location of ITS1, 5.8S and ITS2 rRNA. Of the strains and species studied the ITS sequences from the following strains, *P. superbus,* SN103, PS1579, *P. davidi*, AS03, AF36, Galileo desert sp., JB115, AS01 and PS1732 were successfully amplified and sequenced. The ITS1 and ITS2 sequences were analysed against the GenBank database. Of the nematodes that have ITS sequences in the database none had 100% similarity for either region. This may be because this region tends to be highly polymorphic and evolve rapidly. The manner that certain ITS1, 5.8S and ITS2 sequences are annotated in the database may also differ from the method that was used in this study.

3.2.7 Construction of a phylogenetic tree for the genus *Panagrolaimus* based on 18S sequences.

The 18S (SSU) sequences for the *Panagrolaimus* species and strains were aligned using the MUSCLE alignment software with *Panagrellus redivivus* as the outgroup. *P. redivivus* was selected as a suitable outgroup, because it is a member of the superfamily Panagrolaimidae and in the expectation that it might therefore be possible to align the sequences of *P. redivivus* with those of *Panagrolaimus*. This alignment was further improved by performing a structural alignment with *Saccharomyces cerevisiae* as a reference using the software RNAsalsa (Stocsits et al., 2009a). Once the sequences were aligned, they were trimmed to equal length. Figure 3.6 shows the bootstrap consensus phylogeny inferred using the Maximum Likelihood (ML) model with 1000 bootstrap replicates. The Bayesian model was also used for the same group of nematodes for comparison. The ML and Bayesian trees both have the same topology, however branch supports derived from posterior probabilities tended to be higher. AS01 and Galileo desert sp. are grouped outside from the other strains and species with high support (81% ML, 94% Bayesian). The remaining nematodes may be separated into three clades. PS443, AF36 and *P. superbus* form one clade with medium support (81% ML, 74% Bayesian). *P. paetzoldi*, AS03 and PS1732 form a second clade with high support (80% ML, 100% Bayesian). SN103, PS5056, JB115, JB051, PS1159, PS1579 and *P. davidi* form a third clade with very high support (90% ML, 100% Bayesian). The 18S tree does not show any clear link in these strains and species between phylogeny and freezing or desiccation survival.

3.2.8 Construction of a phylogenetic tree for the genus *Panagrolaimus* based on D3 sequences.

The D3 region (LSU) sequences for the *Panagrolaimus* species and strains were aligned using the MUSCLE alignment software with *P. redivivus* as the outgroup. ML and Bayesian methods were also used to construct the phylogenetic trees (Figure 3.7). The topology of both trees is the same. The support values for the majority of the nodes are very low (less than 70%). The sequences of these genes have high similarity, therefore there are too few nucleotide substitutions give any high resolution and support values for the relationships between these strains and species. The majority of the strains and species (PS443, JB115, AS03, PS1159, PS1579, *P.*

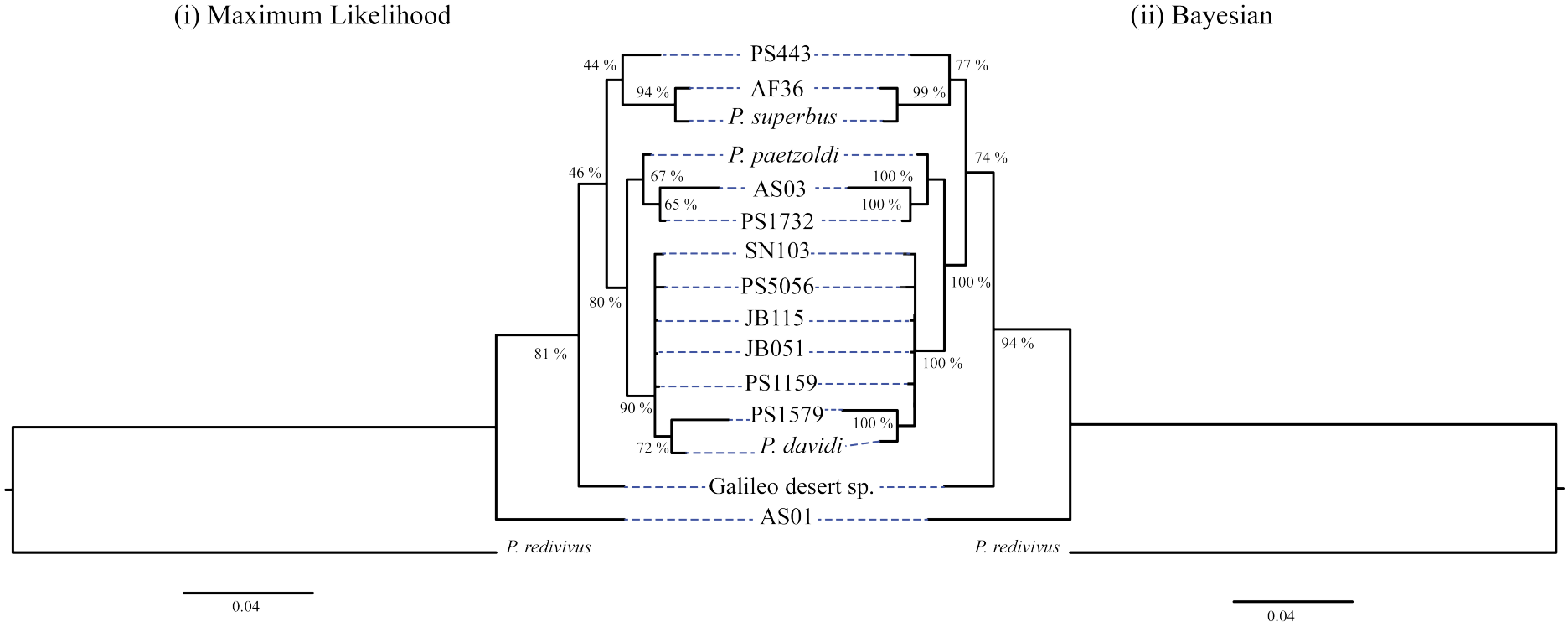


Figure 3.6. Phylogenetic trees of species and strains within the *Panagrolaimus* genus constructed using 18S sequences. (i) The phylogenetic tree obtained using the Maximum likelihood (ML) method. (ii) The phylogenetic tree obtained using the Bayesian method. The scale bar on each tree indicates the number of substitutions per site. The dotted blue lines show a comparison between the topologies generated using each method. The bootstrap support values are indicated at each node.

Figure 3.7. Phylogenetic trees of species and strains within the *Panagrolaimus* genus constructed using D3 sequences. (i) The phylogenetic tree obtained using the Maximum likelihood (ML) method. (ii) The phylogenetic tree obtained using the Bayesian method. The scale bar on each tree indicates the number of substitutions per site. The dotted blue lines show a comparison between the topologies generated using each method. The bootstrap support values are indicated at each node.

davidi, JB051, PS1732) fall into one clade that has medium support (65% ML, 89% Bayesian).

3.2.9 Construction of a phylogenetic tree for the genus *Panagrolaimus* based on 18S and D3 sequences.

The alignment of the 18S and D3 sequences were concatenated and used to construct a third phylogenetic tree (Figure 3.8) using both ML and Bayesian methods. Both methods returned phylogenetic trees having the same topology. The nodes closest to the root have high support values, reducing towards the closer related nematodes. Similarly to the separate phylogenies AS01 is grouped outside of the other nematodes. *P. superbus* and AF36 form a single clade with very high support (99% ML and Bayesian). There is another clade formed that contains the majority of the strains (*P. paetzoldi*, PS1732, *P. davidi*, PS1579, SN103, JB051, PS1159, JB115, Galileo desert sp. and PS5056). This node has high support (81% ML, 99% Bayesian) but the inner relationships lack any significant resolution. The strains AS03 and PS443 are just outside of this main clade. This phylogenetic analysis has demonstrated that the 18S and D3 are not the ideal genes for resolving the internal relationships in the genus *Panagrolaimus*.

3.2.10 Construction of a phylogenetic tree for the genus *Panagrolaimus* based on ITS sequences.

As the 18S gene or the D3 regions did not resolve the closer relationships, the ITS region was sequenced. Only the nematodes *P. superbus,* SN103, PS1579, *P. davidi*, AS03, AF36, Galileo desert sp., JB115, AS01, PS1732 sequences were obtained as it is very difficult to amplify and sequence. Previously annotated ITS regions were aligned against the sequences to determine the location of ITS1, 5.8S and ITS2. The individual ITS1 and ITS2 sequences were each aligned using the MUSCLE alignment software. Both alignments contained regions where the sequence was conserved but the sequences have high levels of nucleotide substitutions, along with several insertions and deletions, as clearly seen in Figure 3.9 and 3.10. These INDELs cause problems in aligning the sequences and these resulting trees were not informative and had poor resolution so are not presented here.

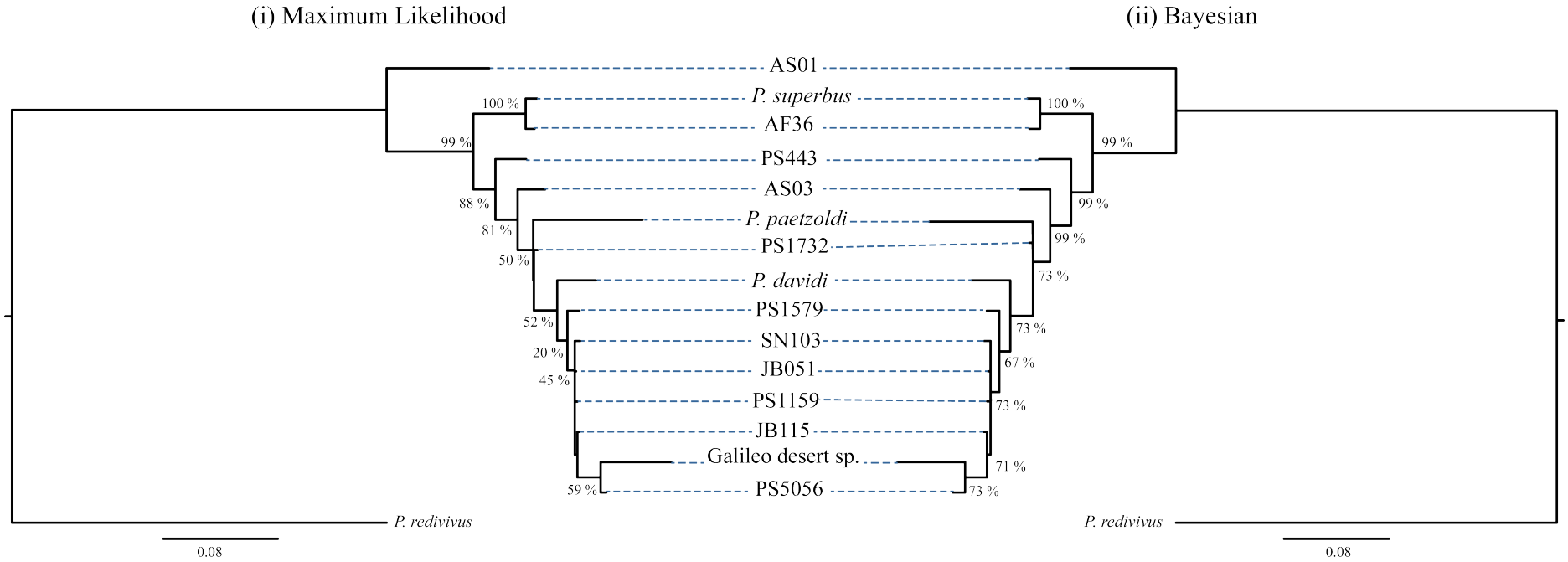


Figure 3.8. Phylogenetic trees of species and strains within the *Panagrolaimus* genus constructed using 18S and D3 sequences. (i) The phylogenetic tree obtained using the Maximum likelihood (ML) method. (ii) The phylogenetic tree obtained using the Bayesian method. The scale bar on each tree indicates the number of substitutions per site. The dotted blue lines show a comparison between the topologies generated using each method. The bootstrap support values are indicated at each node.

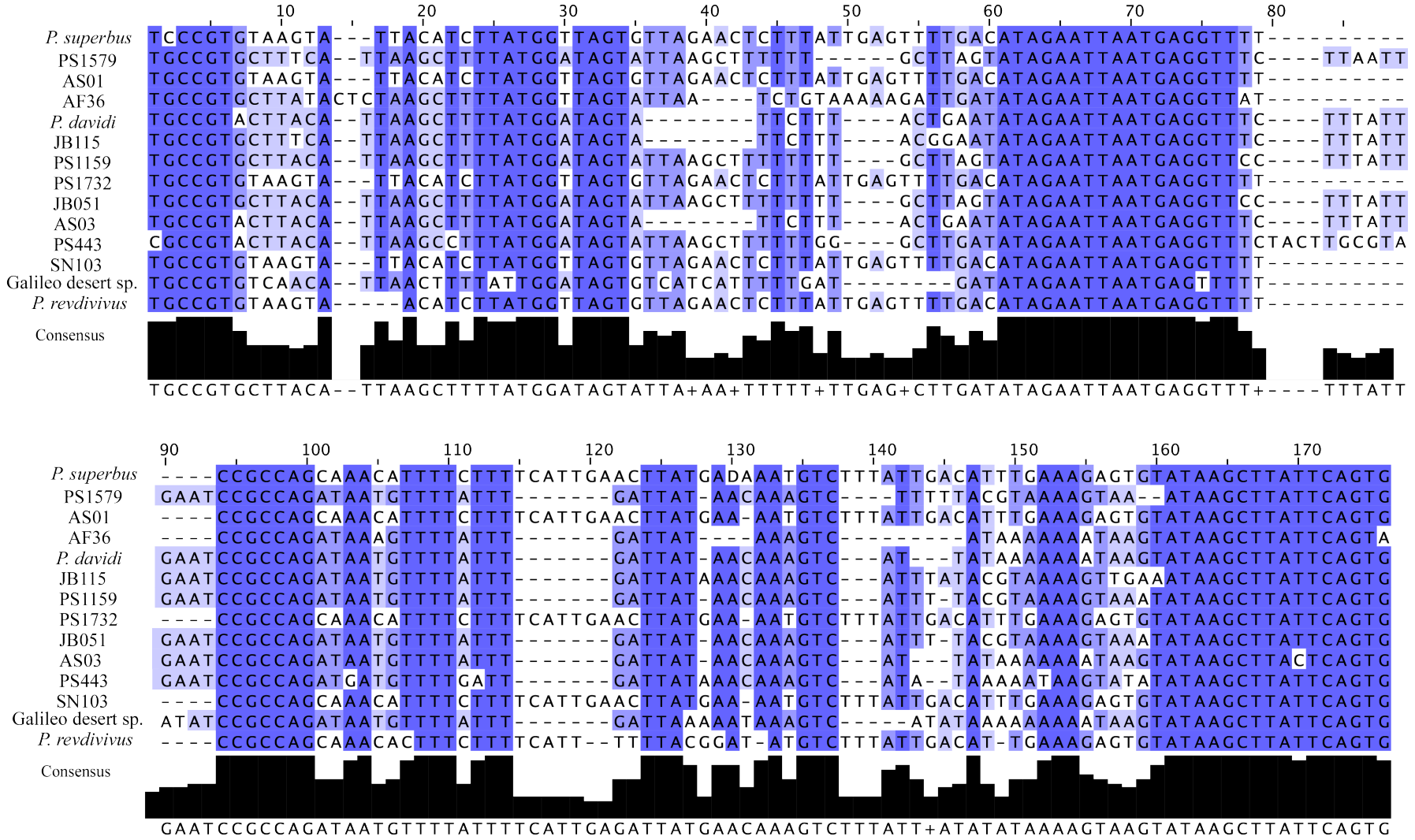


Figure3.9. Multiple sequence alignment of the internal transcribed spacer 1 region. The alignment is coloured according to percentage identity where regions with high conservation are darker purple. Areas of conservation are also denoted by the black graphic at the bottom of the alignment.

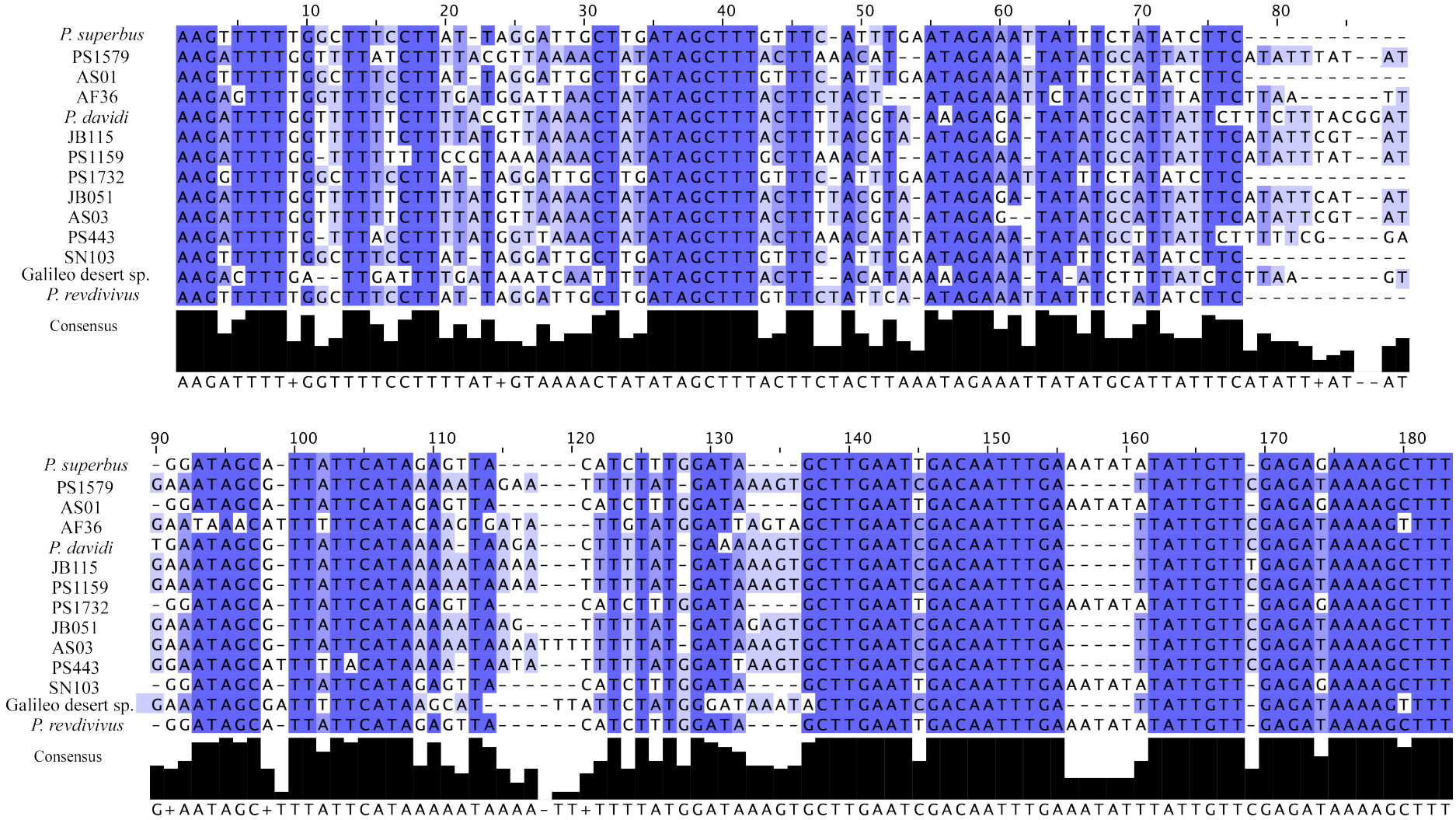


Figure3.10. Multiple sequence alignment of the internal transcribed spacer 2 region. The alignment is coloured according to percentage identity where regions with high conservation are darker purple. Areas of conservation are also denoted by the black graphic at the bottom of the alignment.

3.3 Discussion

Nematodes are often exposed to periods of environmental stress such as freezing and desiccation. Several nematodes species have evolved behavioural, biochemical and physiological adaptations to cope with these stresses. This has enabled them to exploit diverse ecological niches in unfavourable or unpredictable environments ranging from desert Antarctic soils to plant foliage.

The Antarctic nematode *P. davidi* is a model species for the study of freeze tolerance in the genus *Panagrolaimus* (Wharton and Brown, 1991). It can survive temperatures down to -80 ^{o}C and is one of the few animals that can survive internal ice formation (Wharton and Ferns, 1995). The fifteen strains of *Pangarolaimus* investigated in this study display a wide range in their cryotolerance ability. Thirteen of the fifteen strains show some survival upon direct exposure to very low temperature, so these nematodes could be described as freeze survivors to some degree. Cold acclimation has been shown to improve survival in *Meloidogyne hapla* larvae and the animal-parasitic nematode *Nematodirus battus* eggs (Forge and MacGuidwin, 1990), *S. carpocapsae* (Jagdale and Grewal, 2003) and *P. redivivus* (Hayashi and Wharton, 2011). In this study cold acclimation at 10 ^{o}C for 10 days significantly improved survival following exposure to -80 ^{o}C of eleven of the fifteen strains, including two of the freezing sensitive strains (PS5056 and AS01). The strain PS1732 cannot survive freezing and its survival was not improved by cold acclimation. Cold acclimation did have a significant effect on the three medium survivors, PS443, JB051 and SN103. The survival values for unacclimated and cold acclimated *P. davidi* are lower than previous studies (Wharton *et al.*, 2000; Smith *et al.*, 2008). The survival values observed in this project for *P. davidi* were 35% (unacclimated) and 70% (cold acclimated) whereas Wharton *et al.* (2000) and Smith *et al.* (2008) had much higher values. This may be due to a difference in methods and temperatures used in the experiments. If the rate of cooling is sufficiently slow at high subzero temperatures, enough water is lost from the nematode to prevent freezing by cryoprotective dehydration (Wharton et al., 2003). In this study the rate of freezing was 0.05 ^{o}C per second. This is slow so survival of freeze-tolerant nematodes in this study is likely caused by cryoprotective dehydration. The improved freezing survival following acclimation may be due to an increase in low-molecular weight cryoprotectants and ice-active proteins. Trehalose and glycerol protection increase is associated with

acclimation in *P. davidi* (Wharton et al., 2000) and trehalose is also increased in cold acclimation in *S. feltiae*, *S. carpocapsae* and *S. riobrave* (Jagdale and Grewal, 2003) so may also be increased in these strains and species. Chapter 7 describes an experimental investigation of the genes that are up-regulated in response to cold acclimation in *P. superbus.*

While in the anhydrobiotic state, nematodes and other anhydrobiotic animals show increased resistance to many other environmental stresses including high and low temperatures, ionising radiation and metabolic poisons (Freckman *et al.*, 1980; Demeure and Freckman, 1981; Jonsson, 2003; Horikawa *et al.*, 2009). There are two broad categories of anhydrobiotic nematodes recognised: fast and slow dehydration strategists (Womersley, 1987). Fast-dehydration strategists can inhabit environments exposed to frequent desiccation and they can survive immediate and prolonged exposure to rapid dehydration. Fast-dehydration strategists may be pre-adapted at a cellular level to survive desiccation, possibly by constitutively expressing protective molecules required for anhydrobiosis, or by sequestering the mRNAs required for the synthesis of these protective molecules. Alternatively they may have behavioural (coiling, clumping) or morphological adaptations which slow the rate of water loss (Perry, 1999).

The genus *Panagrolaimus* contains both fast and slow dehydration strategists (Wharton and Barclay, 1993; Shannon *et al.*, 2005). The fifteen strains of *Panagrolaimus* investigated in this study also showed a great diversity in their anhydrobiotic ability. The 73% survival observed for *P. superbus* following direct exposure to activated silica gel for 48 h is remarkable and provides an excellent example of a fast dehydration strategist. The majority of free-living anhydrobiotic nematodes found in soil habitats experience slow rates of water loss and are believed to be slow dehydration strategists. Slow-dehydration strategists are unable to survive exposure to extreme desiccation unless they have first experienced a period of preconditioning to moderate reductions in RH. This provides the nematodes with sufficient time to induce the necessary biochemical changes required to survive desiccation. All the *Panagrolaimus* strains used in this study except PS1732 can survive direct exposure to silica gel to a certain degree, so they could be described as fast dehydration strategists using Womersley's description. However with the

exception of the "fast desiccation strategist" *P. superbus* and the "poor desiccators" PS1732 and *P. paetzoldi* all of the strains had minimal to medium survival when exposed directly to silica gel but show substantial improvements in their desiccation tolerance to low RH following preconditioning at 98% RH (SN103, *P. davidi*, AF36, JB051, AS03, PS1159, PS443, JB115, Galileo desert sp., PS1579, AS01, PS5056). These strains are probably best described as slow-dehydration strategists. Only 24 h of preconditioning was sufficient to give a significant improvement in survival to several of the strains. Some of these strains (*P. superbus*, AF36, PS1159, PS1579, PS443, PS1732) were used in a previous study (Shannon et al., 2005), slightly different survival values were obtained but the same trend is apparent in both datasets. Thus it appears that within the genus *Panagrolaimus* there is an obvious variety in survival abilities and induction strategies among the strains.

Among the many environmental stresses that nematodes encounter, low temperature and desiccation are thought to be the most closely linked (Crowe et al., 1983). Regression analysis comparing maximal desiccation and freezing survival levels in the fifteen *Panagrolaimus* sp. suggested that a correlation may exist between desiccation tolerance and freezing tolerance in these nematodes. Cryoprotective dehydration seems to be an important freeze tolerance strategy used by invertebrates in cold deserts (Ring and Danks, 1994) so many of the physiological and molecular responses to cold may have originally been adaptations for desiccation stress (Danks, 2000). At the cellular level both stresses cause similar consequences, increases in solute concentration and osmotic stress, reduction in the supercooling point and membrane shrinkage. Therefore it is not surprising that generally the nematodes that can survive freezing can also survive desiccation.

Many freeze-tolerant organisms synthesise proteins that assist their survival by interacting with the growing ice. These ice-binding proteins may include AFPs and RIPs. The AFPs inhibit the growth of ice crystals, producing a lowering of the freezing point below the melting point termed thermal hysteresis (reviewed by (Ewart et al., 1999). The RIPs are subset of AFPs that do not have thermal hysteresis activity but have important functions in controlling the size, shape and location of ice crystal growth (Yu et al., 2010). The nematode *P. davidi* has been shown to produce an ice-binding protein that has minimal thermal hysteresis activity, so it is believed to be a

RIP (Wharton et al., 2005). Whole cell protein extracts that contain ice-binding proteins shape single ice crystals into distinct "facet" shapes than can be seen in a nanolitre osmometer ice binding experiment. In this experiment protein extracts from six strains including *P. davidi* were observed for thermal hysteresis activity and the ability to shape single ice crystals. In agreement with the previous data for *P. davidi*, the strains used in this experiment do not have significant thermal hysteresis activity. The protein extract from the freeze-sensitive strain PS1732 does not shape the ice and the ice crystals generated in the presence of ice crystals from this nematode have the same round shape as those of the buffer control. There is a correlation between the strains that produce more distinct shaping and those that have higher survival rates during freezing. These results suggest that freeze-tolerant strains produce an ice-binding protein that many have a role in freezing survival. The more sensitive strains may also produce an ice-binding protein but with less affinity for the ice, resulting in a more rounded ice-crystal shapes. Thus the freezing-tolerant nematodes used in this study most likely synthesize a protein which prevents the growth of ice crystals (possibly a RIP), but which is lacking in TH activity.

Due to their conserved morphology nematode phylogenetic relationships have been unresolved for decades. However, the SSU (Blaxter *et al.*, 1998; Holterman *et al.*, 2006; van Megen *et al.*, 2009) and the LSU (Litvaitis et al., 2000) rRNA-sequences have proved to be very useful in resolving the internal relationships within the nematodes. The 18S gene and the D3 expansion region were used to construct a phylogeny of the 15 strains and species used in this study. A similar approach was used to study the origins of parthenogenesis and hermaphroditism in 31 *Panagrolaimus* nematodes (Lewis et al., 2009). These neighbour joining and maximum parsimony phylogenetic trees generated by Lewis *et al.* were derived from the concatenated alignments of 436 nucleotides from the 18S and 623 from the 28S rRNA. Eleven of the strains used in this current phylogenetic study were also used by Lewis *et al.* (2009), however in this current study the full 18S rRNA-sequence and a shorter sequence from the 28S rRNA D3 region were used. The results in these phylogenies reported in this project do differ slightly from this previous study. In the separate and concatenated phylogenies the strains SN103, PS1159, JB115, *P. davidi* and PS1579 do form a single clade with *P. davidi* and PS1579 grouping together in the 18S phylogeny, in agreement with that of Lewis *et al.* (2009). The difference

between the two phylogenies is between the location of the nematodes PS1732 and *P.paetzoldi*. In this study we find that these nematodes group between the main clade and the group containing *P. superbus*, AF36 and PS443, whereas in Lewis *et al.* (2009) these nematodes were grouped outside. This result combined with the low identity of our *P. paetzoldi* 18S sequence to that in GenBank makes us question whether the identity of this nematode is correct. Due to time constraints and the difficulty in growing this strain this nematode 18S rDNA was not resequenced. The rDNA sequences that were obtained for the strain PS1732 did not have matches to PS1732 sequences in GenBank. However this nematode was recently re-acquired from its original source (Dr. John Demodena), cultured and sequenced immediately so we believe that this strain of PS1732 is authentic.

Previously a relationship has been shown to exist between phylogeny and desiccation survival in *Panagrolaimus* using alignments of the D3 expansion region and the ITS1 and ITS2 (Shannon et al., 2005). If there were a clear relationship between phylogeny and stress tolerance it would be expected that the desiccation and freezing sensitive strains would be phylogenetically distinct from the tolerant strains. The strain AS01 is a weak freezing and desiccation survivor when not exposed to a period or acclimation or preconditioning and it is located on a branch outside of the other nematodes. But the strains that had 0% survival when exposed to either freezing or desiccation (*P. paetzoldi*, PS1732 and PS5056) are found within clades that contain stress tolerant nematodes. These phylogenetic analyses do present an interesting insight into the relationships in the genus. The vast difference in stress tolerance between closely related strains and species within the same genus may indicate that the number of genes needed to achieve successful freezing or desiccation survival may be relatively small. The positioning of weak stress survivor strains within the phylogenetic trees and variation in survival values also suggest that the ability to acquire desiccation and freezing tolerance may evolve rapidly.

The phylogenetic trees gave very poor resolution and support values for the closely related nematodes. The SSU and the LSU are very conserved and slowly evolving molecules so they contain a lack of sufficient informative sites to distinguish between the closely related sequences. The fast evolving ITS regions were sequenced in an attempt to further resolve some of the closely related strains. These regions were

difficult to amplify but after several attempts a more robust primer set than the general nematode primers published were designed specific for *Panagrolaimus.* Both of the ITS alignments had areas of high conservation and areas with insertions, deletions and mutations. These alignments cause problems in phylogenetic reconstruction. Trees were constructed using these rDNA ITS sequences but they showed no consistency with the D3 or 18S trees or previously published trees. Thus the rDNA ITS region appears to be too polymorphic and fast evolving to be used as a reliable molecular marker for nematode phylogeny.

Chapter IV Survival, Phylogenetics and Biogeography tropical *Panagrolaimus* sp.

4.1 Introduction

The opportunity arose to study stress tolerance of another group of *Panagrolaimus* sp. after the data for the other nematodes used in Chapter 3 was already complete. This resulting dataset was analysed separately. In the previous collection, *Panagrolaimus* strains were from a wide range of biogeographically distinct regions includes temperate, continental, subarctic and Antarctic regions where the nematodes would be exposed to a variety of stresses. In the new collection, all of the nematodes are from more tropical regions that would rarely, if ever, experience freezing conditions. These nematodes were kindly donated from Dr.Marie-Anne Félix (Institut Jacques Monod). For the purposes of this chapter this new group of nematodes was named the Félix collection.

The strains in this collection were collected from China, India, the Indian Ocean islands and Cape Verde. These regions range from having dry, arid climate to a humid, monsoon climate. The strain JU765 was sampled at the edge of a rice paddy in Guangxi, China. Guangxi has a subtropical climate that is warm and wet. The strains JU1365, JU1366, JU1367 and JU1369 were collected from rotting plants and soil in the Indian state, Tamil Nadu. Tamil Nadu is situated in the southernmost part of the Indian peninsula and has a dry sub-humid to sub-arid climate. It is dependent on the monsoon rains and prone to drought when monsoons fail. The strain JU1361 was isolated from rotting fruit in the Periyar Natural Preserve in the Indian state, Kerala. Kerala is located along the southwest coast of the Indian peninsula to the west of Tamil Nadu. It has a humid, equatorial tropic climate. The strain JU1371 was also isolated in India. It was isolated from soil and leaf litter sampled in Pondicherry. Pondicherry has a similar climate to coastal Tamil Nadu, it has a warm to humid climate that is not as extreme as a desert climate. The strain JU1387 was isolated from rotting velvet apples sampled in the island La Réunion. This Indian Ocean Island is located east of Madagascar and southwest of Mauritius, it has a tropical and humid climate. The strains JU1645 and Ju1646 were both isolated from rotting leaves and

fruit from the Cape Verde islands, Santo Antão and Santiago respectively. Cape Verde has a dry, sandy, subtropical and semi-desert climate.

In this study of the Félix collection the nematodes were analysed for their ability to survive freezing and desiccation stress. Phylogenetic trees using a partial 18S sequence and the D3 region sequence of the 28S from these nematodes were constructed using Maximum Likelihood and Bayesian methods (as described in Sections 2.2.3.4.1 and 2.2.3.4.2 respectively). In the previous study (Chapter 3) there was a correlation between freezing and desiccation tolerance among the *Panagrolaimus* strains investigated, but not between their stress survival and phylogeny. The relationship between desiccation and freezing survival and between stress tolerance and phylogeny was also investigated for the Félix collection. The sequences from the Félix collection were also combined with those from the previous collection and phylogenetic trees constructed. The geographic locations where the nematodes used in this complete phylogeny were isolated are presented in Figure 4.1.

4.2 Results

4.2.1 Freezing tolerance and the effect of cold acclimation on freezing survival

The effect of freezing on the ten recently described Félix strains of *Panagrolaimus* was investigated (Figure 4.2). Nematode survival following thawing was estimated by microscopic observation. Statistical analysis using the ANOVA test suggested that significant differences existed between survival levels among the strains, but overall the initial freezing survival ability in this collection is low (max 23%). Acclimation at 10 ^{o}C for 10 days significantly improved survival (P<0.01) when compared to unacclimated nematodes for half of the strains: JU1369, JU1366, JU1367, JU1365, JU1361. The most significant difference is the 40% increase in freezing survival following acclimation for JU1369. Acclimation had no effect on the survival of JU1371, JU1387, JU1645 and JU1646 (P>0.05). The strain JU765 is sensitive to initial freezing, it can survive after a period of acclimation but the improvement is non-significant. All of the other strains do have some level of freeze tolerance that may be improved by acclimation but in comparison to a similar study in Chapter 3 the level is very low with most of the strains achieving less than 30% survival even with

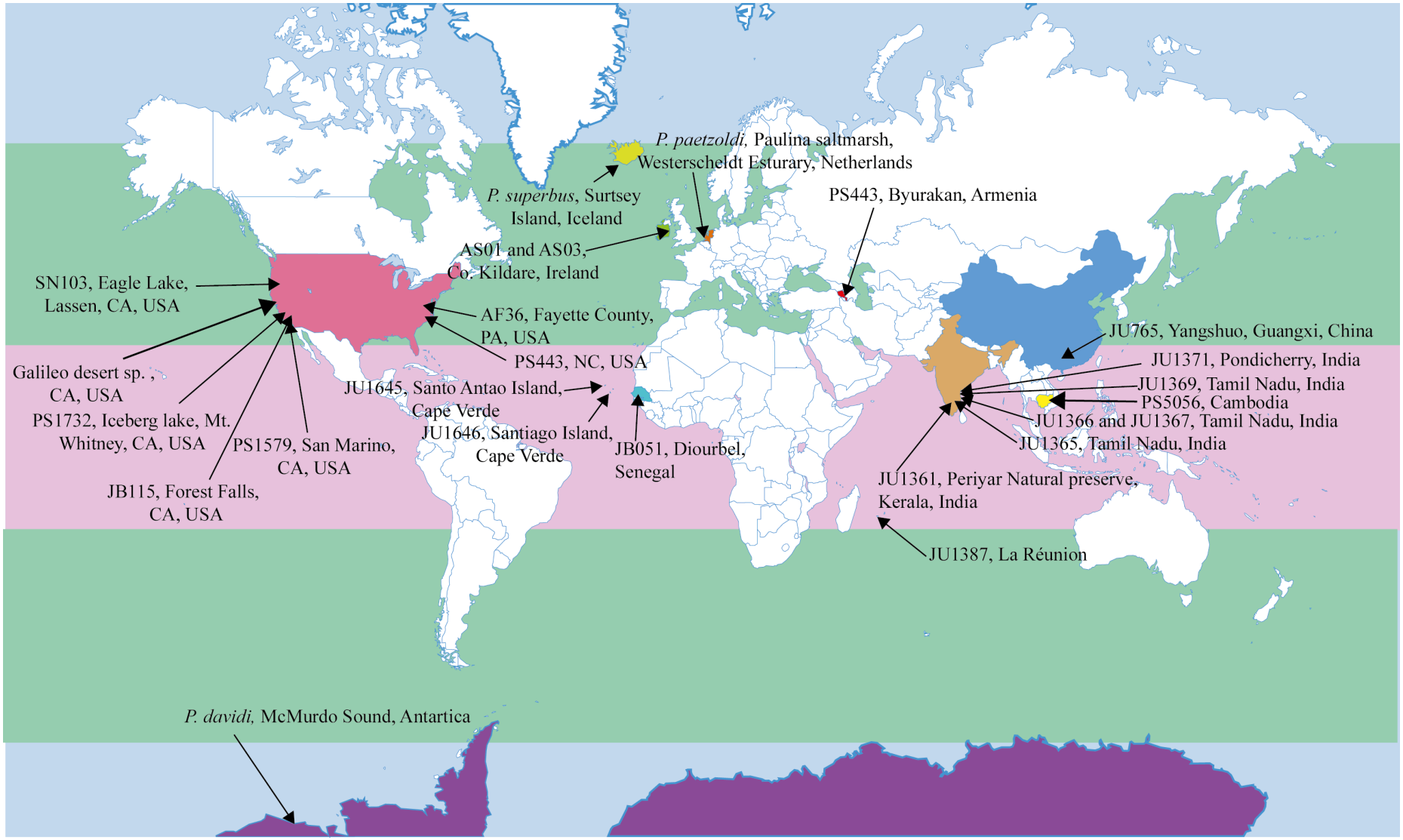


Figure 4.1. Maps of the world with the locations of all of the strains and species used in this study. The temperate latitudes on the globe are indicated in green and the tropical latitude is indicated in pink. The location of each strain within each country is approximate.

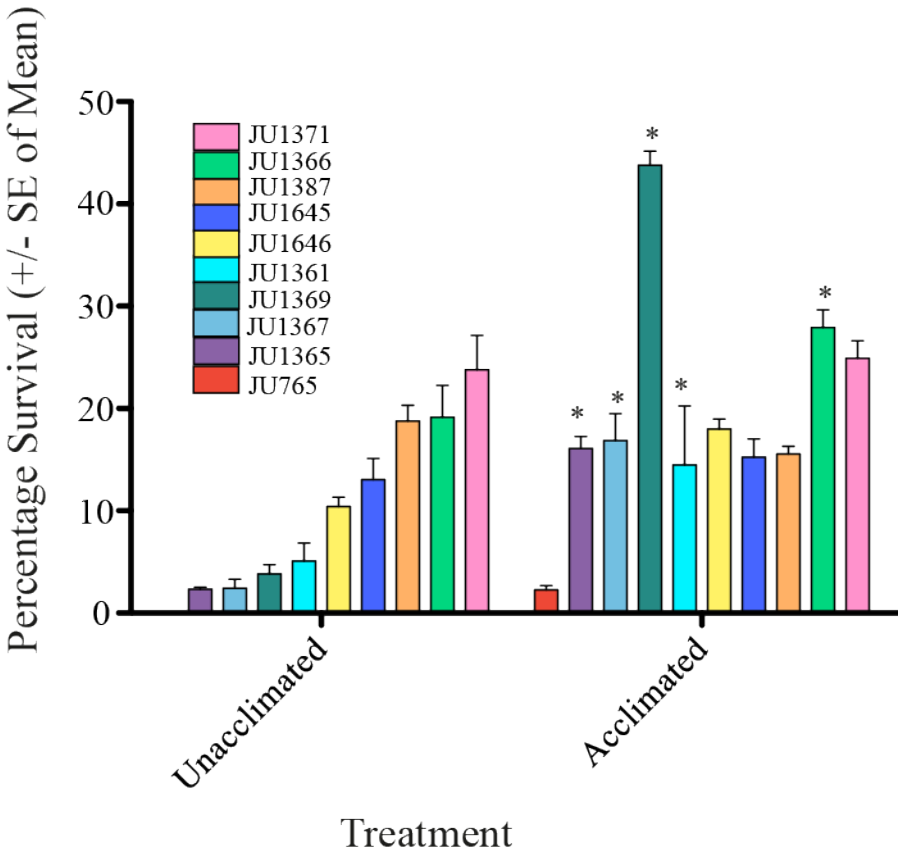


Figure 4.2. The effect of freezing treatment on the survival of ten tropical strains of *Panagrolaimus*. Unacclimated: Direct exposure to -80 ^{o}C for 24 h. Acclimated: Acclimation at 10 ^{o}C for 10 days followed by exposure to -80 ^{o}C for 24 h. The nematode samples are thawed for 24 h before survival was determined. Survival values are the mean survival value of four replicates +/- the standard error. Mixed stage populations were used for all strains. The strain JU765 showed 0% freezing survival in the unacclimated treatment. (* = P<0.05 in Bonferonni *post-hoc* test)

acclimation (seven of fourteen strains tested in Chapter 3 had post acclimation freezing survival values greater than 60%).

4.2.2 Effect of desiccation and preconditioning on survival

The ten *Panagrolaimus* strains in the Félix collection were preconditioned at 98% RH for various time periods (0-96 h) followed by exposure to freshly activated silica gel for 48 h. Nematode survival following rehydration for 24 h was estimated by microscopic observation (Figure 4.3). This collection of nematodes had minimal ability to survive direct exposure to silica gel with four of the eleven strains lacking the capacity to survive immediate exposure (0-11%). Preconditioning improves survival in the strains JU1361, JU1365, JU1366, JU1367, JU1369 and JU1387 and there is a significant correlation between time preconditioned and percentage survival ($P<0.05$). The minimum preconditioning time of 24 h at 98% RH was sufficient to significantly improve survival of the strains JU1366, JU1367, JU1369, JU1371, JU1645 and JU1646 ($P<0.01$). The strains JU1361, JU1365, JU1387 required 48 h at 98% RH to significantly increase desiccation survival. The strain JU765 is a "poor desiccator" and only has minimal non-significant survival after 96 h at 98% RH. With the exception of the desiccation sensitive JU765, all of the strains in this collection are excellent examples of "slow dehydration strategists" and have very high survival values (>75%) after a period of preconditioning.

4.2.3 The relationship between freezing tolerance and desiccation tolerance

The data for the freezing and desiccation survival clearly demonstrate that there is no clear link between freezing and desiccation survival in this collection (Figure 4.2 and Figure 4.3). The only strain that has any correlation is the freezing and desiccation sensitive strain JU765. The other strains are excellent slow dehydration strategists but do not have similar freezing survival even after acclimation (<20%). Regression analysis was carried out to confirm the lack of a relationship between these two stresses in this collection. The *P*-value of 0.1 and the R-squared value of 0.5492 are non-significant confirming that for these strains maximal freezing and desiccation tolerance are not related.

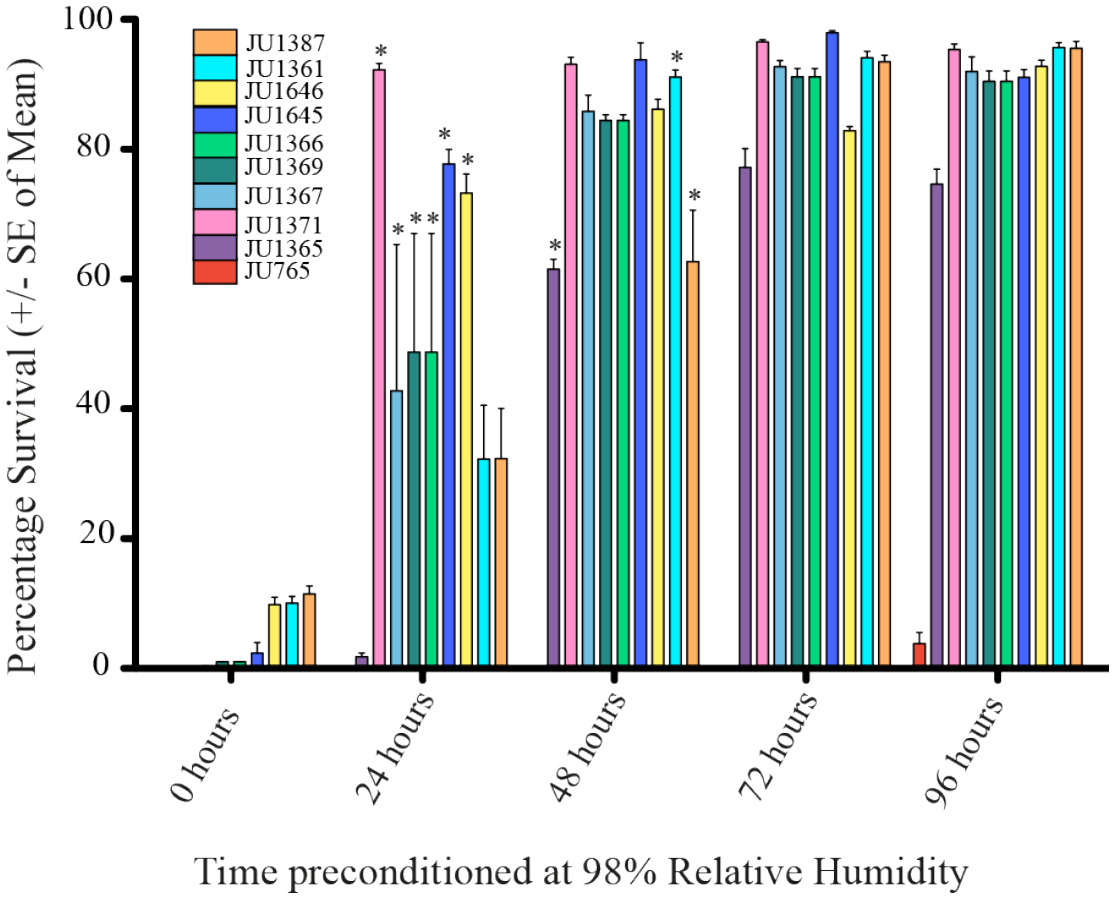


Figure 4.3. The effect of preconditioning time at 98% relative humidity (RH) on survival anhydrobiotic of ten tropical strains of *Panagrolaimus*. Nematodes per preconditioned at 98% RH at 20 ^{o}C for 0-96 h, desiccated for 48 h over freshly activated silica gel and hydrated in H_2O for 24 h before survival was determined. Survival values are the mean of three replicates +/- the standard error. Mixed stage populations were used for all strains. JU765 had 0% desiccation survival after 0, 24, 48 and 72 hour preconditioning treatment. JU1365, JU1371 and JU1367 had 0% desiccation survival after 0 h preconditioning treatment. (* = $P<0.05$ in Bonferonni *post-hoc* test)

4.2.4 Ribosomal RNA-sequencing

A 900 bp region from the 18S gene, together with D3 and ITS ribosomal regions were sequenced from these new *Panagrolaimus* strains. Dr. Marie-Anne Félix morphologically identified the strains as members of the genus *Panagrolaimus* but sequencing was required for confirmation. All of the sequences had similarity with *Panagrolaimus* nematodes and none had 100% identity to any sequence in Genbank, this confirmed that they have been correctly classified and are new strains. These sequences were used for subsequent phylogenetic analysis.

4.3.5 Construction of a phylogenetic tree based on 18S sequences

The partial 18S gene sequences used for strain classification were used to construct a phylogeny. The sequences were aligned using the MUSCLE alignment software and trimmed to equal lengths. A bootstrap consensus phylogeny was inferred using the Maximum Likelihood (ML) method with 1000 bootstrap replicates (Figure 4.4 (i)). The Bayesian method was also used for comparison (Figure 4.4 (ii)). The topologies recovered for both trees are the same with slight differences in branch lengths. The strains JU1365, JU1367 and JU1369 form a single clade that is divergent from the other nematodes with high support (92% ML, 94% Bayesian). The strains JU1645 and JU1646 group together with high support (74% ML, 92% Bayesian). The strains JU1361 and JU1366 are outside of the JU1645 and JU1646 group. JU1371 is outside from JU1361 and JU1366, with JU1387 outside of JU1371 and JU765 outside of JU1371 and JU1387. All of the main nodes are very well supported in both phylogenies (>70%).

4.3.6 Construction of a phylogenetic tree based on D3 sequences

The D3 region sequences were aligned using the MUSCLE alignment software and trimmed to equal length. Both ML and Bayesian methods were implemented. The two methods gave slightly different tree topologies (Figure 4.5). In both phylogenies the strain JU1365 is divergent from the other strains. In the two trees the order in the grouping of JU1367, JU1369 and JU1646 differs. ML groups JU1367 and JU1369 together with medium support (65%) with JU1646 closer to the other nematodes whereas Bayesian groups JU1369 and JU1646 together with high support and JU1367 placed outside, closer to JU765. There are also differences in the resolution of the group of nematodes comparing JU1366, JU1365, JU1361, JU1387 and JU1371. The

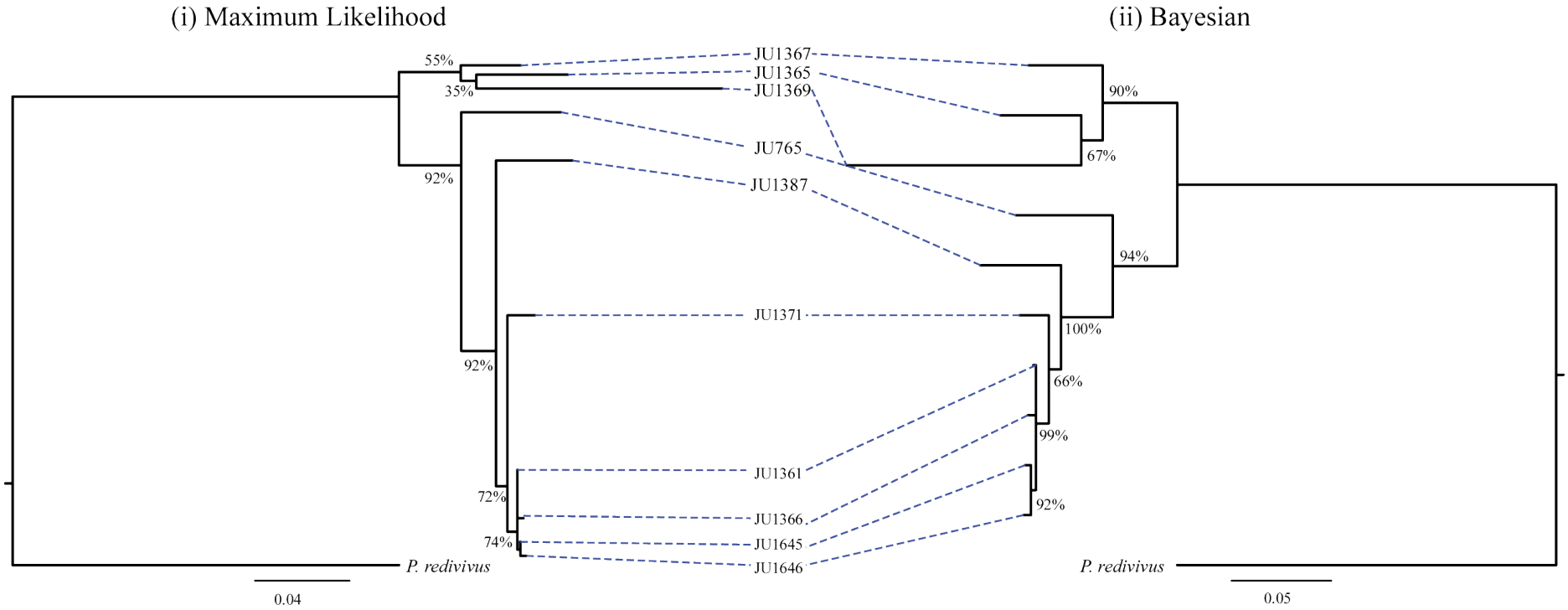


Figure 4.4. Phylogenetic tree constructed using 18S gene sequences. (i) The phylogenetic tree inferred using the Maximum likelihood method. (ii) The phylogenetic tree inferred using the Bayesian method. The scale bar on each tree indicates the number of substitutions per site. The dotted blue lines show a comparison between the topologies generated using each method.

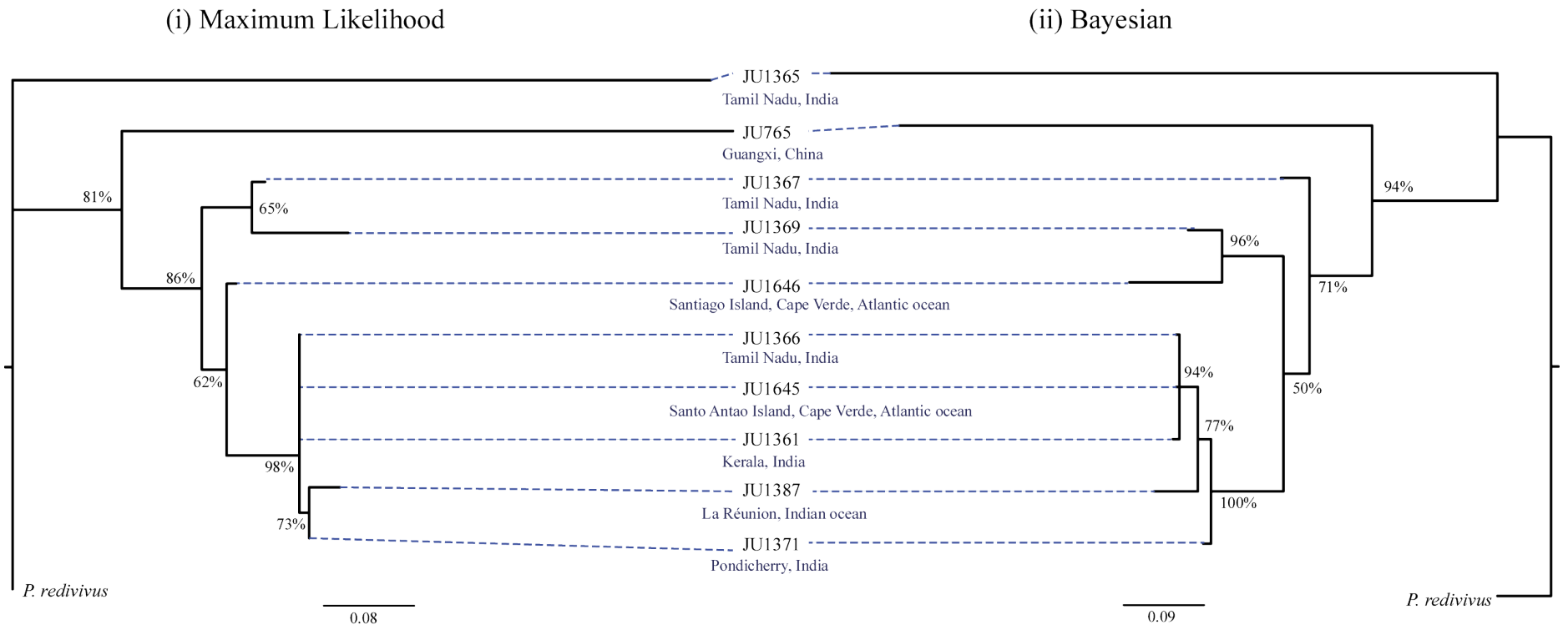


Figure 4.5 Phylogenetic trees constructed using sequences of the D3 region. (i) The phylogenetic tree inferred using the Maximum likelihood method. (ii) The phylogenetic tree inferred using the Bayesian method. The scale bar on each tree indicates the number of substitutions per site. The dotted blue lines show a comparison in the topologies generated using each method.

ML tree is collapsed apart from the grouping of JU1387 and JU1371 whereas the Bayesian tree has more resolution with the strains separated with higher support.

4.3.7 Construction of a phylogenetic tree based on concatenated 18S and D3 sequences.

The sequence alignment of the 18S gene and the D3 region were concatenated and used to construct a third phylogenetic tree (Figure 4.6), using both ML and Bayesian methods. In this phylogeny both methods gave the same topology. The strains JU1367, JU1365 and JU1369 form a clade with high support (71%). There is a second clade formed with JU1361, JU1366, JU1371 and JU1387. Within this clade JU1371 and JU1387 group together with low support in the ML tree (48%) but high support in the Bayesian tree (100%). The strain JU1645 is placed outside of this clade with medium to high support (67% ML, 100% Bayesian). JU1646 is placed outside from JU1645 with high support (97% ML, 100% Bayesian) and JU765 outside from it (92% ML, 100% Bayesian).

4.3.8 Construction of a phylogenetic tree with all available *Panagrolaimus* sequence.

The *Panagrolaimus* 18S gene and D3 region sequences described in Chapter 3 were combined with the sequences from the Félix collection, aligned, trimmed and concatenated. This concatenated alignment was used to construct a ML and Bayesian phylogenetic tree (Figure 4.7). Since the freezing and desiccation survival data suggested that these nematodes are excellent slow dehydration strategists but have quite poor at freezing tolerance, it was of particular interest to see if these strains may be phylogenetically divergent from the previous collection of strains and species. The inclusion of these new sequences into this tree did change the topology slightly from that seen in Figure 3.8 Chapter 3, but the highly supported relationships remained the same. The concatenated 18S gene and D3 region phylogenetic tree in Figure 4.7 clearly demonstrates that the Félix collection with the exception of JU765 are clearly quite closely related but divergent from the previously analysed nematodes. JU1369, JU1367 and JU1365 (all from Tamil Nadu, India) form a group with only AS01, which has a similar freezing and desiccation survival phenotype. JU1646, JU1371, JU1387, JU1645, JU1361 and JU1366 form a second group. The freezing and desiccation sensitive nematode JU765 is found amongst the previously studied strains

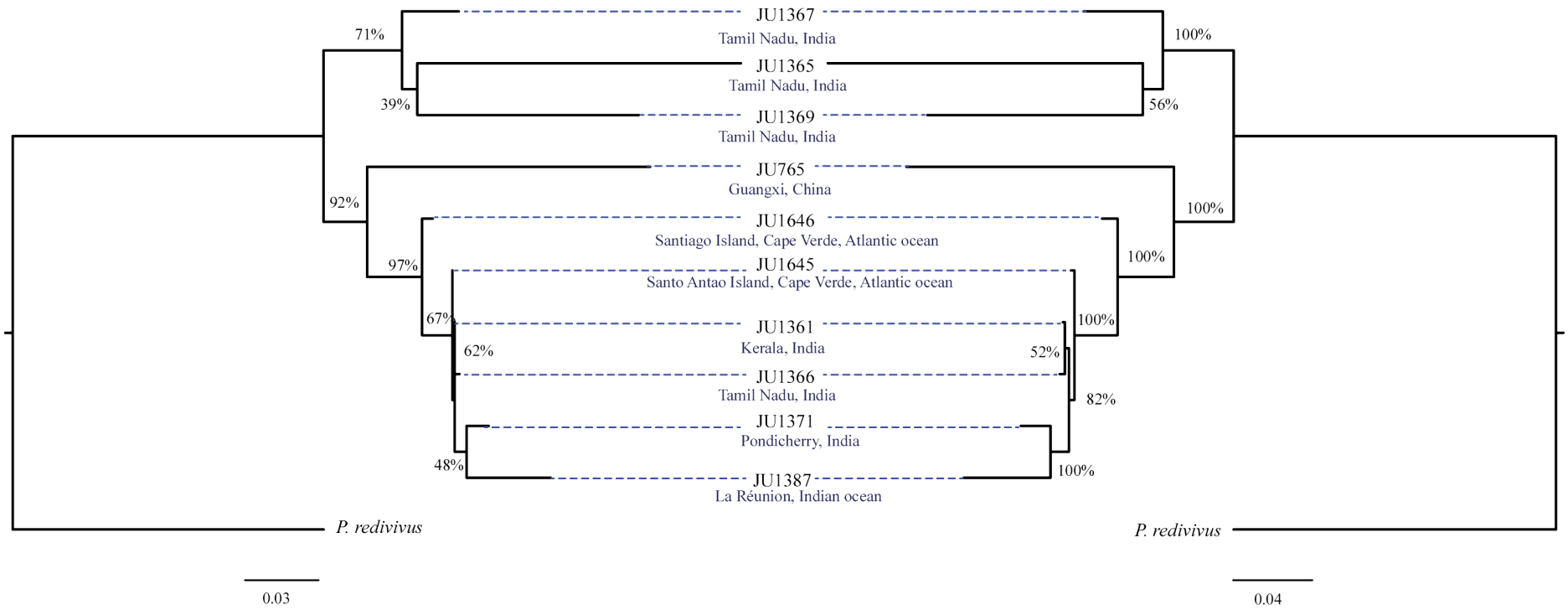


Figure 4.6. Phylogenetic tree constructed using concatenated 18S gene and D3 region sequences. (i) The phylogenetic tree inferred using the Maximum likelihood method. (ii) The phylogenetic tree inferred using the Bayesian method. The scale bar on each tree indicates the number of substitutions per site. The dotted blue lines show a comparison between the topologies generated using each method.

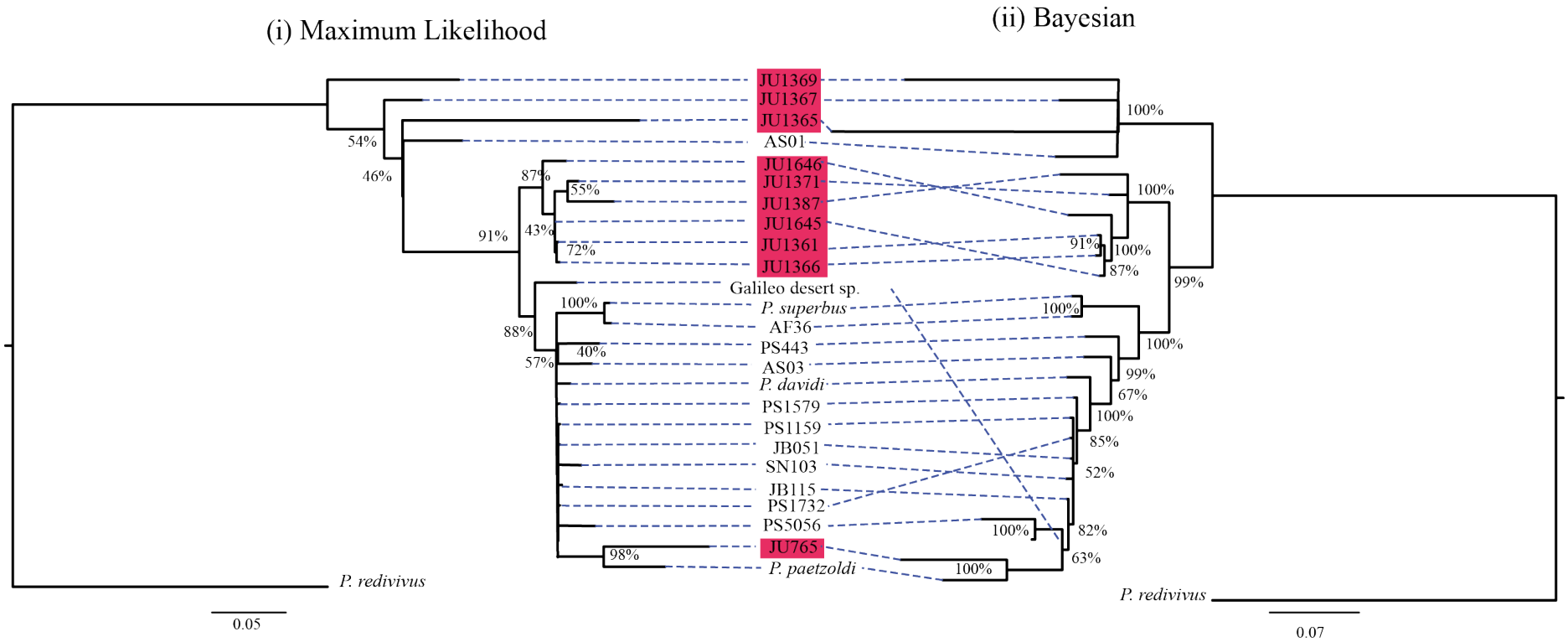


Figure 4.7. Phylogenetic tree constructed using entire collection of concatenated 18S gene and D3 region sequences. (i) The phylogenetic tree inferred using the Maximum likelihood method. (ii) The phylogenetic tree inferred using the Bayesian method. The scale bar on each tree indicates the number of substitutions per site. The dotted blue lines show a comparison between the topologies generated using each

grouping with *P. paetzoldi*.

4.3 Discussion

The Félix collection of nematode strains were isolated from tropical regions of the world whereas the collection analysed in Chapter 3 were mainly from temperate regions (Figure 4.1). A connection between desiccation and cold tolerance has been identified in several species, including nematodes (Wharton and Barclay, 1993; Chen *et al.*, 2005; Adhikari *et al.*, 2009, 2010) and insects. In the temperate nematodes studied in Chapter 3 a correlation was found between those that tolerated freezing and desiccation. The Félix collection of nematodes are from regions that experience conditions ranging from very humid to very dry climates, but they would not typically experience low temperatures. They were also assessed to see whether they have desiccation tolerance that may be required to thrive in these regions and to establish whether anhydrobiotic phenotype gave cross tolerance to freezing.

All of the strains except JU765 showed some level of freezing tolerance when exposed directly to freezing temperatures, so they could described as freezing survivors. However, since the maximum percentage survival was only 23% they are generally poor at freezing survival as compared with the *Panagrolaimus* strains from temperate and cold regions. Many species including nematodes show an improvement their cold tolerance after a period of acclimation (Wharton *et al.*, 2000; Jagdale and Grewal, 2003; Smith *et al.*, 2008; Hayashi and Wharton, 2011). In the nematodes studied in Chapter 3 acclimation improved the survival ability of eleven of the fifteen strains. In the Félix collection acclimation significantly improved the freezing survival of only half of the strains. Acclimation had the greatest effect on the strain JU1369 (from Tamil Nadu, India), increasing survival from 4% to 44%. In the other strains the improvement in freezing tolerance following acclimation was minimal, survival for the majority of the strains was still less than 20% survival. These results show that these nematodes have a limited ability to survive freezing that may be slightly improved by acclimation. As these nematodes are from tropical regions they may have not been required to evolve a cold tolerance strategy. Alternatively the lack evolutionary pressure may have resulted in the loss of genes that are required for cold tolerance or have lost a previous ancestral strategy.

Nematodes are essentially aquatic animals that need to be fully hydrated and covered in a water film for normal activity. The ability to survive conditions of extreme desiccation, known as anhydrobiosis is essential for them to flourish in environments prone to dehydration. The Félix collection nematodes were assessed for their ability to undergo anhydrobiosis to survive desiccation stress. Although these nematodes were isolated from tropical regions where they may be prone to variable levels of desiccation stress. All of the strains, except JU765 were slow dehydration strategists *sensu* Womersley (1987). They have no or very low survival when exposed to extreme desiccation unless they first have a period of preconditioning to moderate reductions in RH (Womersley, 1987). This period of acclimation can allow modulation of metabolic and biochemical processes essential for the successful induction of anhydrobiosis (Burnell and Tunnacliffe, 2011). The freezing and desiccation sensitive strain JU765 was sampled at the edge of a rice paddy in warm and wet region of China. This nematode may not be exposed to prolonged periods of either desiccation or freezing stress so has probably not needed to evolve such adaptations. The strains JU1365, JU1366, JU1367 and JU1369 were all located in the same Indian state, Tamil Nadu. Tamil Nadu has a dry climate that is prone to drought if the monsoon rains fail. After varying periods of preconditioning these strains have high levels of survival to desiccation stress (75-92% after 96 h at 98% RH). The strain JU1371 isolated from the nearby Pondicherry is sensitive to direct exposure to silica gel but after only 24 h of preconditioning at 98% RH it has over 90% desiccation survival. This strain must have very efficient mechanisms for up-regulation of the genes and gene-products required for desiccation adaptations. The strains JU1361 and JU1387 are from different regions (Kerala, India and La Réunion, Indian ocean) but both experience similar humid and tropical climates and have similar levels of survival after preconditioning. The strains JU1645 and JU1646 are from the dry, subtropical, sub-desert conditions of Cape Verde. It may be expected that these strains would have be fast desiccation strategists to survive this environment but they only have low ability to survive direct exposure to silica gel. They required a period of 48 h at 98% RH to achieve significant levels of desiccation survival. Therefore they may experience a microclimate that has reducing levels of humidity that allows them to slowly desiccate in response to dehydration.

Freezing and desiccation survival in the Félix collection of *Panagrolaimus* nematode strains are not correlated. These nematodes can enter anhydrobiosis after a period of preconditioning, but have low freezing survival that is only slightly improved by acclimation. Many physiological and molecular responses to cold may have originally been adaptations for desiccation stress (Danks, 2000). Nematodes have been shown to have cross-tolerance to freezing and desiccation stress and can survive freezing by cryoprotective dehydration (Wharton et al., 2003). A period of mild desiccation may also improve freezing survival (Chen *et al.*, 2005; Hayward *et al.*, 2007; Adhikari *et al.*, 2010). It appears that in this collection that although this may be true the desiccation stress response has not been evolved for use in the freezing response. It is also possible that since these nematodes live in tropical regions that would rarely experience freezing temperatures that they may have lost this cross-tolerance between cold and desiccation stress.

The ten strains were all confirmed by sequencing to be from the genus *Panagrolaimus*. Phylogenetic trees were constructed to investigate whether there was any relationship between stress tolerance and phylogeny or biogeography/climate and phylogeny. In the phylogenies of the Félix collection nematodes there are some loose relationships between phylogeny, stress tolerance and biogeography. In the concatenated 18S gene and D3 region phylogeny with the Félix nematodes the strains JU1365, JU1367 and JU1369 group closely together. These nematodes are all desiccation tolerant and from the same region of India (Tamil Nadu), showing a link between the phylogeny and the biogeography. Then the strains JU1366 and JU1371 are from similar regions, Tamil Nadu and Pondicherry but are distant from these nematodes in the phylogeny. However, these are closely related and have similar desiccation tolerance to the other Indian strain JU1361 (Kerala) and the La Réunion strain JU1361. If there was a clear relationship between biogeography and phylogenetic relationship, the nematodes JU1645 and JU1646 (Cape Verde Islands, Atlantic ocean) should be phylogenetically divergent from the Indian strains. This does not occur: these strains have similar stress tolerances to the other strains and are positioned within the main clade. The one completely freezing and desiccation sensitive strain JU765 is also slightly confusing as it is not divergent from the other nematodes and is placed between the two main clades in the concatenated tree. This

positioning of a sensitive strain amongst the tolerant strains also occurred in the phylogenies in Chapter 3 with PS1732 and *P. paetzoldi*.

The strains in the Félix collection are different to the previous collection in that these strains do not show a correlation between freezing and desiccation tolerance: they are tolerant to desiccation after a period of preconditioning, but have weak freezing tolerance which is typically only slightly improved by acclimation. As these stress responses were found to be correlated in the more temperate *Panagrolaimus* nematodes in Chapter 3 it was interesting to see where the Félix strains would position in a large phylogeny of all the strains. This 18S gene and D3 phylogeny (Figure 4.6) showed with high support that JU1369, JU1367 and JU1365 form a group with the divergent strain AS01 while JU1646, JU1371, JU1387, JU1646, JU1645, JU1361 and JU1366 formed another separate group. The freezing and desiccation sensitive strain JU765 is different to the other nematodes and groups *P. paetzoldi* in the main clade that contains most of the *Panagrolaimus* nematodes. The strain AS01 (Leixlip, Kildare) that was placed among the Félix strains (Figure 4.7) and has similar levels of slow desiccation survival and low maximal freezing survival to the Félix strains. Overall the phylogenies of the Félix collection or the combined phylogeny of all the Panagrolaimus strains in this study show that there tends to be a relationship between high desiccation tolerance and phylogeny. However there are exceptions in that some closely related strains do not show the same stress tolerance phenotypes. This raised the possibility that these stress phenotypes may be dependent on relatively small number of genes, possibly regulatory genes for their expression.

Chapter V Molecular clock analysis on the genus *Panagrolaimus*

5.1 Introduction

Panagrolaimus species have been isolated from several sites in the Antarctic including continental nunataks (Sohlenius *et al.*, 1996; Swart and Harris, 1996). These nematodes are particularly abundant in the bacteria rich soil associated with bird nesting sites in the Antarctic, *P. magnivulvatus* was isolated from material taken from inside snow petrel nests (Swart and Harris, 1996). Two species of *Panagrolaimus* have also been isolated from a gull's nest on a nunatak in Surtsey (Boström, 1998), an Icelandic island formed during 1963 to 1967 from volcanic eruptions (Baldursson and Ingadóttir, 2007). In addition to *P. superbus* from Surtsey and *P. davidi* from Antarctica, other isolates in this study are from high altitude locations from temperate, sub-tropical or continental soils, from roof moss or roof gutters (Table 2.1). The phylogenetic tree presented in Figure 3.8 shows that *P. davidi* from Antarctica is most closely related to *Panagrolaimus* sp. PS1579 which was isolated in Huntington Botanical Gardens, San Mario, California, but *P. davidi* is also closely related to strains of *Panagrolaimus* isolated in other regions of North America, Senegal Africa, the Netherlands and Ireland. *P. superbus* from Surtsey and *Panagrolaimus* sp. AF36 from Pennsylvania, USA form a distinct clade that is clearly separated from the clade containing *P. davidi*. Thus overall there seems to be little correlation between biogeography and phylogeny among these closely related strains of *Panagrolaimus.* *P. davidi* and *P. superbus*, together with their sister species and strains from temperate regions are both anhydrobiotic and cryotolerant. It is apparent that *P. superbus* did not evolve on Surtsey Island but was transported there after the volcanic origin of the island. Similarly it may be possible that *P. davidi* is not endemic to Antarctica but may have been transported there in an anhydrobiotic state by snow petrels or other Antarctic birds. In this chapter I estimate the divergence times for *P. davidi* and *P. superbus* and their sister species *Panagrolaimus* sp. PS1579 *Panagrolaimus* sp. AF36, respectively, using the relaxed molecular clock approach. Data on the estimated age of origin of *P. davidi* may provide an insight into its time of dispersal to Antarctica.

Antarctica was derived from the breakup of the super-continent Gondwana and it has been glaciated since the Eocene (Tripati et al., 2005). Models of the Antarctic ice sheet Last Glacial Maximum (~20,000 years ago) leave no ice free refuges for most terrestrial biota (Huybrechts, 2002) which would imply that contemporary Antarctic biota have colonized the region following the Last Glacial Maximum. The paucity of contemporary Antarctic biota could then be attributed to difficulty of recolonization because of the harsh environment, and the geographic barriers posed by the Southern Circumpolar Ocean and Current, the prevailing Circumpolar westerly winds and the cold winds radiating from the centre to the edges of the continent. Terrestrial communities in Antarctica consist of bryophytes and lichens, cyanophyta and micro-invertebrates (nematodes, tardigrades, rotifers, mites, collembola) and protozoans (Convey *et al.,* 2008). But, surprisingly, the majority of these taxa (with the exception of the bryophytes) show a high degree of endemism which points to the possibility that these endemic taxa are pre-Pleistocene relics (Convey *et al.,* 2008). All currently described Antarctic nematode species are considered to be endemic to the region (Andrássy, 1998; Maslen and Convey, 2006).

In order to estimate divergence times within a given phylogeny, calibration points within the phylogeny are required. Thus the absence of a fossil record has made it difficult to estimate the divergence times of nematode taxa. Nevertheless molecular clock methods have been used to estimate the divergence times of nematode taxa, utilising various strategies: a strict molecular clock, with the molecular clock rate for nematode globin and cytochrome being extrapolated from metazoan (predominantly chordate and arthropod) phylogenies (Vanfleteren et al., 1994); using a single calibration point the time of divergence of nematodes from arthropods (Coughlan and Wolfe, 2002; Stein *et al.*, 2003) and for the divergence time of *Caenorhabditis elegans*/*C. briggsae* an estimate based on the neutral mutation rate and an estimated 60 day average generation time (Cutter, 2008).

In this chapter the relaxed molecular clock approach is used to estimate the divergence times of five strains and species of *Panagrolaimus* using DNA sequences for the 18S rRNA gene and the D3 region of the 28S rRNA gene. This analysis was carried out using two different molecular clock models: the autocorrelated CIR model (Lepage et al., 2007) and the uncorrelated Ugamma model (Drummond et al., 2006).

See Chapter 1 Section 1.2.2.2 for further discussion on molecular clock calibrations. This analysis included the nematode species, which had been used in the previous molecular dating experiment by Vanfleteren *et al.* (1994) and two species of *Halicephalobus*. *Halicephalobus* is a member of the superfamily Panagrolaimidae and these two species were included in the study because *H. mephisto* which can tolerate high temperatures. It is a newly described species which was recently isolated from fracture water in a mine borehole (Borgonie et al., 2011). As calibrator species to date the tree rRNA sequences from selected pairs of arthropods whose minimum and (soft) maximum evolutionary divergence dates have been published by Benton *et al.* (2009) were used. Thus far there are no reports in the literature on the use of relaxed molecular clock methodology combined with minimum and maximum fossil-based dates to estimate nematode divergence times.

5.2 Results

5.2.1 Alignment of sequences used in the molecular clock

The 18S gene and D3 regions of five *Panagrolaimus* strains and species were sequenced as described in Chapter 3 and used for molecular clock analysis. The 18S gene and D3 sequences from the other nematodes, arthropods and annelids were obtained from GenBank. The accession numbers for the relevant sequences are shown Table 5.1. In certain cases the database contains the sequence for the entire region that spans the SSU and LSU (same accession numbers in Table 5.1). In these cases the 18S and D3 regions were manually isolated from the larger sequence by aligning to a known sequence. The D3 and 18S sequences were aligned using the MUSCLE alignment software (Edgar, 2004). The 18S alignment was further improved by performing a structural alignment with the program RNAsalsa, (Stocsits et al., 2009a), using *S. cerevisiae* as a reference, as described in Section 2.2.3.4. The RNAsalsa structural alignment resulted in an improved alignment with considerably fewer gaps.

5.2.2 Phylogenetic tree of obtained using 18S and 28S D3 region rDNA sequences

The 18S structural alignments and the D3 alignments were trimmed and concatenated together into a single alignment file. This alignment was used to infer ML phylogeny

Table 5.1. GenBank accession numbers of sequences used in this work.

Organism	D3 region accession number	18S region accession number
Halicephalobus mephisto	GU811759	GQ918144
Halicephalobus gingivalis	AF202156	DQ145637
Caenorhabditis elegans	X03680	X03680
Caenorhabditis briggsae	AY604481	FJ380929
Nippostrongylus brasiliensis	AM039748	AJ920356
Trichostrongylus colubriformis	AM039743	AJ920350
Ascaris suum	FJ418792	U94367
Pseudoterranova decipiens	AY821763	U94380
Trichodorus primitivus	AM180729	AJ439517
Nasonia vitripennis	GQ374784	GQ410677
Apis mellifera	AY703551	AB126807
Chironomus tentans	X99212	X99212
Aedes albopictus	L22060	X57172
Musca domestica	AJ551427	DQ133074
Drosophila melanogaster	M21017 M29800	M21017 M29800
Daphnia magna	AF532883	AM490278
Artemia salina	AF169697	X01723
Glycera dibranchiata	AY995208.1	AY995207
Eisenia fetida	X79872	AF212166

with the MEGA 5.0 software suite. The expected topology was obtained (Figure 5.1). The *Panagrolaimus* nematodes form a single clade (blue), *H. gingivalis* and *H. mephisto*, which are also members of the superfamily Panagrolaimidae are grouped outside the *Panagrolaimus* species. All of the other nematodes are grouped outside of the Panagrolaimidae clade. The arthropods also form a single clade (green). The Diptera (*Musca domestica, Drosophila melanogaster, Chironomus tentans* and *Aedes albopictus*) form a clade, with the Hymenoptera (*Nasonia vitripennis* and *Apis mellifera*) forming a separate group. The crustaceans (*Daphnia magna* and *Artemia salina*) from a distinct clade outside of all of the insects. The bootstrap values are high for all nodes (< 81%), so the topology for this phylogeny has high support. The effect of using a structural alignment of the 18S rDNA sequences can be seen in the bootstrap values. The values for the non-structural alignments are indicated in brackets in Figure 5.1. The structural alignment generally yielded higher bootstrap values for the majority of the nodes.

5.2.3 Molecular clock analysis

The concatenated 18S (structural) and D3 sequence alignment (Section 5.2.1) and its corresponding ML phylogenetic tree (Section 5.2.2), a calibration file (Table 5.2) and outgroup file were used as an input for the molecular clock analysis with the program PhyloBayes 3.0 (Lartillot et al., 2009). The maximum and minimum node dates to calibrate the clock are presented in Table 5.2. These data were obtained from (Benton et al., 2009).

5.2.4 Final molecular timescale

In this study two models for the rate of evolution were tested, the autocorrelated model CIR and the uncorrelated model Ugamma. The effects of root age prior (Section 5.2.6.) and of soft-bound priors (Section 5.2.7) on molecular clock were also investigated. Based on these tests the molecular divergence time estimated under the autocorrelated CIR model was selected using soft bounds (5% default in PhyloBayes) for the calibration points and a prior root age of 550 MYA with a standard deviation of 50 was set. The resulting chrono-phylogeny is illustrated in Figure 5.2. The analysis was replicated 20 times and the mean calculated for each node. The divergence times for each individual node under the CIR model are indicated in Table 5.3. The estimated divergence dates for the nodes with calibration points were

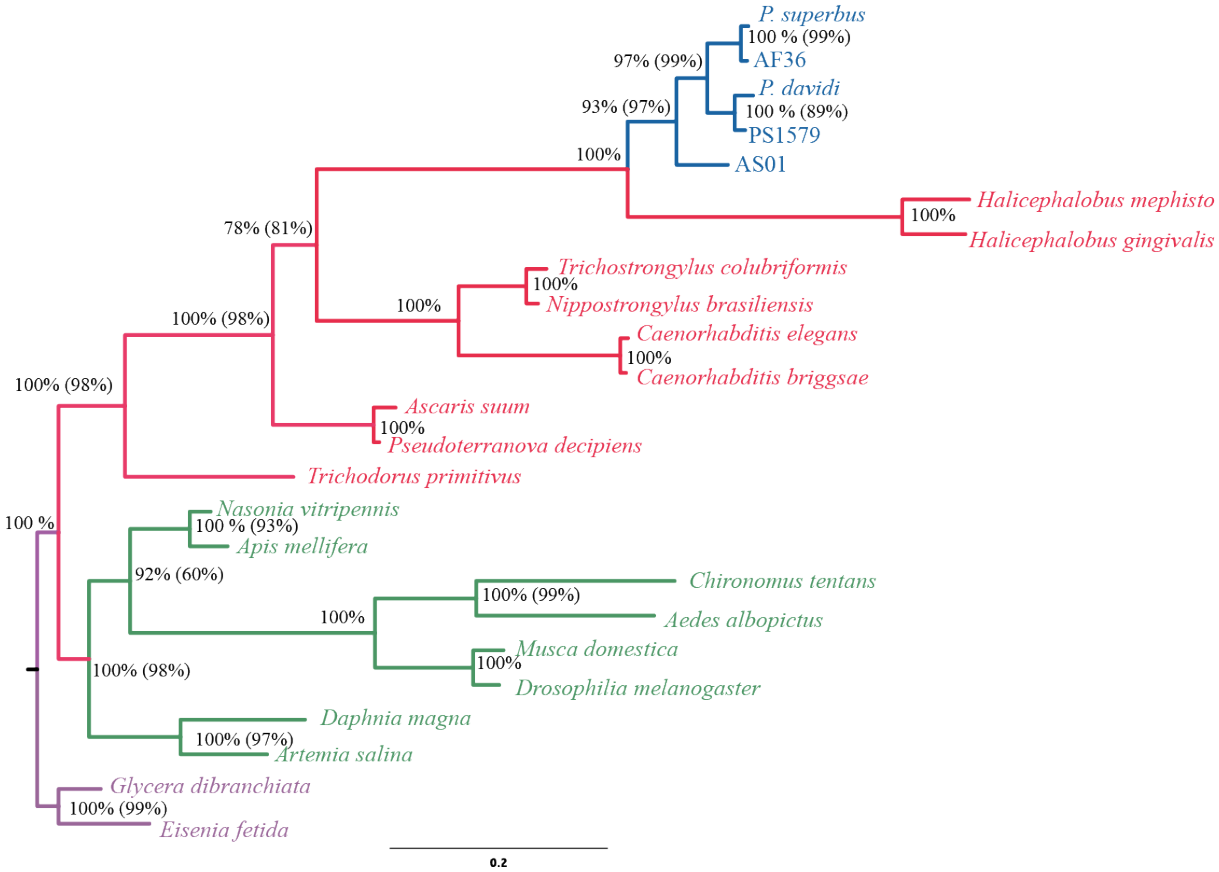


Figure 5.1. Maximum Likelihood phylogenetic tree constructed using concatenated structurally aligned 18S rDNA and 28S rDNA D3 sequences aligned using the MUSCLE program. Bootstrap values are indicated at each node. The bootstrap values in brackets were obtained when a non-structural alignment of the 18S sequences was used to infer ML phylogeny. Two annelid species *Glycera dibranchiata* and *Eisenia fetida* were used as the outgroup. The scale bar on the tree indicates the number of substitutions per site.

Table 5.2. The minimum constrains and maximum date estimates used to calibrate the molecular clock (obtained from Benton *et al.*, 2009).

Node	**Maximum node age (MYA)**	**Minimum node age (MYA)**
Nematode and Arthropod (*Caenorhabditis elegans* and *Artemia salina*)	581	520.5
Crustaceans and Insects (*Daphnia magna* and *Nasonia vitripennis*)	543	510
Hymenoptera and Diptera (*Apis mellifera* and *Aedes albopictus*)	307.2	238.5
Within Diptera (*Drosophila melanogaster* and *Chironomus tentans*)	295.4	238.5
Within Hymenoptera (*Apis mellifera* and *Nasonia vitripennis*)	243	152

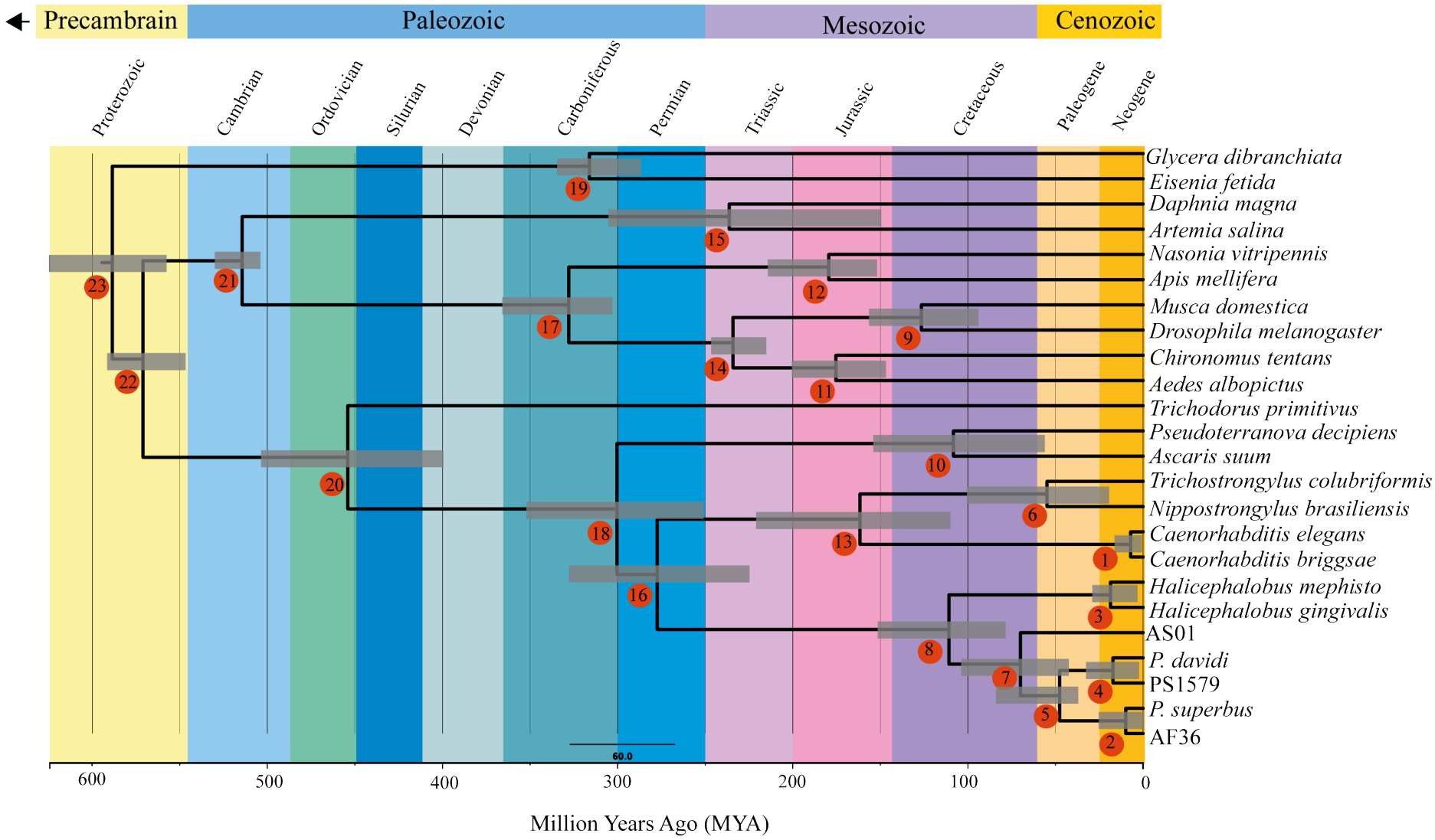


Figure 5.2. Molecular divergence time estimated under the autocorrelated CIR model and using soft bounds (5%) default in Phylobayes) for the calibration points. The numbers at each node correspond to the specific ages for each node as listed under CIR model in Table 5.3

Table 5.3. Divergence ranges and time period for each node in molecular clock.

Node		Divergence time	Era	Period
1	*C. elegans* and *C. briggsae*	2.58-18.38	Cenozoic	Neogene
2	*P. superbus* and AF36	3.544-22.55	Cenozoic	Neogene
3	*H. mephisto* and *H. gingivalis*	10.58-31.55	Cenozoic	Neogene-Paleogene
4	*P. davidi* and PS1579	7.315-35.19	Cenozoic	Neogene-Paleogene
5	*P. superbus*/AF36 and *P. davidi*/PS1579	25.34-80.52	Cenozoic-Mesozoic	Paleogene-Cretaceous
6	*T. colubriformis* and *N. brasiliensis*	22.15-101.9	Cenozoic-Mesozoic	Neogene- Cretaceous
7	*P. superbus*/AF36/*P. davidi*/PS1579 and AS01	41.16-108.5	Cenozoic-Mesozoic	Paleogene-Cretaceous
8	*P. superbus*/AF36/*P. davidi*/PS1579/AS01 and *H. mephisto*/*H. gingivalis*	77.3-150.5	Mesozoic	Cretaceous-Jurassic
9	*M. domestica* and *D. melanogaster*	92.21-154.5	Mesozoic	Cretaceous-Jurassic
10	*P. decipiens* and *A. suum*	60.54-157.1	Cenozoic-Mesozoic	Paleogene- Jurassic
11	*C. tentans* and *A. albopictus*	148.3-200	Mesozoic	Jurassic
12	*N. vitripennis* and *A. mellifera*	151.4-217.1	Mesozoic	Jurassic-Triassic
13	*C. elegans*/*C. briggsae* and *T. colubriformis* /*N. brasiliensis*	113.1-217.4	Mesozoic	Cretaceous-Triassic
14	*M. domestica* /*D. melanogaster* and *C. tentans*/*A. albopictus*	219.6-242.3	Mesozoic	Triassic
15	*D. magna* and *A. salina*	150.5-311.2	Mesozoic-Palaeozoic	Jurassic-Carboniferous
16	*C. elegans*/*C. briggsae*/*T. colubriformis* /*N. brasiliensis* and Panagrolaimidae	224.4-330.4	Mesozoic-Palaeozoic	Triassic- Carboniferous
17	Hymenoptera and Diptera	308.8-350.7	Palaeozoic	Carboniferous
18	*C. elegans*/*C. briggsae*/*T. colubriformis* /*N. brasiliensis*/Panagrolaimidae and Ascaridida	249.7-351.5	Mesozoic-Palaeozoic	Triassic- Carboniferous
19	*G. deibranchtata* and *E. fetida*	254.7-390.4	Palaeozoic	Permian-Devonian
20	*C. elegans*/*C. briggsae*/*T. colubriformis* /*N. brasiliensis*/*Panagrolaimidae*/*Ascaridida* and *T. primitivus*	400.1-504.8	Palaeozoic	Devonian-Cambrian
21	Crustaceans and Insects	508.1-528.3	Palaeozoic	Cambrian
22	Arthropods and Nematodes	548.1-583.4	Neoproterozoic	Ediacaran
23	Annelids and Arthropods/Nematodes	559.3-624	Neoproterozoic	Ediacaran

generally within the range specified by these points. The estimated divergence time of 570.8 MYA (548.1-583.4) for the arthropods and nematodes lies within the 520.5-581 MYA used to calibrate this point. Between the crustaceans and the insects the estimated divergence time of 514.5 MYA (508.1-528.3) lies within the calibration date (510-543). The estimated divergence of the Hymenoptera and the Diptera is 327.9 MYA (308.8-350.7) so it is slightly older than the maximum calibration date (238.5-307.2) provided by Benton *et al.* (2009). However Wiegmann *et al.* (2009) present phylogenetic evidence indicating that the Hymenoptera are the earliest branching holometabolan lineage, with a divergence time from the other holometabolous insects (including the Diptera) of ca. 350 MYA, a date which is considerably older than existing fossil estimates. Within the Diptera the estimate of divergence is 243 MYA (219.6-242.3) lying within the calibration dates of 238.5-295.4 MYA. The estimated divergence of *Nasonia* and *Apis* within the Hymenoptera was 178.7 MYA (151.4-217.1) this also sits within the calibration dates of 152-243 MYA.

Molecular divergence estimates for the nematodes from the arthropods and annelids were Ediacaran at 570.8 MYA (548.1-583.4). Within the nematodes the time of divergence of the plant parasitic Trichodoridae from the remainder of the nematodes was Ordovician at 454.4 MYA (400.1-504.8). The order Ascaridida diverged from the order Rhabditida during the Permian at 299.4 MYA (249.7-351.5). Within the order Rhabditida the divergence of the Panagrolaimidae occurred in the Permian at 275.7 MYA (224.4-330.4). The divergence of *Panagrolaimus* sp. from *Halicephalobus* sp. was in the Lower Cretaceous at 109.8 MYA (77.3-150.5). Within the *Panagrolaimus* sp. the divergence of AS01 from the remainder of the strains is occurred in the Upper Cretaceous 70.12 MYA (41.16-108.5). The *P. davidi* and its sister species PS1579 diverged from *P. superbus* and its sister species AF36 the Eocene epoch at 47.77 MYA (25.34-80.52). The divergence between *P. davidi* and PS1579 occurred in the Miocene at 17.18 MYA (7.315-35.19). The divergence between *P. superbus* and AF36 occurred in the late Miocene at 9.969 MYA (3.544-22.55). Other noticeable dates are the late Miocene divergence time of 7.173 MYA (2.58-18.38) between *C. elegans* and *C. briggsae* and the early Miocene divergence time of 18.34 MYA (10.58-31.58) between *H. mephisto* and *H. gingivalis.*

5.2.5 Effect of rate of evolution model on molecular clock

In this study two models for the rate of evolution were tested, the autocorrelated model CIR and the uncorrelated model Ugamma. The analysis was replicated 20 times. There was no significant difference for each node between the 20 replicates. The mean minimum, maximum and optimal ages for each node are presented in Table 5.4. The optimal age between each model is very significantly different (ANOVA $P<0.0001$) for every node on the tree. The difference between the optimal node ages in each model is clearly demonstrated in Figure 5.3 (i). The maximum node ages differ very significantly between each model (ANOVA $P<0.0001$) except for nodes 11 and 22. These correspond to the divergence between the insects *Chironomus tentans* and *Aedes albopictus* and between the annelid outgroup and the arthropods/nematodes. The minimum node ages also differ very significantly between each model (ANOVA $P<0.0001$) with the exception of nodes 1, 2 and 21. These correspond to the divergences between *C. elegans* and *C. briggsae,* between *P. superbus* and AF36 and between the arthropods and the nematodes. Figure 5.3 (ii) displays the maximum and minimum node ages for each model. This Figure clearly demonstrates that for our data the Ugamma model generates a much wider range of maximum and minimum dates, and for the majority of nodes the much tighter CIR dates sit within the Ugamma. The stricter CIR data were therefore selected for the final analysis as outlined in Section 5.2.4.

5.2.6 Effect of root age prior on molecular clock

A prior age of 550 MYA was specified for the divergence time of the root. A standard deviation is also specified for the mean root age. The effect of changing the standard deviation on the root age was analysed (Figure 5.4). This analysis was repeated 10 times and the mean divergence dates calculated. Changing the standard deviation of the root divergence time did not significantly change divergence time for the majority of the nodes (ANOVA $P<0.001$). The nodes closest to the root were significantly changed to a slightly younger date, but this had no effect on the nematode divergence time, the subjects of this study. The prior age of 550 MYA with a standard deviation of 50 was used in the final analysis as outlined in Section 5.2.4.

Table 5.4. The maximum, minimum and optimal dates for each node using the CIR and Ugamma models in the phylogeny.

Node	CIR Model			Ugamma Model		
	Max age (MYA)	Min age (MYA)	Optimal age (MYA)	Max age (MYA)	Min age (MYA)	Optimal age (MYA)
1	18.38	2.58	7.173	81.82	2.182	18.91
2	23.55	3.544	9.969	69.77	2.168	17.87
3	31.58	10.58	18.34	117.2	16.03	50.46
4	35.19	7.315	17.18	79.59	4.505	24.88
5	80.52	25.34	47.77	150	22.83	69.31
6	101.9	22.15	54.59	124.4	5.487	37.41
7	108.5	41.16	70.12	211.9	46.78	112.5
8	150.5	77.3	109.8	272.3	95.35	163.6
9	154.5	92.21	126.5	187.5	10.35	66.75
10	157.1	60.54	108.2	231.4	5.108	51.94
11	200	148.3	175.9	202.9	58.53	129.3
12	217.1	151.4	178.7	233.1	149.4	176.7
13	217.4	113.1	161.4	272.1	58.4	140.4
14	242.3	219.6	234	266.3	235.9	246.2
15	311.2	150.5	234.4	434.5	36.77	160.4
16	330.4	224.4	275.7	436.7	210.8	316.8
17	350.7	308.8	327.9	311.4	281.7	300.7
18	351.5	249.7	299.4	472.9	241.5	356.7
19	390.4	254.7	314.6	571.5	24.99	177.6
20	504.8	400.1	454.4	551.5	349.7	483.3
21	528.3	508.1	514.3	542	509.6	522.8
22	583.4	548.1	570.8	581.4	528.9	558.8
23	624	559.3	588.5	651.5	543.2	586.3

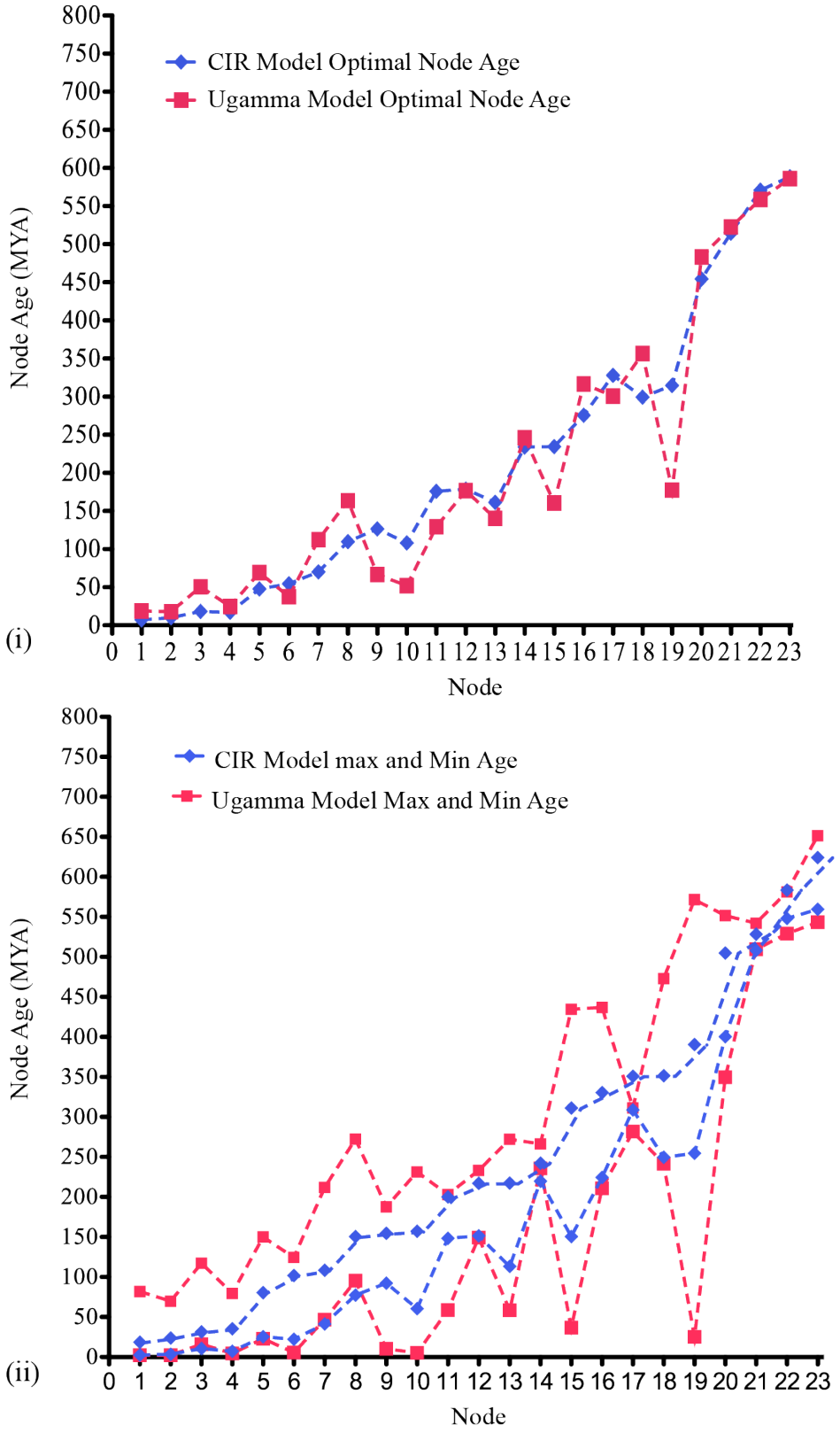


Figure 5.3. Effect of the autocorrelated model CIR and the uncorrelated model Ugamma on the molecular clock divergence estimates. Comparison between the optimal node ages (i) and a comparison between the maximum and minimum node ages using both models (ii).

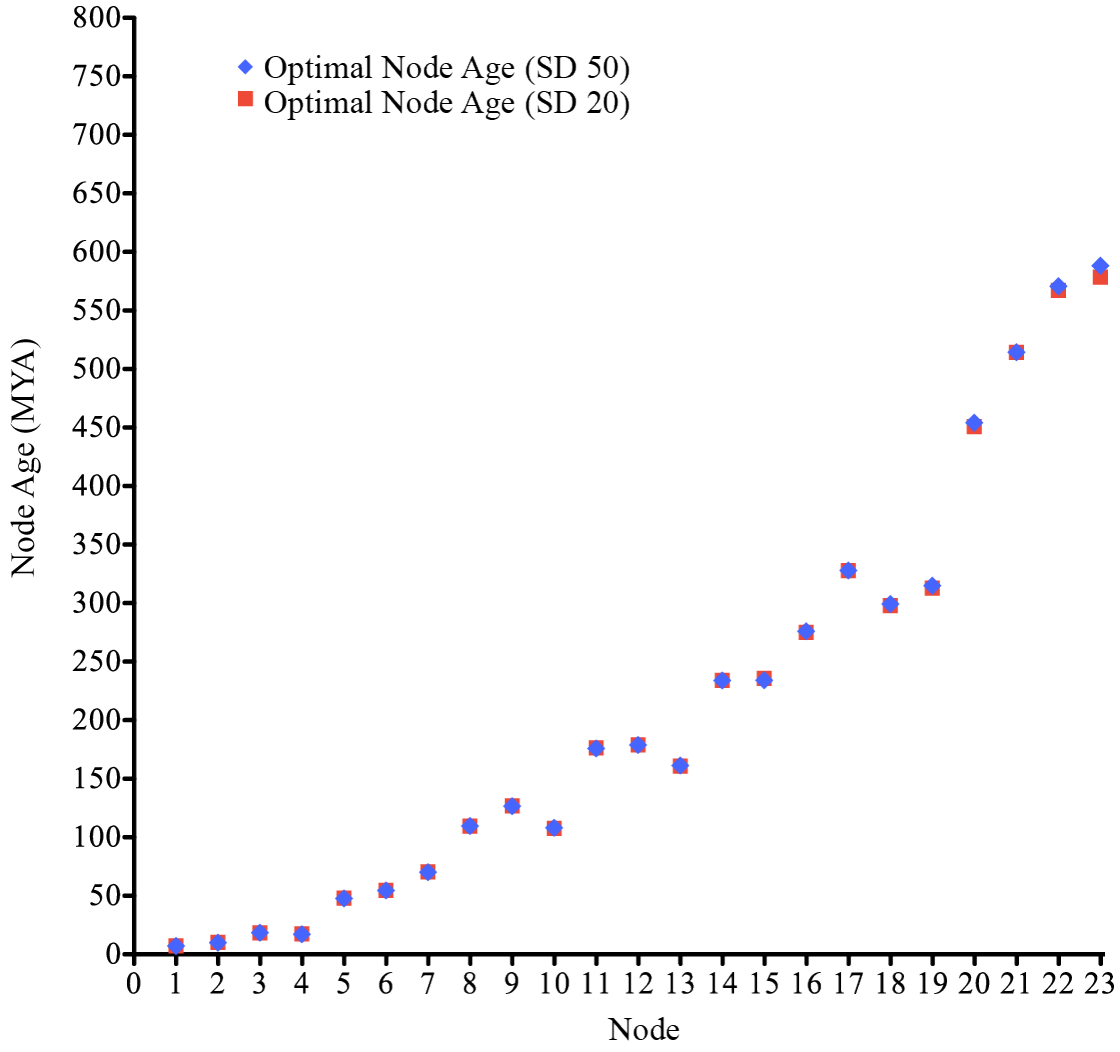


Figure 5.4. Effect of varying the standard deviation of the root age. The difference between the root divergence times when the standard deviation is varied is not significant for nodes 1-17 and is very significant for nodes 18-23. The changes do not affect any of the nematode divergence times. This analysis was repeated 10 times and the optimal ages averaged.

5.2.7 Effect of the soft-bound prior probability distribution on molecular clock calibration dates

A soft-bound prior probability distribution may be applied to the calibration dates. The default setting is 0.025, this means 2.5% of the total probability mass is allocated outside of each side of the specified lower and upper calibration points. The effect of varying the soft-bound prior on the divergence dates for each node was analysed (Figure 5.5). This analysis was replicated 10 times and the mean divergence times calculated. Changing the soft-bound prior does significantly change the node ages. Although changing this prior does significantly affect all of the nodes, it appears to have a larger impact on the nodes that have been calibrated rather than nodes near the root or the nematode nodes. As we do not have any calibration points within the nematode section of our tree it was decided that the strictest setting of the default value of 0.025 would be used in the final analysis as outlined in Section 5.2.4.

5.3 Discussion

Members of the Phylum Nematoda are soft-bodied and most are only millimetres in length, and thus fossilise badly. There are two main sources of nematode fossils. They can occur in amber, with records of insect parasites extending back 130 million years (Poinar, 2003). The other nematode source is coprolites of subfossils. Nematode eggs identified as being members of the family Ascarididae were found in Early Cretaceous dinosaur coprolites from Belgian strata (Poinar and Boucot, 2006). Poinar *et al.* (2008) have also described a fossil nematode *Palaeonema phyticum* (order Enoplia) recovered from the stomatal chambers of an early Devonian (396 MYA) fossilized prevascular land plant *Aglaophyton major* from the Rhynie chert fossil bed in Scotland. Since the nematode fossil record is poor and paleontological evidence of divergence is lacking for the vast majority of nematodes, including the nematodes used in this study, other methods to calculate divergence times are required. The molecular clock approach allows for divergence times to be estimated for the nematodes by using fossil data from other species as calibration points which allows the calculation of divergence time estimates other parts of the tree which lack calibration points.

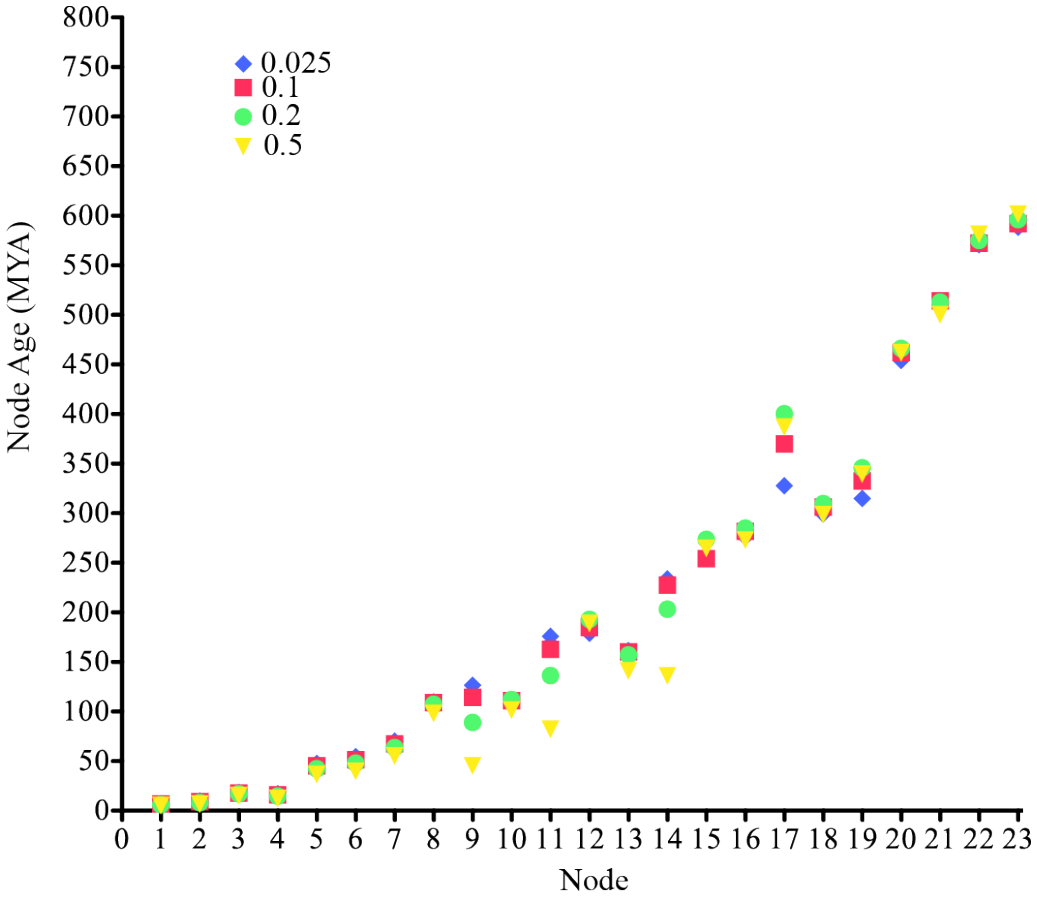


Figure 5.5. Effect of varying the soft-bound prior on divergence times. Changing the soft-bound prior significantly effected the divergence dates at all of the nodes, this effect is more evident at the nodes in the middle of the tree where they are calibrated.

Data from Borgonie *et al.* (2011) show by ^{14}C dating that the fracture water in the mine borehole from which the nematode *Halicephalobus mephisto* was isolated is 4,413-6,247 years old. These dates provide minimum constraints for the divergence of *H. mephisto* from *H. gingivalis*, but these dates were not included in this analysis as no maximum constraint data for the divergence of these two nematode species were available. Five dates were used to calibrate the tree (Table 5.2). These date estimates were deduced by Benton *et al*. and published in the *The Timetree of Life* (Benton et al., 2009). The clock analysis did not change the dates used to calibrate the clock significantly across the tree. The estimated molecular divergence times for the calibration nodes obtained using the autocorrelated CIR model of molecular evolution was very similar to the fossil calibration dates, with the optimal divergence dates lying within the range of the minimum and maximum dates proposed by (Benton et al., 2009). As the analysis did not change the calibration dates significantly it gives confidence that the dates on the tree between each of the calibration points are accurate. The only date that slightly differed was the estimated date of 327.9 MYA (308.8-350.7) for the divergence between the Hymenoptera and the Diptera. This date is older than the date used for calibrating the tree (238.5-307.2 MYA). However Wiegmann *et al.* (2009) present phylogenetic evidence that the Hymenoptera have a divergence time from the other holometabolous insects (including the Diptera) of ca. 350 MYA, and this date is within the range which was obtained for the divergence between the Hymenoptera and the Diptera using the autocorrelated CIR model of molecular evolution.

There have been estimates of the divergence dates within the Phylum Nematoda but this is an area of much debate. Cytochrome c and globin amino acid sequences were used to estimate the divergence between *C. elegans*, *T. colubriformis*, *N. brasiliensis*, *A. suum* and *P. decipiens* (Vanfleteren et al., 1994). This analysis estimated that the rhabditids (*C. elegans* and *C. briggsae*) and strongylids (*T. colubriformis* and *N. brasiliensis*) diverged from the ascarids over 500 MYA, the genera *Nippostrongylus* and *Trichostrongylus* diverged over 200 MYA and the *Ascaris* and *Pseudoterranova* diverged 150-250 MYA. These dates are all considerably older than the dates that were estimated for these nodes in this study. Nematode eggs identified as being members of the family Ascarididae were found in Early Cretaceous dinosaur coprolites from the Bernissart Iguanodon beds in Belgium (Poinar and Boucot, 2006).

These dinosaur beds belong to the Upper Barremian lowermost Aptian of the Creatatious period 124-127.24 MYA (Schnyder *et al*., 2009). The estimated divergence time for *Ascaris* (Ascaraidoidea, Ascaridae) and *Pseudoterranova* (Ascaraidoidea, Anasacidae) in the study under the CIR model is 60.54-157 MYA, with an optimal date of 108.2 MYA, which is in good agreement with the data for the Benissart nematode fossils.

Several groups have estimated the date for the divergence between *C. elegans* and *C. briggsae*. Butler *et al.* (1981) estimate the date of divergence to be 10-100 MYA, using a combination of 5S rRNA-sequences, anatomical differences, and protein electrophoretic motilities. Subsequent estimates based on DNA sequence data were 30-60 MYA (Prasad and Baillie, 1989), 23-32 MYA (Heschl and Baillie, 1990), 54-58 MYA (Lee et al., 1992) and 40 MYA (Kennedy et al., 1993). These dates are based on data from 1 to 7 genes. A date of 50-120 MYA was estimated by comparing the chromosomal rearrangements between the genomes of *C. elegans* and *C. briggsae* (Coughlan and Wolfe, 2002). Cutter (2008) estimated the time of the most recent common ancestor of *C. elegans* and *C. briggsae* ~18 MYA (range 11.6-29.9 MYA) by calculating the rate of neutral mutation accumulation and estimating a 60 day generation time for *C. elegans*. The estimated divergence time of 9.969 MYA (range 2.58-18.38 MYA) for *C. elegans* and *C. briggsae* in the analysis presented here overlaps with that of Cutter (2008). It is likely that Cutter's estimate of a recent divergence between *C. elegans* and *C. briggsae* is correct since nematodes appear to experience a higher mutation rate than other eukaryotes. Also a divergence time as old as 100 MYA would require an unrealistically slow generation time for these nematodes.

Within the order Rhabditida our molecular clock estimates place the divergence of the Panagrolaimidae in the Permian at 275.7 MYA (224.4-330.4). The estimated divergence time of *Panagrolaimus* sp. from *Halicephalobus* sp. (both members of the superfamily Panagrolaimidae) was in the Lower Cretaceous at 109.8 MYA (77.3-150.5). Within the *Panagrolaimus* sp. the divergence of AS01 from the remainder of the strains is predicted to occur in the Upper Cretaceous 70.12 MYA (41.16-108.5), with further diversification of the anhydrobiotic Panagrolaimids occurring during the Eocene and Miocene epochs of the Cenozoic Paleogene period. The Cretaceous was a

complex period that saw episodes of abrupt greenhouse warming and cooling and the end of the Cretaceous is marked by the global mass extinction of many faunal and floral groups, most notably the dinosaurs (Keller et al., 2009). During the Cretaceous the supercontinent Gondwana continued to break up, as South America separated from Africa and India began to drift away from the landmass containing Antarctica and Australia. Dinosaurs and palm trees were present in Antarctica and southern Australia during the Cretaceous period, and there were no polar ice caps. Figure 5.6 shows how the world looked at this time compared to how it is currently formed.

Our data predict that *P. davidi* and its sister species PS1579 diverged from *P. superbus* and its sister species AF36 during the Eocene epoch at 47.77 MYA (25.34-80.52) and we estimate that mean divergence times for the Antarctic *P. davidi* and the sub-Arctic *P. superbus* from their closest temperate congeners are 17.18 MYA (Miocene) and 9.96 MYA (Miocene), respectively. During the Eocene, Australia began to separate from Antarctica and drift northward, allowing the Antarctic Circumpolar Current to develop which ultimately resulted in the thermal isolation of the Antarctic continent (Kennett, 1977), and the progressive development of the Antarctic cryosphere through the Cenozoic era (Anderson et al., 2011). Thus, during the Cenozoic, Antarctica evolved from being a warm mainly ice-free continent, through transitional states of expanding and contracting ice sheets to extreme polar conditions, such as observed today. High rates of cooling occurred in Antarctica during the middle Miocene (ca. 14 MYA), a step, often referred to as the middle-Miocene climatic transition (Flower and Kennett, 1993). Since the late Miocene, the Antarctic ice sheets have repeatedly advanced onto the continental shelf, the frequency of ice sheet advance and retreat having increased during the Pliocene and Pleistocene (Anderson et al., 2011). Our data indicate that *P. davidi* separated from its temperate congeners during a period when the Antarctic cryosphere was expanding rapidly. Lewis *et al.* (2009) inferred divergence estimates of 14,000-140,000 years for *P. davidi* and the Californian PS1579 based on an external calibration derived from data on mutation rates in *C. elegans*. This analysis assumed one to six nematode generations per annum. It is difficult to estimate the number of generation times per year for a genus that can survive for more than eight years in an anhydrobiotic state

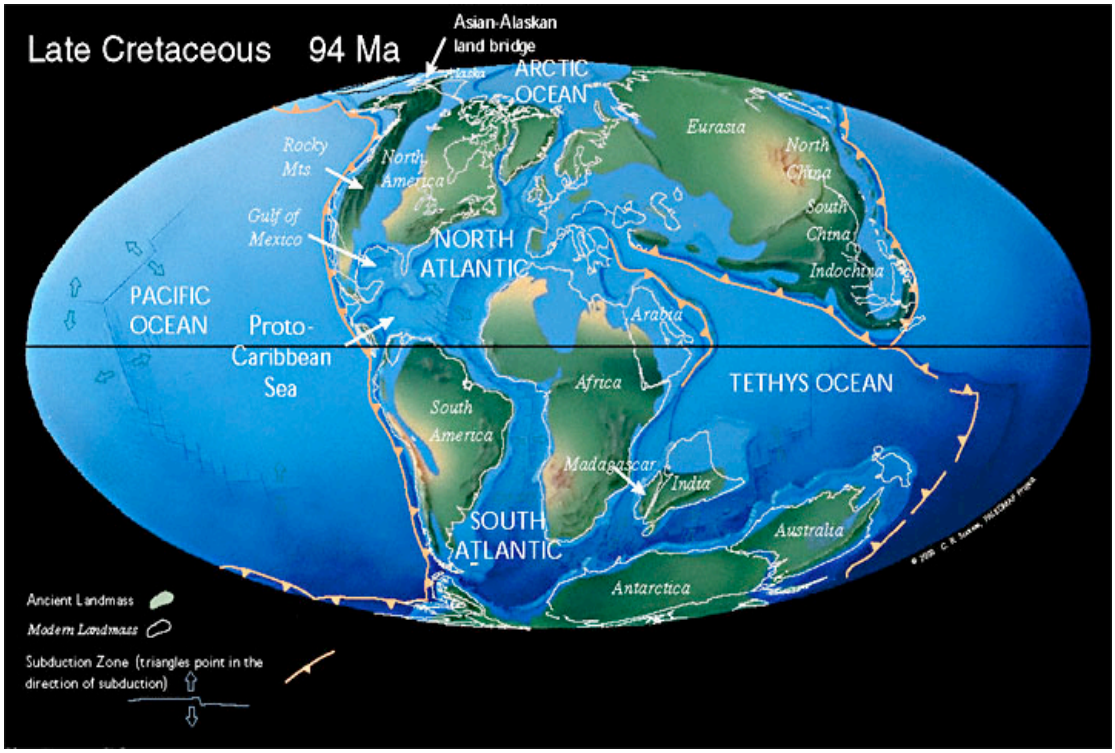


Figure 5.6. Atlas of the Earth during the Late Cretaceous period compared with its current form (outlined in white). The source of this image is www.scotese.com/cretaceo.html.

(Aroian et al., 1993). Our analyses provide a significantly older divergence date for these two nematodes.

Models of the Antarctic ice sheet Last Glacial Maximum (~20,000 years ago) leave no ice-free refuges for most terrestrial biota (Huybrechts, 2002), which would imply that contemporary Antarctic biota have colonized the region following the Last Glacial Maximum. But surprisingly the majority of Antarctic micro-invertebrate taxa show a high degree of endemism which points to the possibility that these endemic taxa are pre-Pleistocene relics (Convey et al., 2008). All currently described Antarctic nematode species are considered to be endemic to the region (Andrássy, 1998; Maslen and Convey, 2006).

The Antarctic specialist nematode *Scottnema lindsayae*, the sole member of a monotypic genus is abundant and widespread where suitable ice free soil occurs. In the laboratory its optimal growth temperature is 10 ^{o}C. Its reproductive cycle takes a year and its reproductive capacity declines if grown at higher temperatures (Overhoff et al., 1971). By contrast the optimal temperature for the polar isolates *P. davidi, P. superbus* and *P. detritophagous* is ~25 ^{o}C (Sohlenius, 1988; Brown *et al.*, 2004) and their life cycles are ~8 days at the optimal temperature. These physiological and life history traits do not show evidence of an evolved response to polar environmental conditions. Thus based on its physiology, phylogeny, and estimated time of divergence it seems unlikely that *P. davidi* has existed in Antarctica since the mid Miocene period. On balance it seems more likely that *P. davidi* was transported to Antarctica in anhydrobiotic state by a wide-ranging polar sea bird. Since *P. davidi* is a hermaphrodite, it would be possible to establish a nematode population from a single worm. The Phylum Nematoda contains a great diversity of anhydrobiotic taxa. Anhydrobiosis aids nematode dispersal and protects against desiccation and may provide cross-tolerance to cold stress. These specialist features may be the reason why nematodes are among the most abundant and diverse invertebrates in Antarctic soils and vegetation and why they were also early invertebrate colonizers of the nunatak at Surtsey, Iceland after the volcanic formation of this island.

Chapter VI Ice affinity protein purification

6.1 Introduction

Several proteins are associated with the response to cold and freezing temperatures. This Chapter describes the attempted purification of an ice-binding protein (IBP) /antifreeze protein (AFP) from the desiccation and freezing tolerant nematode *P. superbus* using ice affinity purification. AFPs are ice-binding proteins that are used as a freeze-avoidance strategy in some fish, insects, plants and microbes. AFPs adsorb to a growing ice crystal and restrict the growth of the ice front. This causes a non-colligative, non-equilibrium lowering of the freezing point of tissue fluids, without significantly altering their melting point. This phenomenon is termed thermal hysteresis (TH) (Raymond and DeVries, 1977). Certain AFPs known as recrystallisation inhibition proteins (RIPs) may have low TH activity, but can control the growth of ice limiting the growth of small ice crystals into large damaging ones (Knight et al., 1984).

The Antarctic nematode *P. davidi* is believed to produce an IBP (Wharton et al., 2005). This protein can produce a hexagonal ice crystal but has very low TH activity; therefore it is thought that this AFP is a RIP. The cold adapted nematode *Plectus murrayi* is the only nematode that is has a documented DNA sequence with homology to an existing AFP. A molecular study found a *P. murrayi* expressed sequence tag (EST) that had high similarity a Type II AFP from the Atlantic herring (Adhikari et al., 2009). Fish Type II AFPs are considered to have evolved from C-type lectins (Ewart and Fletcher, 1993) and lectin genes are very abundant in nematodes (Drickamer and Dodd, 1999). Thus functional characterisation of the protein encoded by the *P. murrayi* EST will be required to determine whether this protein has ice-binding activities or corresponds to a C-type lectin. Neither of these proteins or any other nematode IBP have been isolated to date. The ability of the freezing and desiccation tolerant *P. superbus* protein extracts to shape ice as seen in Chapter 3 (Figure 3.5) suggested that *P. superbus* may possess an ice-binding protein.

The remarkable diversity of the antifreeze protein sequences and structures makes their identification and isolation very difficult. A detailed review of the AFPs that

have been characterised to date is provided in Chapter 1 Section 1.5.3. Often an AFP that is present in one species may not be present in a closely related species, or it may be completely different to all other known AFPs. The fish AFPs may be divided into five classes, Type I-V. There is no correlation between AFP type and the phylogeny of the host fish. The Type I AFPs all have very similar 11 amino-acid repeat sequences, but they are found in fish from three different taxonomic orders (Davies and Hew, 1990; Harding *et al.*, 1999). This convergent evolution is also seen in the Type II AFP host fish. Type II AFPs have been found in the Atlantic herring, rainbow smelt, Japanese smelt, sea raven and longsnout poacher (Ewart *et al.*, 1992; Sonnichsen *et al.*, 1995; Ewart *et al.*, 1996; Gronwald *et al.*, 1998). The Type III AFPs are found in eelpouts and wolfish - these fish are actually closely related, further demonstrating the lack of a pattern in which fish will have AFPs and which will not (Schrag *et al.*, 1987, Scott *et al.*, 1988, Cheng and DeVries, 1989, Wang *et al.*, 1995). The purified insect, plant, algal and microbial AFPs are all different in sequence and structure to the fish AFPs. However, the yeast AFP has been shown to have high sequence similarity to the fungal, algal and bacterial AFPs (Lee et al., 2010).

The identity of an AFP may also not be predicted by extrapolation from the evolutionary precursor of previously isolated AFPs. There are no clear similarities between the precursors proteins from which isolated AFPs appear to have evolved. Type II AFP is homologous to the carbohydrate recognition domain of calcium dependent (C-type) lectins (Ewart and Fletcher, 1993). Type III AFPs have some sequence similarity with the extreme C-terminal region of the enzyme sialic acid synthase (SAS) (Baardsnes and Davies, 2001). Type IV AFPs are similar to apolipoproteins (Deng et al., 1997) and the Type V AFPs are derived from a gene encoding a pancreatic trypsinogen (Chen et al., 1997b). The evolutionary precursors for the Type I AFP and the insect AFPs are not known. Plants AFPs have very interesting evolutionary precursors. Seven AFPs have been isolated from winter rye and individual AFPs from rye have homology to diverse pathogenesis-related proteins, β-1, 3-glucanases, chitinases and thaumatins (Hon et al., 1995). Carrot AFP has sequence similarity to polygalacturonase inhibitor proteins (PGIPs) that contain a leucine-rich repeats (LRR). The carrot AFP cannot inhibit fungal polygalacturonase activity so it may have evolved from a LRR region that acquired the ability to bind ice

and lost its primary function (Worrall *et al.*, 1998; Meyer *et al.*, 1999). The evolutionary origins of bacterial, fungal and algal AFPs have not yet been identified (Raymond, 2009; Lee *et al.*, 2010).

The diversity of AFPs has made their purification and identification very difficult. Even where the genome sequence is available for a species with an AFP, it can be still difficult to identify the protein as it may have a novel sequence and structure. The majority of the AFPs that have been isolated so far have been obtained through biochemical methods. The AFPs are often purified from crude protein extracts through a combination of gel-exclusion and ion-exchange chromatography. The eluted fractions that have TH activity are pooled and may be further purified using reverse-phase HPLC (Graham *et al.*, 1997; Tyshenko *et al.*, 1997; Worrall *et al.*, 1998; Simpson *et al.*, 2005).

A method that has been developed for the purification of AFPs from microbes and algae takes advantage of the only trait that defines AFPs, their affinity for ice (Raymond and Fritsen, 2000). With this method the AFP is purified from the culture medium supernatant using three freeze-thaw cycles. In each cycle, the sample is adjusted to an osmolarity of approximately 300 mOsm kg^{-1} and frozen overnight in 500 ml centrifuge bottles at -15 ^{o}C. The bottles are centrifuged upside down expel brine and unbound proteins from the medium. The thawed ice fraction contains the IBPs. This method has been successful in purifying or semi-purifying AFPs from an Antarctic sea bacterium (Raymond et al., 2007) and algae (Raymond and Fritsen, 2001; Raymond and Knight, 2003; Janech *et al.*, 2006; Raymond, 2009). However this method is only suitable in cases where the antifreeze protein would be released out of an organism into its surrounding environment. This is unlikely to occur in species other than microbes, as AFPs are generally produced to protect internal structures from the damaging effects of ice growth.

Another ice-affinity method that can be applied to intracellular AFPs has been developed by Kuiper *et al.* (2003). In this method a hollow metal finger is kept at temperatures below freezing by circulating a mixture of ethylene glycol and water from a low temperature water bath through the central cavity of the "cold finger" (Figure 6.1). The cold finger is seeded with a thin layer of ice and placed into the

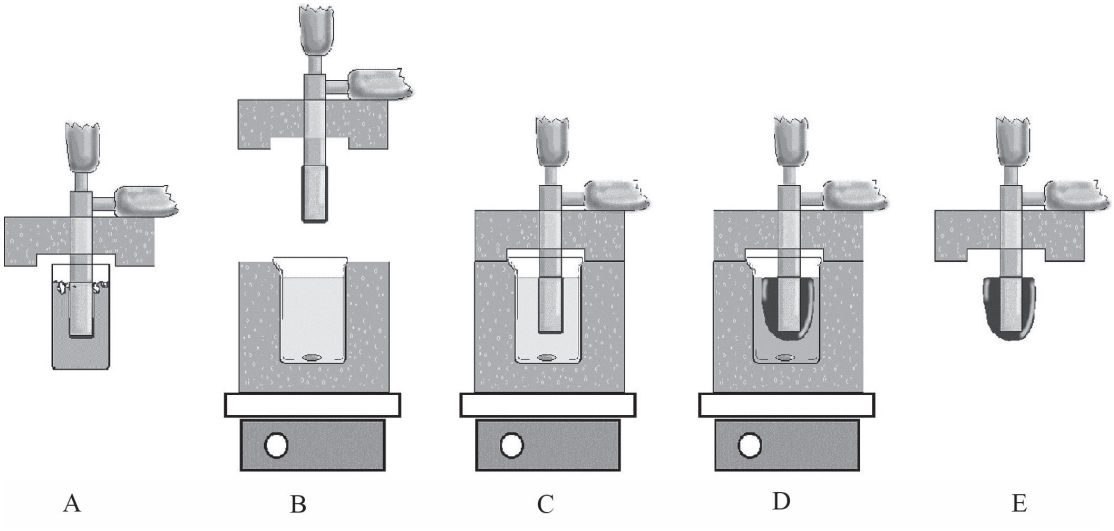


Figure 6.1. General method for ice-affinity purification. (A) Seeding the cold finger with ice; (B-D) growing the ice; (E) the ice fraction containing the bound AFPs. A magnetic stirrer is used to keep the protein extract constantly mixing and the beaker containing the protein extract is insulated with polystyrene. The source of the illustration is Kuiper *et al.* (2003).

crude protein extract. Under conditions where starting protein extract concentration is low, the temperature is slowly reduced and the solution is constantly mixed any IBPs will adsorb into the ice and become trapped, while the other proteins which lack affinity for ice will become excluded. This method has been successfully used in the purification of AFPs from the snow flea (Graham and Davies, 2005). It has also been used to purify recombinantly produced AFPs (Middleton et al., 2009).

Biochemical purification relies on an effective TH assay to determine the fractions that contain the IBP. In cases where the IBP may have very low TH activity or be only active at high concentrations IBP purification using biochemical fractionation would be a very time-consuming and difficult task. For these reasons IAP was used in this study, in an attempt to purify IBPs from *P. superbus*. Protein bands that appeared to be selectively adsorbed into the ice fractions were analysed using liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS). LC-MS/MS is a highly sensitive method for the identification of proteins. The two key components of this process are the ion source that generates the ions, and the mass analyser that sorts the ions. Protein mixtures are digested into peptides with the enzyme trypsin. These peptides are injected into a high performance liquid chromatography column that is coupled with the mass spectrometer. The peptides are eluted from the column in order of their hydrophobicity, the hydrophilic peptides will elute first. When the peptide flows to the end of the column it flows through a needle. At the end of the needle the liquid is nebulised and the peptide is ionised by a process is called electrospray ionisation (ESI) (Fenn et al., 1989). ESI operates at atmospheric pressure and produces tiny charged droplets when a high electric potential is set in the chamber between the chromatography column and the mass analyser. By using heated drying gas, the charged droplets shrink and the charge concentration in the droplets increases. The repulsive force between ions with the same charges will cause the ions to be ejected from the droplets into the gas-phase where they are attracted towards the mass analyser. The ESI uses a steady stream of solvent to produce a continuous beam of ions (Lin et al., 2003). Electrosprayed peptide ions enter the mass analyser through a small hole or a transfer capillary. The mass analyzer separates the ions by their mass-to-charge (m/z) ratios. Magnetic and/or electric fields in a vacuum are used to manipulate the ions in a mass-dependent manner (Lin *et al.*, 2003; Steen and Mann, 2004).

From the m/z value the peptide mass fingerprint (PMF) may be generated. The mass accuracy and the percentage of the target protein sequence covered are analysed using scoring algorithms that calculate the level of confidence for the match. The MASCOT MS/MS ion search (http://www.matrixscience.com/) may be used to search the various databases for identities to raw MS/MS data. The Spectrum Mill software provided with Agilent LC/MS systems may also be use to search custom made databases using the MS spectral data. The PMFs of *P. superbus* peptides were searched against the NCBI database (http://www.ncbi.nlm.nih.gov/) and also against a custom database of *P. superbus* peptides.

6.2 Results

6.2.1 Optimisation of ice-binding purification using a positive control

A modified version of the ice-binding finger described by Kuiper *et al.* (2003) was constructed by Mr. Joe O'Sullivan (Biology Dept., NUIM). The apparatus was cooled to –1.5 ^{o}C and immersed in chilled distilled water containing pieces of ice that nucleate freezing, forming a thin layer of ice around the finger. A recombinant globular fish type III AFP (HPLC 12) from the eel pout *Macrozoarces americanus* was acquired from Prof. Peter Davies (Queen's University, Kingston, ON, Canada) to use as a positive control. The protein was expressed in *E. coli* and the crude protein extract (extracted as described in Sections 2.2.5.5 and 2.2.5.6) used as the starting material for ice affinity purification. The ice was allowed to grow slowly on the ice finger for 2 days or until the sphere of ice (Figure 6.2) contained 50% of the crude bacterial protein extract. The frozen ice fraction has a visible lattice pattern indicating the presence of an ice binding protein. This sphere of ice, the "ice fraction", was melted off the finger and frozen at -80 ^{o}C. To maximise the concentration of the proteins contained in the ice fraction the remaining liquid fraction was used again as a starting point for another round of affinity purification and diluted to bring the volume to 80 ml. This was repeated a total of three times and the ice fractions were pooled together. An aliquot (6 x 1ml) of this ice fraction was taken for analysis on a SDS-PAGE gel and the remainder was used as a starting material for a second round of purification. The second round of purification was also repeated three times. The three ice fractions were combined from round 2, freeze-dry lyophilised and resuspended in water. 1D SDS-PAGE was used to separate the proteins present in

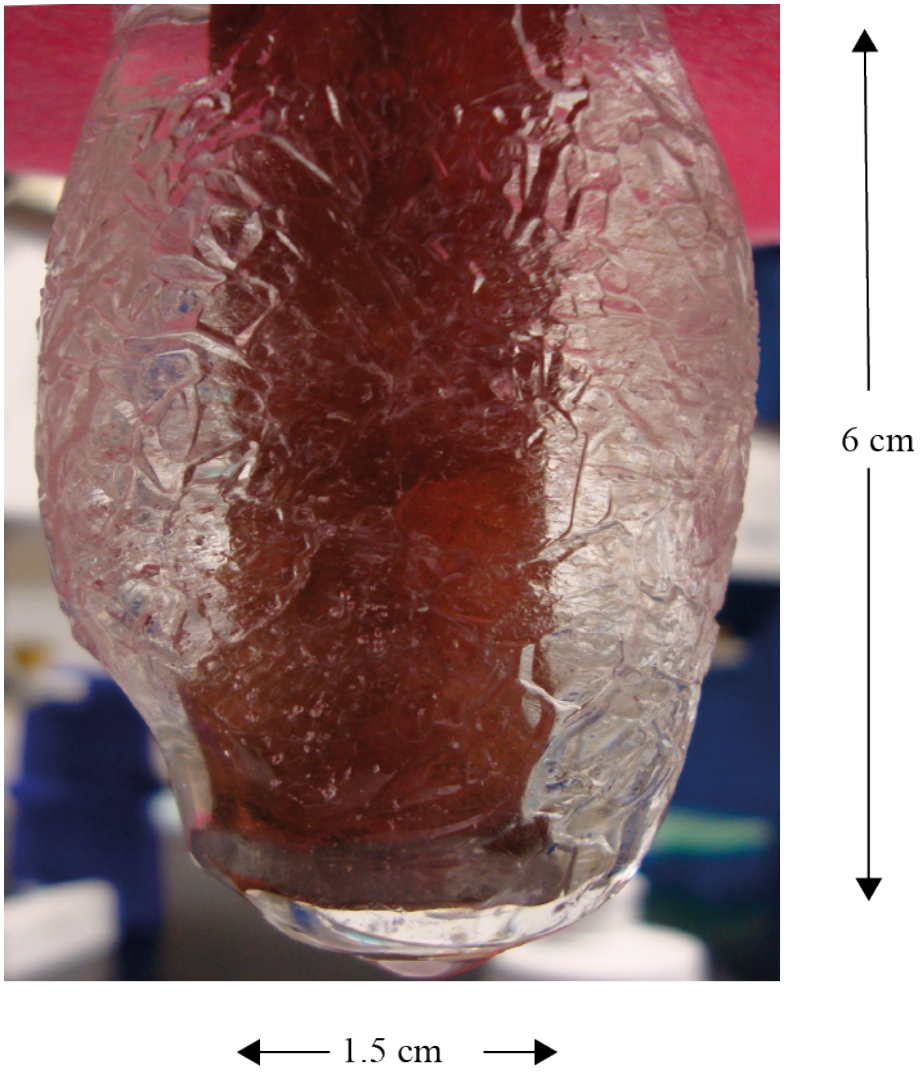


Figure 6.2. An ice sphere grown from an *E. coli* protein extract containing recombinant Type III AFP from eel pout. The grooved pattern on the ice surface suggests the presence of an ice-binding protein.

each of the fractions (Figure 6.3). The 7.0 kDa band corresponding to HPLC 12 (Sonnichsen et al., 1993) was not clearly visible in the liquid fractions after each round of purification. It is clearly visible in the ice fraction after one round of purification and is highly enriched in the ice fraction after two rounds. The non-specific binding of the other *E. coli* proteins are reduced after a second round of ice affinity purification. This purification protocol was repeated independently with a second batch of *E. coli* recombinant protein extract and yielded similar results.

6.2.2 LC/MS/MS of the positive control

The two protein bands corresponding to HPLC 12 from separate ice affinity purification experiments were excised from the protein gel (Figure 6.3). They were digested with trypsin and analysed by LC-MS/MS. The peptides were identified by searching against the NCBI database with MASCOT MS/MS online search software (Table 6.1). The resulting peptides had hits to the desired Type III AFP HPLC 12. There were also hits to a major outer membrane lipoprotein and keratin. The major outer membrane lipoprotein is non-specific adsorbent to the ice fraction. The keratin likely corresponds to human protein contamination, as it is found in all layers of human skin including the stratum corneum (the outermost layer of the epidermis).

6.2.3 Ice-binding purification of *P. superbus*

Soluble proteins (approximately 1 mg/ml in a total of 80 ml) were extracted from *P. superbus* as outlined in Section 2.2.5.1, and these were subjected to IAP as described in Section 2.2.5.8. The frozen ice fraction showed some grooves, indicating the possible presence of an IBP shaping the ice, but this etching was significantly less than that seen in the positive control (Figure 6.4). The SDS-PAGE protein profiles differed between the starting, liquid and ice fractions. It appeared that bands that were in the starting fractions are less concentrated in the liquid fraction and that the proteins corresponding to these bands are being adsorbed into the ice (Figure 6.5 (i)). The experiment was repeated a second time (Figure 6.5 (ii)) and the similar banding patterns were obtained. There was a large amount of non-specific binding of proteins to the ice, as has also been observed in the lysates from recombinant *E. coli*. A second round of ice-binding purification was attempted after pooling several round 1 ice fractions together as a starting fraction but no proteins were visible on the resulting SDS-PAGE gel. Protein bands that appear to be adsorbing strongly into the ice were

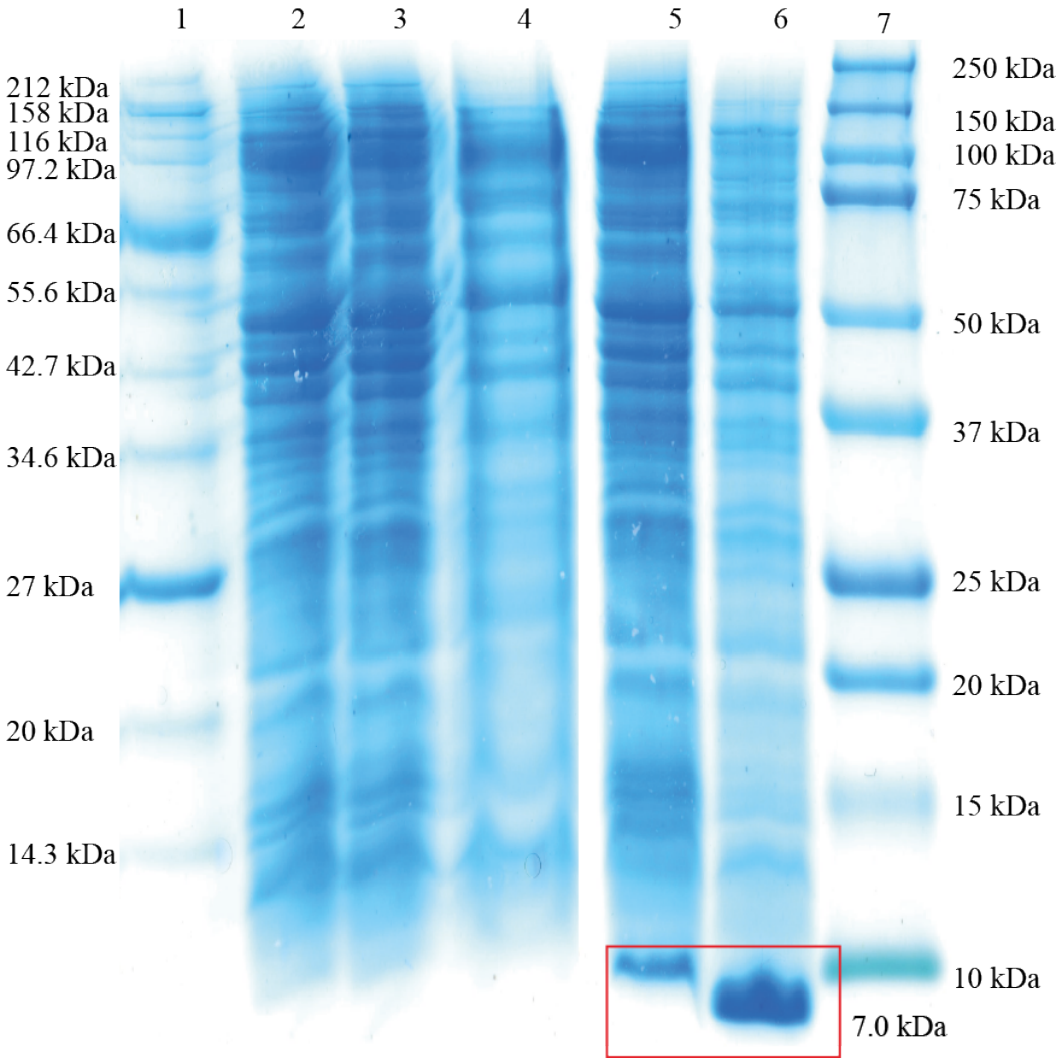


Figure 6.3. 1D SDS-PAGE analysis showing separation of proteins from lysates of recombinant *E. coli* that were recovered during ice-affinity purification experiments. The gel was loaded as follows: Lane 1; Broad range marker (NEB), Lane 2; Starting fraction, Lane 3; Liquid fraction (round 1), Lane 4; Liquid fraction (round 2), Lane 5; Ice fraction (round 1), Lane 6; Ice fraction (round 2), Lane 7; Precision plus protein marker (Biorad). Each lane contained 40 μg of protein. The desired 7.0 kDa bands for HPLC12 in each of the ice fractions are indicated by the red box.

Table 6.1: LC-MS/MS data for the enriched 7.0 kDa protein found in the ice fraction from an *E. coli protein* extract containing recombinant Type III AFP from eel pout.

Sample	Peptides found	Protein ID	Accession number	Sequence coverage by the peptide	Score
1	R.SEVVTPVGIPAEDIPR.L R.LVSMQVNR.A R.AVPLGTTLMPDMVK.G R.AVPLGTTLMPDMVK.G	Type-III ice-structuring protein HPLC 12 (*Macrozoarces americanus*)	P19614	57%	157
	R.ALEESNYELEGK.I	Keratin, type I cytoskeletal 10 (*Homo sapiens*)	NP_000412	2%	99
	K.IDQLSSDVQTLNAK.V	Major outer membrane lipoprotein (*Serratia mercescens*)	P02938	18%	71
2	-.MNQASVVANQLIPINTALTLVMMR.S R.SEVVTPVGIPAEDIPR.L R.LVSMQVNR.A	Type-III ice-structuring protein HPLC 12 (*Macrozoarces americanus*)	1HG7_A	93%	179
	K.IDQLSSDVQTLNAK.V K.IDQLSSDVQTLNAK.V K.VDQLSNDVNAMR.S K.VDQLSNDVNAMR.S	Major outer membrane lipoprotein (*Serratia mercescens*)	P02938	33%	148

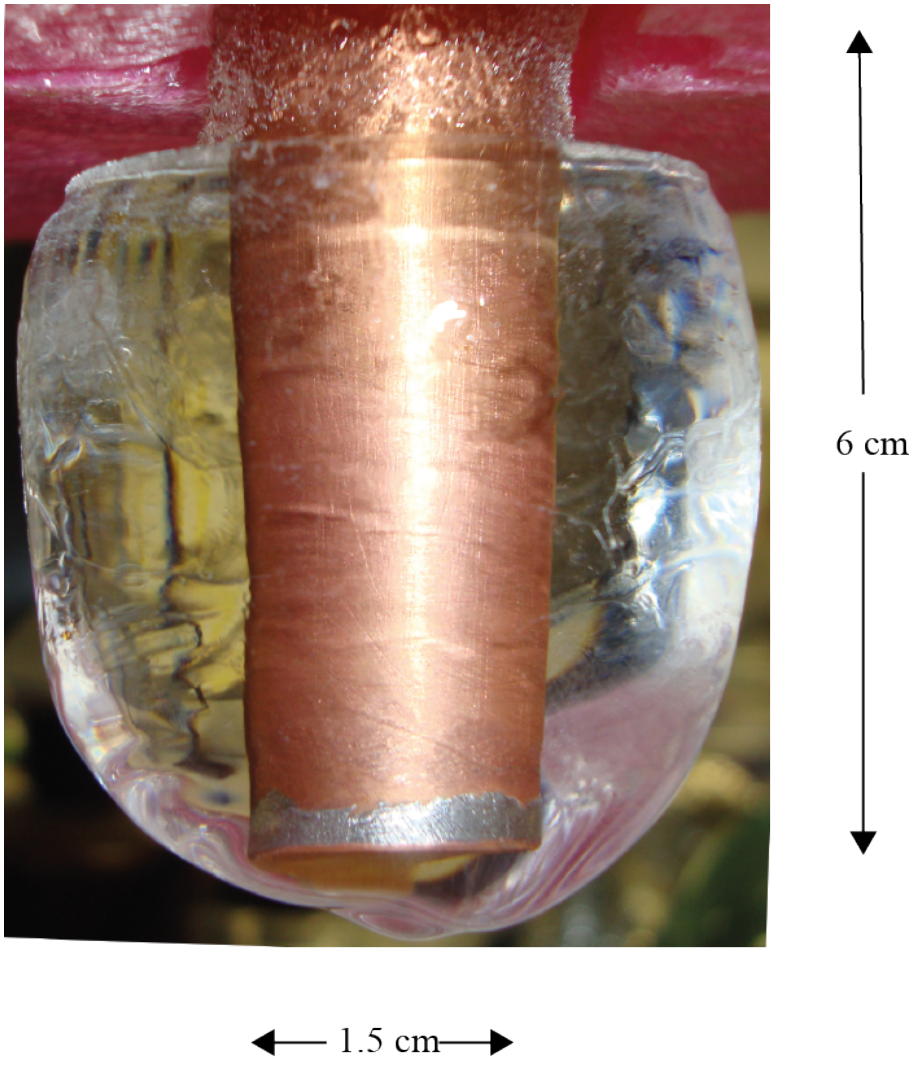


Figure 6.4. An ice sphere grown from *P. superbus* protein extract. A grooving pattern on the ice surface suggests the possible presence of an ice-binding protein.

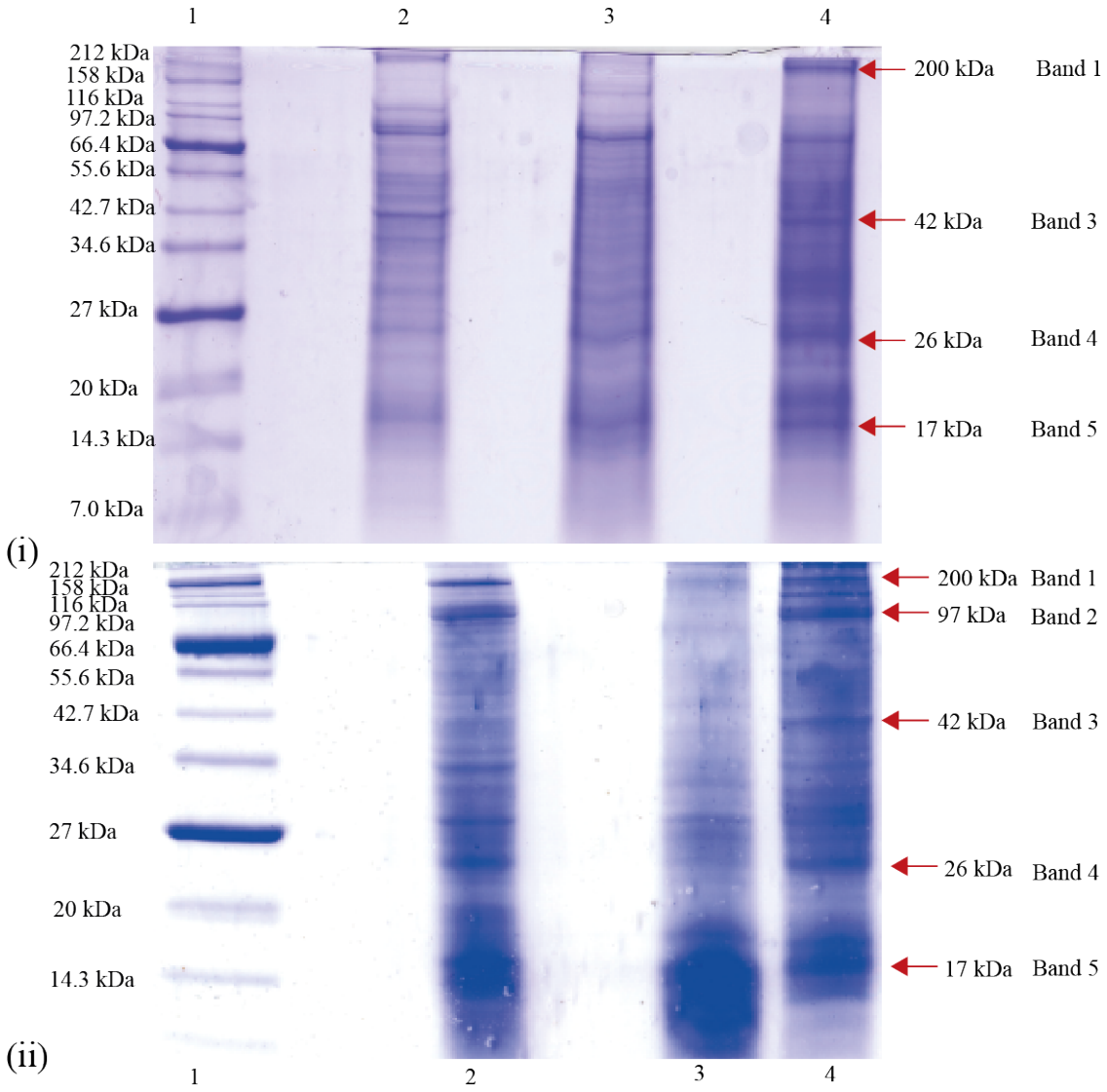


Figure 6.5. 1D SDS-PAGE analysis of protein profiles of ice-binding fractions from *P. superbus*. The experiment was carried out twice from two independent starting fractions and one gel from each experiment is shown in panel (i) and (ii). The gel was loaded as follows: Lane 1; Broad range marker (NEB), Lane 2; Starting fraction, Lane 3; Liquid fraction and Lane 4; Ice fraction. Arrows indicate the most intensely staining bands from the ice fraction, which were selected for mass spectrometry analysis.

excised and analysed by LC-MSMS.

6.2.4 LC-MS/MS of *P. superbus* ice-adsorbing proteins

Peptides obtained using LC/MS/MS were initially searched against the NCBI database using the MASCOT MS/MS online search software. Our *P. superbus* peptides matched none or very few of the proteins in the database. The sequencing of the *P. superbus* transcriptome and genome is ongoing (Georgina O'Mahony, NUIM). The latest assembly of the *P. superbus* transcriptome was processed through the prot4EST pipeline by Dr. Stephen Bridgett (University of Edinburgh). Prot4EST is a suite of programs that takes the expressed sequences and translates them optimally to produce putative peptides (Wasmuth and Blaxter, 2004). This custom *P. superbus* database (PST) was searched using the Spectrum Mill software provided by Agilent. The *P. superbus* sequences that had matches to the peptide fragments were searched for identities against the NCBI protein database using the BLASTp program.

The identities of the proteins found in each band excised from the gel are shown in Table 6.2. Peptides from band 1 (whose molecular mass was ca. 200 kDa) found hits in our *P. superbus* database to three proteins. There were several peptide matches to these proteins sequences, giving confidence to the identification. Two proteins were identified as vitellogenin, an abundant component of yolk, and the third high molecular weight protein identified was paramyosin, a major component of the muscle thick filament of many invertebrates including *C. elegans* (Moerman and Fire, 1977). None of the peptides isolated from band 2 had hits to the *P. superbus* protein database. Peptides from band 3 (molecular mass ca. 42 kDa) found hits to two proteins, but both of these were identified using only a single peptide match. One peptide match was to vitellogenin; the second match was to UDP-glucuronosyltransferase. Band 4 (molecular mass ca. 42 kDa) peptides had hits to eleven proteins; but only two of the proteins were identified with matches to more than one peptide. These were elongation factor 1 gamma and a myosin family protein. All 11 proteins putatively associated with band 4 were either enzymes or were proteins involved in protein interactions and thus any one of these proteins could potentially be the evolutionary precursor of an AFP protein. However none of the 11 proteins were highly enriched in band 4 and it seems more likely that these proteins

Table 6.2. LC-MS/MS data for the peptides recovered from enriched protein bands from the *P. superbus* ice fraction. The peptides were searched against the *P. superbus* database (PST) using Spectrum mill. These PST sequences were then searched for identities against the NCBI database using BLASTp.

Band	Peptides found	*P. superbus* transcript ID	BLAST ID of PST	BLAST Percentage Identity of PST	Accession number
1	(K)EESLLYSSEWDLIEK(R) (K)HLSLSIDSLFNGLFEK(D) (K)AIGNAGNELAVLNLEK(I) (R)AILNLVQLNLK(D) (R)MTAFSMLMATFPDK(V) (K)HLSLSIDSLFNGLFEK(D) (R)MTAFSMLMATFPDK(V) (K)DTVSFSLTQQNLEDLYHR(F) (R)QLSSLPAISDK(Q) (K)ALSTSMIPEER(E)	PST01866	Vitellogenin-6 (*Ascaris suum*)	31%	ADY40227
	(K)AVLDMIMTEMIR(E) (K)TFMQGILSDAMAVAGTK(H) (R)VQITAEHLEK(L) (K)LNLPFQIR(F)	PST20106	Vitellogenin (*Oscheius tipulae*)	25%	AAB49749
	(R)LEDAEGTTDSQIEANR(K) (K)YQLAQQLEEAR(R) (R)ALAELQQLR(V)	PST12780	Paramyosin, UNC-15 (*Caenorhabditis elegans*)	90%	NP_492085
3	(R)AILNLVQLNLK(D)	PST01866	Vitellogenin-6 (*Ascaris suum*)	31%	ADY40227
	(K)IGILVPDIAGSQFLFNLR(V)	PST00450	UDP-glucuronosyltransferase, UGT-50 (*Ascaris suum)*	48%	ADY43319

Table 6.2.

Band	Peptides found	*P. superbus* transcript ID number	BLAST ID of PST	BLAST Percentage Identity of PST	Accession number
4	(R)GYSFTTTAER(E)	PST00010	Actin (*Panagrellus redivivus*)	100%	AAM47606
	(R)AVPLALSLLSISNPQLNVLETLSK(Y)	PST20617	26S proteasome non-ATPase regulatory subunit 2 (*Ascaris suum*)	70%	ADY41709
	(K)TFFPIFSNYLK(E)	PST10293	Glutathione S-transferase 1 (*Ascaris suum*)	41%	ADY45818
	(K)KKPAAAAEEPEGGDETPQEK(F) (R)FSYADASVALDLLPAFQNVLTQK(E)	PST21698	Elongation factor 1 gamma	53%	XP_001901841
	(R)VGELWILGDVFIGURER(Y)	PST00767	Cathepsin D (*Ascaris suum*)	68%	ADY43078
	(R)DVLQELLATLENAK(Y)	PST07527	Cystathionine gamma-lyase 2 (*Ascaris suum*)	62%	ADY41727
	(K)GPAAPASPYYGGDTR(L)	PST06482	Cytochrome b-c1 complex subunit 2 (*Ascaris suum*)	49%	ADY44125
	(K)GEETSTNPIASIFAWTR(G)	PST22916	Isocitrate dehydrogenase NADP cytoplasmic (*Ascaris suum*)	83%	ADY46768
	(K)LEQQLDDLEDNVER(E)	PST29541	Myosin heavy chain (*Haemonchus contortus)*	71%	ABC40752
	(K)QLASQGTSGDALEIAK(K)	PST06482	Cytochrome b-c1 complex subunit 2 (*Ascaris suum*)	49%	ADY44125
	(R)LDEAGGATAAQIETNK(K) (R)LDEQGGATAAQLEVNK(K) (R)LEDEQSLVGK(L)	PST18737	Myosin family protein (*Brugia malayi*)	73%	XP_001901629

Table 6.2.

Band	Peptides found	*P. superbus* transcript ID	BLAST ID of PST	BLAST Percentage Identity of PST	Accession number
5	(K)TVLIMELINNVAK(A)	PST16759	ATP synthase subunit protein (*Caenorhabditis briggsae)*	88%	XP_002642779
	(K)ILPYLLFLIGK(N)	PST23219	Cytosolic glutathione S-transferase (GST) (*Caenorhabditis briggsae)*	35%	XP_002630613
	(K)EHLYDNLETYLPVFSTFLK(E)	PST22216	Cytosolic glutathione S-transferase 2 (*Oesophagostomum dentatum*)	38%	ACA30415
	(K)VPILEFDGITLVESAAISR(Y)	PST01874	Cytosolic glutathione S-transferase 1 (*Ascaris suum*)	45%	P46436
	(R)LILHHANVDFEDVR(I)	PST04300	Cytosolic glutathione S-transferase 1 *(Ascaris suum)*	42%	P46436
	(K)VHINIVVIGHVDSGK(S)	PST10075	Eukaryotic translation elongation factor 1A (*Ancylostoma ceylanicum*)	92%	DAA05868
	(K)EFEIIDLLIPNLRDEVLK(I)	PST28714	40S ribosomal protein S2 (*Bursaphelenchus xylophilus*)	86%	ACZ13328
	(K)VLEQLTGQTPVLSK(A)	PST22290	60S ribosomal protein L11 (*Loa loa*)	72%	XP_003138893
	(K)SLDLLNYGEPR(L)	PST17672	Cytochrome c oxidase subunit II (*Caenorhabditis briggsae*)	83%	ADB96816
	(R)LVQAFQFVDK(H)	PST01066	Thioredoxin peroxidase (*Dirofilaria immitis*)	82%	AAC38831
	(R)GYSFTTTAER(E)	PST00010	Actin (*Panagrellus redivivus*)	100%	AAM47606
	(K)LVIPLLIIR(A)	PST09681	Sigma non-opioid intracellular receptor 1 (*Ascaris suum*)	41%	ADY43258
	(R)IISMQNGGNVGQVLER(L)	PST08732	Arginine kinase (*Heterodera glycines*)	92%	AAO49799
	(K)AGPIVAMVWEGLDAVR(Q)	PST11397	Nucleoside diphosphate kinase (*Ajellomyces capsulatus* G186AR)	70%	EEH05396

were present in the ice fraction as a result of non-specific binding. Peptides from band 5 (molecular mass ca. 17 kDa) had hits to fourteen proteins. All of these were identified by matches to only one peptide each, although four peptides gave hits to the enzyme glutathione S-transferase, these hits corresponded to three distinct *P. superbus* transcripts. The proteins putatively associated with band 5 corresponded to enzymes, ribosomal proteins and actin. Similarly to the band 4 proteins, any one of these proteins could potentially be the evolutionary precursor of an AFP protein. However, none of the band 5 proteins were highly enriched, with the possible exception of glutathione S-transferase, and it also seems more likely that these proteins were present in the ice fraction as a result of non-specific binding.

6.3 Discussion

Data presented in Chapter 3 showed that *P. superbus* has high freezing tolerance and its protein extracts have the ability to control the growth of ice. AFPs have this property and through their ice-binding capability, AFPs are able to influence the freezing survival of several fish, insects, plants and microbes (Ewart et al., 1999). The diversity of AFPs is great and there is no effective and reliable way of identifying an AFP from amino acid sequence information alone. This has made the isolation and purification of AFPs very difficult. The ice-affinity purification method utilised in this Chapter takes advantage of the one thing that defines and an AFP, the ability to bind to ice.

Based on the measurements provided by Kuiper *et al.* (2003) the IAP cold finger was constructed from brass at NUIM. The finger needs to be able to maintain a predetermined sub zero temperature and sufficient pressure is needed to circulate the coolant through the finger. This is necessary to ensure that when the finger is placed in icy water the entire finger is cool enough to become nucleated and maintain a slow growing layer of ice around the finger. The length of the cold finger needed to be adjusted several times before an optimal size was achieved that would nucleate and grow an ice sphere.

A recombinant AFP was kindly provided by the Prof. Peter Davies for use as a positive control to confirm that the apparatus was working and for optimisation of the

ice affinity binding protocol. This 7.0 kDa protein, HPLC 12 (also referred to as QAE) is a globular Type III AFP was originally purified from the ocean pout, *Macrozoarces americanus.* The growing ice fraction had a growth pattern that indicated the presence of an IBP. When AFPs are present in the starting solution, they will adsorb to the ice surface and at temperatures below the non-colligative freezing points, and become trapped by the growing ice front. They will shape the region around where they are trapped making apparent groves in the ice (see Figure 6.2). The IAP method semi-purified the AFP from the crude mixture of *E. coli* proteins. SDS-PAGE data comparing the ice fraction to the starting and liquid fractions clearly showed enrichment of the AFP in the ice fractions, but there were also several contaminating proteins. With a second round of IAP the concentration of contaminating proteins were reduced and the AFP fraction was further enriched. Further rounds of IAP would likely improve the purity but this level of purification was sufficient to show that the apparatus worked. The LC-MS/MS analysis confirmed the identity of the enriched band as the Type III AFP, HPLC 12.

The IAP method did not purify IBP from *P. superbus* as effectively as hoped. The growing ice fraction did appear to be shaping the ice to some degree but this was not as extensive as that seen in the positive control. This suggested that the concentration of potential IBPs in the *P. superbus* protein extract is quite low or that *P. superbus* IBPs have a low affinity for the ice. The *P. superbus* nemtaodes used to extract protein for ice-binding purification where grown at normal conditions (20 ^{o}C). A period of cold-acclimation at low non-freezing temperatures may have increased the concentration of potential ice-binding proteins in the starting extract.

SDS-PAGE data showed that there are some protein bands present that were in the starting fraction that appear to be less concentrated in the liquid fraction but are seen in the ice fraction. This experiment was done twice and similar bands had this characteristic in both cases. The SDS-PAGE protein profiles also showed that there was non-specific binding of proteins to the ice. Ideally the rate of freezing would have been regulated using a programmable water bath. The sudden although small manual reductions in the temperature of the finger probably led to rapid ice growth trapping many non-specific proteins in the ice. Optimisation specifically to purify IBP from the *P. superbus* protein extract was not possible due to the amount of time required to

obtain sufficient protein for this experiment (1 mg/ml in a total volume of 80 ml). Two independent rounds of this experiment required approximately 40 ml of compact nematode pellets, this took 3 months to collect. With larger quantities of protein the starting, liquid and ice fractions would have been optimised for separated by 2D-PAGE, followed by quantitative image analysis. This may have given a clearer identification of proteins enriched in the ice fraction.

The protein bands that appear to be enriched in the ice fraction compared to the liquid fraction were analysed by LC-MS/MS. This analysis did not reveal any proteins that had homology to existing antifreeze proteins in the *P. superbus* database (PST) or any of any proteins in the NCBI database. There were a large number of proteins found in four of the five bands that were investigated; band 2 peptides had no sequence similarity to existing proteins. Peptides found in band 1 had sequence similarity to three proteins in the PST database. Paramyosin is likely a contaminating protein and not an ice-binding protein. The other two proteins are vitellogenins. Vitellogenin is a yolk protein that is synthesised and secreted in the intestine during the egg-laying stage. In *C. elegans* adult stages vitellogenin has been showed to have antioxidant properties by binding to toxic metals (Nakamura et al., 1999) and has also been shown to have a role in immunity, it can recognise pathogen-associated molecular patterns in fish (Zi et al., 2008).

There were two proteins found in band 3, Vitellogenin and UDP-glucuronosyltransferase. Each protein was only identified using one peptide, this reduces the confidence that these proteins were actually present in these bands. UDP-glucuronosyltransferases are responsible for the enzymatic addition of sugars to fat-soluble molecules, both endogenous and xenobiotic. This is an important process that increases the solubility of these compounds in water and aids in their excretion. The substrates for UGTs are numerous, including diverse dietary constituents such as fatty acids and steroids (McElwee et al., 2004) so it is possible that UDP-glucuronosyltransferase could have evolved ice-binding capabilities.

In bands 4 and 5 there were a large number of proteins, the majority of which were also only identified by one peptide match. There were several muscle-associated proteins, ribosomal proteins and enzymes that are involved in different metabolic

pathways present in the excised bands. Since these proteins are involved in protein or protein ligand interactions it is possible that any one of these proteins could potentially be an evolutionary precursor of an AFP. However none of these proteins were highly enriched and it is more likely that these proteins were present in the ice fraction as non-specific binding. In band 5 there is also a sigma non-opioid intracellular receptor. This is a transmembrane chaperone protein that modulates calcium signalling. This protein has an affinity for several different compounds (Su and Hayashi, 2003). It may be possible that this protein also has affinity for ice.

In conclusion it is unknown whether any of the proteins that were identified are ice-binding proteins. These proteins may simply be hydrophilic or have some affinity for ice but may not have any influence in controlling the growth of ice. However the evolutionary precursors for the antifreeze proteins that have been identified so far are very diverse and some are very unlikely candidates for ice-binding, thermal hysteresis or recrystallisation inhibition activity. There are the C-type lectins (Ewart and Fletcher, 1993), sialic acid synthase (Baardsnes and Davies, 2001), apoliproteins (Deng et al., 1997), trypsinogen (Chen et al., 1997b), pathogenesis related proteins (Hon et al., 1995) and polygalacturonase inhibitors (Worrall *et al.*, 1998; Meyer *et al.*, 1999). Therefore it is entirely possible that some of the proteins listed in Table 6.2 are AFPs. There is also the possibility that the molecules that are shaping the ice may be interfering with the ice formation in some manner preventing its further growth without actually binding to the ice (Wharton, Personal Communication).

Although all previously described antifreeze and thermal hysteresis (TH) producing biomolecules are proteins, Walters *et al.* (2009b) isolated a non-protein TH-producing biomolecule active from the freeze-tolerant beetle, *Upis ceramboides*, by means of ice affinity. Amino acid analysis, polyacrylamide gel electrophoresis and NMR spectroscopy indicated that the THF contained little or no protein, yet it produced 3.7 ^{o}C of TH at 5 mg/ml, comparable to that of the most active insect antifreeze proteins. Compositional and structural analyses indicated that this antifreeze contains a disaccharide core comprising repeated units of β-mannopyranosyl-(1→4) β-xylopyranose and a fatty acid component, although the lipid may not be covalently linked to the saccharide. Walters *et al.* (2011) subsequently reported the isolation of

xylomannan-based antifreeze glycolipids from a plant (*Solanum dulcamara*) and 6 additional insect species. Thus it is possible that the ice structuring activity of *P. superbus* extracts described in Chapter 3 might be produced by a non-protein biomolecule. Recombinant proteins would need to be produced to test the ice-binding activity in the identified proteins. Because of the large number of candidate AFPs, the high risk of negative results and the large amount of experimental time required making and testing recombinant proteins, these downstream experiments were not attempted.

Chapter VII RNA-seq analysis of *P. superbus* freezing survival

7.1 Introduction

In this Chapter the ability of *P. superbus* to survive low temperatures is further investigated. Chapter 3 showed that *P. superbus* has improved freezing survival after a period of cold-acclimation (Section 3.2.1). This may occur due to the expression of genes that render the nematode more resistant to the damaging effects of freezing. The genes that are differentially expressed in response to cold-shock and cold acclimation were determined using the next generation sequencing method RNA-seq.

Over the past few years there has been a shift from the application of the "first generation" Sanger sequencing method towards newer high throughput parallel sequencing methods that are referred to as next-generation sequencing (NGS) (Metzker, 2010). The major commercial pioneers of the NGS platforms were Roche (454) (Margulies et al., 2005) and Illumina (Bentley et al., 2008). The major advantage of NGS is that it offers the ability to produce large amounts of DNA sequence data in a short period of time at relatively low cost. The throughput varies from 1,00,000 single reads with the Roche FLX Titanium XL+ system to 3 billion read with the Illumina HiSeq2000 in a single run. The read lengths also vary, the Roche FLX Titanium XL+ system generates up to 1,000 bp while the Illumina platforms give 100-150 bp. These sequencers can also perform "paired end sequencing", where the template can be sequenced in the forward and reverse direction, doubling the amount of data produced. NGS has revolutionised the approaches undertaken to answer many biological questions (Metzker, 2010). The most widely used application is the sequencing and resequencing of genomes. NGS has also facilitated previously unimaginable projects such as the 1000 Genomes Project, which aims to discover, genotype and provide accurate haplotype information on all forms of human DNA polymorphism in multiple human populations (The 1000 Genomes Consortium, 2010). A service to sequence personal genomes is now available from Illumina for less than $10,000 (www.everygenome.com).

In addition to genomics, NGS has also changed transcriptomics through the development of RNA-sequencing (RNA-seq) (Nagalakshmi et al., 2008). The

transcriptome comprises all of the transcripts that are in a cell and is specific for a developmental stage or physiological condition. An understanding of the transcriptome is essential in the understanding of gene regulation in development and in response to disease. Until the advent of NGS technologies, hybridisation and sequence-based approaches were used in transcriptome analysis. Hybridisation-based approaches include the microarray platforms (Lockhart et al., 1996). Microarray hybridisation is high-throughput and relatively inexpensive but it has several disadvantages. Microarray hybridisations suffer from background and cross-hybridisation noise and can only allow measurement of the abundance of RNA transcripts whose corresponding sequences have been spotted onto the array, so novel sequences cannot be analysed (Zakharkin *et al.*, 2005; Okoniewski and Miller, 2006). Novel sequences can be investigated using tiling arrays but these are expensive and suffer from problems with binding affinity so reliable probes can be difficult to design (Bradford et al., 2010). With microarrays, comparing expression levels across different experiments is difficult requiring complicated normalisation protocols. They are also unable to detect subtle differences in gene expression.

Prior to NGS the sequence-based approaches to transcriptomics were through direct sequencing of cDNA libraries (Adhikari et al., 2009). Tag-based methods were also developed and these include serial analysis of gene expression (SAGE) (Velculescu et al., 1995), cap analysis of gene expression (CAGE) (Shiraki et al., 2003), and massively parallel signature sequencing (MPSS) (Brenner et al., 2000). These high throughput methods can provide precise gene expression data, but they are all based on the expensive Sanger sequencing technology. Additionally only a portion of the transcriptome is analysed and isoforms are generally indistinguishable.

The NGS RNA-seq approach to transcriptomics has the power to analyse and compare expression levels, differential splicing, allele-specific expression, RNA editing and fusion transcripts under a variety of conditions (Costa *et al.*, 2010b; Marguerat and Bahler, 2010; Ozsolak and Milos, 2011). It has several advantages over microarrays and sequencing of cDNA libraries. In comparison with microarrays RNA-seq is not limited to the detection of known transcripts, it can allow the identification, characterisation and quantification of new splice isoforms and for the correct gene annotation to be defined to a single nucleotide resolution. RNA-seq also

has very low background signal when compared to microarrays and has a larger dynamic range over which transcripts can be detected. RNA-seq also shows high level of reproducibility for both technical and biological replicates (Marioni et al., 2008b).

The method by which each the NGS platforms produces RNA-seq data are different but they all have similar work flows for the preparation of libraries. Briefly, the RNA sample is initially fragmented to reduce possible secondary structures and to obtain a size that is compatible with the sequencers. The RNA is converted into double stranded cDNA and ligated to unique adaptors for sequencing and amplification. The adaptors will allow the cDNA fragments to be singled out, either on beads or on a slide to be sequenced in parallel.

The NGS platforms each use different sequencing chemistry and methodological procedures for RNA-seq. The sequencing method employed by the Illumina platform (Costa et al., 2010a) will be described, as it is was used in this study. The prepared cDNA library is placed on one of eight lanes of a flow-cell (slide). Individual fragments of cDNA attach onto the surface of the lane and undergo an amplification step where they are converted into clusters of double stranded cDNA. The flow-cell is placed in the sequencing machine and each cluster is sequenced in parallel. Illumina uses the Cyclic Reversible Termination (CRT) method of sequencing, this means that it uses reversible terminators in a cyclic manner. A DNA polymerase bound to the primed template adds one fluorescently modified nucleotide per cycle. The remaining unincorporated nucleotides are washed away and image capture is performed. A cleavage step occurs before the next cycle to remove the terminating group and the fluorescent dye, this is followed by another washing step. For each flow-cell this process is repeated for a given number of cycles. The fluorescence intensities are converted into base calls. The number of cycles the sequencer is capable of determines the length of the reads and the number of clusters determines the number of reads. With this sequencing method substitutions are the most common error type and there may also be an underrepresentation of AT-rich and GC-rich regions (Dohm et al., 2008).

When the reads have been obtained from the sequencer they are generally assessed for quality before they are mapped to a reference genome. Recently, with the ability to do paired-end sequencing there is sufficient read coverage for *de novo* transcriptome assembly as well as for gene expression and polymorphism analysis (Crawford *et al.*, 2010; Grabherr *et al.*, 2011). There are several programs for mapping reads to the reference genome or transcriptome, including Bowtie, SOAP2, MAQ, BWA, Mosaik, Novoalign and ZOOM (Trapnell and Salzberg, 2009). In this study Bowtie was used to perform the alignments (Langmead et al., 2009).

A *P. superbus* transcriptome constructed from high-throughput Roche 454 Titanium paired end reads was used as a reference in this project (by Georgina O'Mahony). RNA-seq data was aligned to this reference using Bowtie. NGS sequencers can produce billions of reads in one run, this makes mapping the reads to a reference using traditional alignment algorithms very computationally expensive. Bowtie uses a computational strategy known as "indexing" to speed up the mapping process. Like the index in a book, an index of the large reference sequences allows shorter sequences to be found within it rapidly. This approach involves the use of a technique known as the Burrows-Wheeler transform that was originally developed for compressing large files (Burrows and Wheeler, 1994). This memory-efficient data structure allows Bowtie to scan reads in a very fast and efficient manner on a standard computer. Following the alignment there are three types of reads, those that map uniquely to the reference, those that map to multiple regions in the reference genome or transcriptome and those that do not map to the reference. Reads that map to multiple locations are typically removed, as the region they were sequenced from cannot be determined. The use of paired-end protocols reduces this problem. Short reads that are identical to each other are also removed as they are thought to be PCR artefacts. The problem with the removal of PCR artefacts is that abundant identical reads could be a genuine reflection of abundantly expressed RNA species.

In order to calculate gene expression and compare gene expression across samples the reads that align to each gene must be counted and normalised to adjust for varying lane sequencing depths. The read count has been found to be approximately linearly related to the abundance of the target transcript (Mortazavi et al., 2008). In this analysis the reads were counted with Python package HTSeq-count (www-

huber.embl.de/users/anders/HTSeq/doc/overview.html). This software simply counts how many reads from aligned sequencing reads map to each feature from a given annotation. This raw count data was normalised by the R package DEseq (Anders and Huber, 2010). The program Cufflinks may also be used to count and normalise the data (Trapnell et al., 2010). This method normalises read length and total read number based on reads per kilobase of exon model per million mapped reads (RPKM) (Mortazavi et al., 2008). However with this method a few very strongly and differentially expressed reads could make up a large proportion of the total reads and lead to the RPKM number been skewed (Bullard et al., 2010). The DEseq package estimates a size factor from the count data and transforms the data. When the abundance is estimated from raw counts the same gene from multiple samples can be compared assuming that the statistical model that will be used can take the variation in sample abundance into account. DEseq estimates these size factors (*sj*) by taking the median of the ratios of observed counts. Each size factor estimate is calculated as the median of the ratios of each samples counts to those of a reference sample. The reference (*f_j*) is the geometric mean over all samples and counts. To calculate the size factor for a sample (*j*), the counts (*k_ij*) for the genes (*i*) are divided by the reference values (*f_j*), and use the median of all those ratios (*k_ij/f_i*) as the size factor (Anders and Huber, 2010).

Following normalisation of the read count (abundance) by size factors, the DEseq package was used to estimate differential expression (DE) (Anders and Huber, 2010). When using only technical replicates RNA-seq data has been shown to follow a Poisson distribution (Marioni et al., 2008a), however this distribution has been found to predict smaller variations between replicates than what is actually seen in the data (Anders and Huber, 2010). DEseq and a similar R package edgeR (Robinson et al., 2010) model count data and test for DE with the related negative binomial distributions. The core assumption of the DEseq method is, like the Poisson distribution, that the mean is a good predicator of the variance (that genes with similar expression levels also have similar variances across replicates). This assumption is necessary as typically a RNA-seq experiment may have only two or three replicates; this is too low to reliably estimate the variance for each gene separately. DEseq estimates a function for each condition that allows the variance to be predicted from the mean. This is done by calculating the sample mean for each gene and the variance

between the replicates and then fitting a curve to this data. The package DEseq has various tests with visual outputs that analyse whether the variance function estimates are accurate and are dependent on the mean. As the DEseq normalises from raw counts, it can also estimate the level of noise in the data. The "shot noise" is the noise that almost always present, even with the best possible replication, and this can be modeled by a Poisson distribution. Shot noise is a term, which was originally coined to describe the electronic noise arising in an electric current because of the number of electrons flowing through a circuit is sufficiently small that fluctuations occur due to the Poisson distribution. It is dependent on the size factor and tends to be the dominant noise for the weakly expressed genes. This noise can only be reduced by deeper sequencing. In addition to shot noise there may also be technical and biological noise in the date. Technical noise is introduced in sample preparation and sequencing and ideally should not be present. This can be overcome by removing the affected samples from the dataset. Biological noise is the noise present in samples that cannot be account for by technical noise. It is the dominant noise for the most strongly expressed genes and may only be resolved by doing more biological replicates (Anders and Huber, 2010).

7.2 Results

7.2.1 Selection of conditions for RNA-seq

Three conditions were selected for RNA-seq, a control (nematodes grown at 20 ^{o}C) and two treatments, a 4 ^{o}C cold-shock for 24 h and 10 ^{o}C acclimation for 10 days. Quantitative PCR (qPCR) was used to estimate the time point that would yield the highest level of differential gene expression for the cold-shocked nematodes. Four putative cold-responsive genes were tested, *hsp-90*, *tps* (trehalose-6-phosphate synthase), *lea-3* (late embryogenesis abundant) and *pyk* (pyruvate kinase). The nematodes were placed on plates at 4 ^{o}C for 12, 24, 36 or 48 h before washing, pelleting and freezing at -80 ^{o}C under Trizol. Each sample was thawed; RNA was extracted and converted into cDNA before analysis by qPCR (as described in Sections 2.2.2.15). Exposing the nematodes to 4 ^{o}C for 24 h resulted in the highest mean relative expression (fold change) for all four genes (Figure 7.1). Two-way ANOVA (P<0.001) and *post-hoc* testing between the crossing points of the control and each

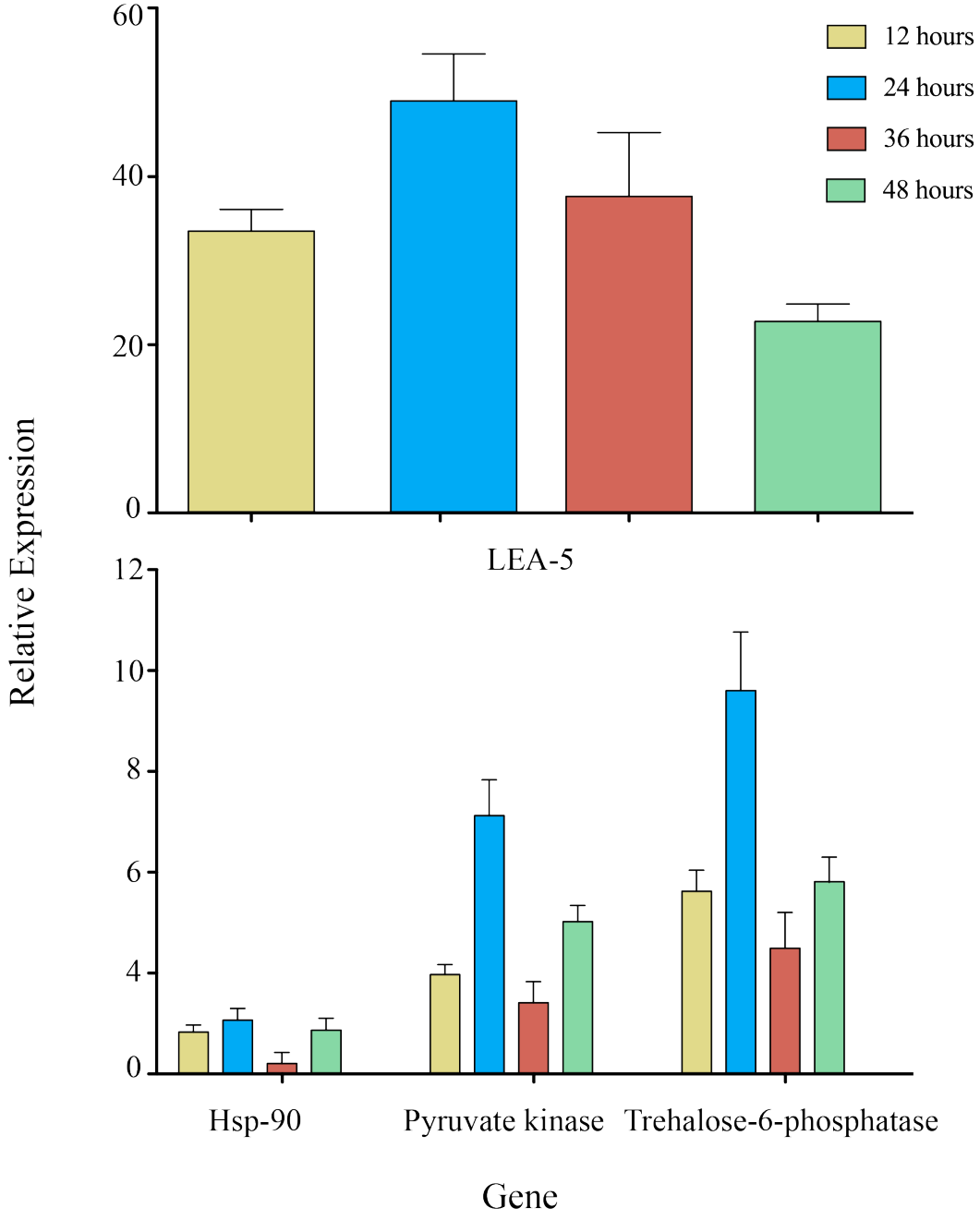


Figure 7.1. Relative expression of cold-responsive genes over time.

time point also found the 24 hour time point to be the most significant. This cold-shock treatment was therefore selected for the RNA-seq experiments. A cold acclimation regime was 10 ^{o}C for 10 days as it had previously been found to give the highest survival freezing values for *P. superbus* when compared to the control (Section 3.2.1).

7.2.2 RNA extraction

The RNA was extracted as described in Section 2.2.2.2. Following extraction the RNA concentration was calculated using the Qubit® fluorometer (Invitrogen). The integrity of the RNA was assessed using a Bioanalyzer (Agilent) (Figure 7.2). The Bioanalyzer software assigns an RNA integrity number (RIN) value of 1 to 10, where a value of 10 indicates completely intact RNA (Schroeder et al., 2006). Samples used to construct a RNA-seq library required of 7.0 or more. Out of the nine samples used for our analysis, they were 6.9 or above, this was considered high enough for use. These RNA samples were brought to a final volume of 50 μl, each sample contained more than 10 μg of RNA. The TrinSeq laboratory at Trinity College Dublin performed library preparation and sequencing for RNA-seq.

7.2.3 Quality control analysis

The program Fast QC (www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) was used to perform quality control on the raw sequence data from the Illumina sequencing runs. The graphs displaying the quality scores across all bases (Figure 7.3(i)) and the quality score distribution (7.3 (ii)) for one of the samples are presented as all samples gave similar results. The Phred quality score (Ewing et al., 1998) across the read including the error was within 30 and 40 across all bases, this means the probability of calling an incorrect base was very low (between 1 in 1000 and 1 in 10000). The quality of the calls does reduce slightly as the run progresses. This is common across most sequencing platforms but the quality score is still within the good quality range (Figure 7.3 (i), green zone). The skewed distribution of the quality scores towards high Phred scores further demonstrated that all of the reads are of high quality (Figure 7.3 (ii))

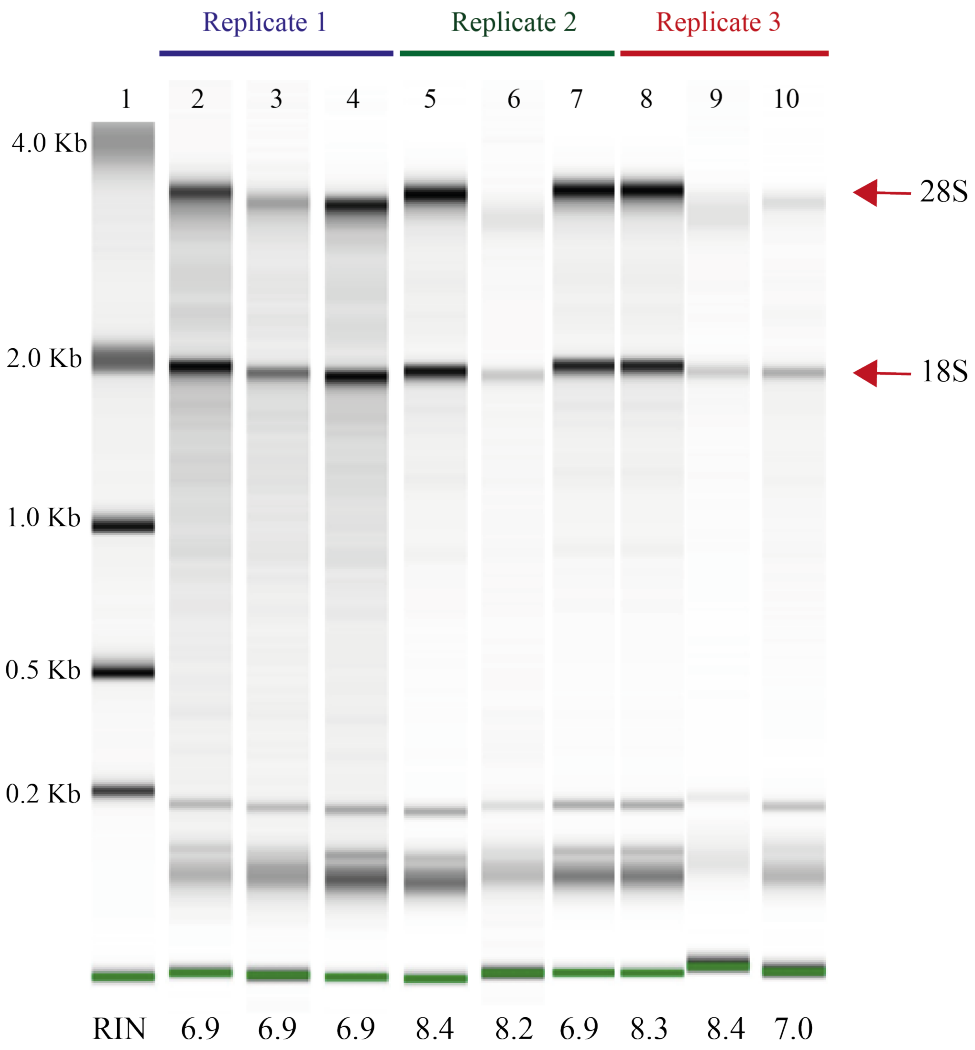
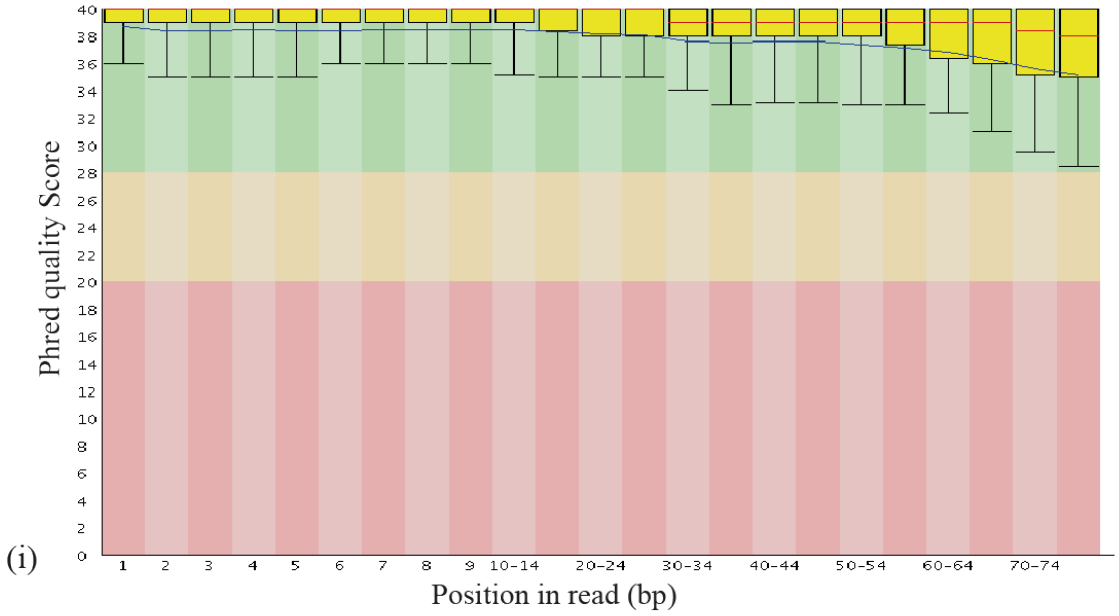
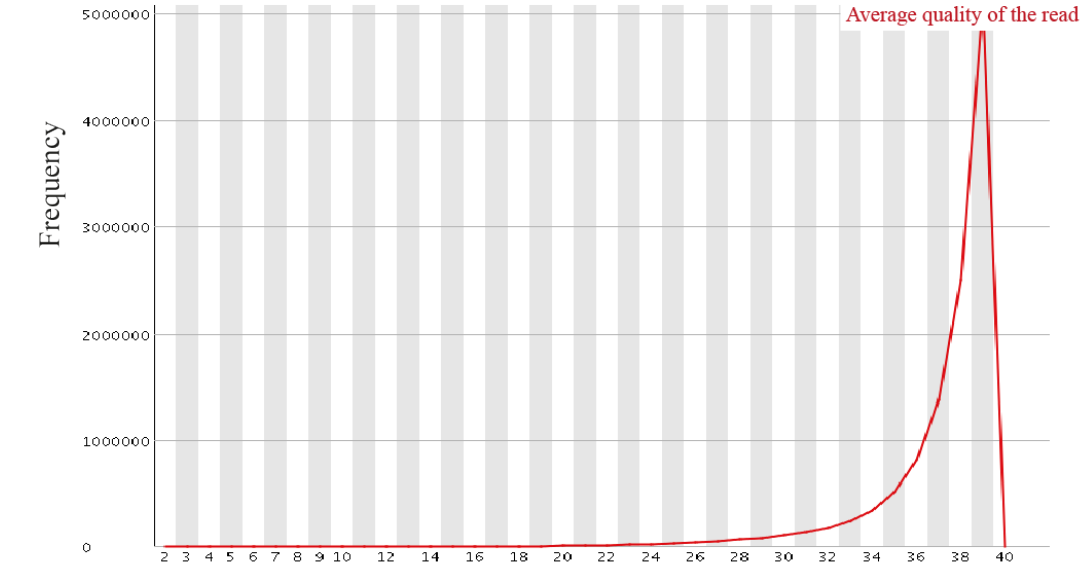


Figure 7.2. Digital image and RNA integrity numbers of the samples used for RNA-seq experiments. The location of the bands corresponding to the 18S and 28S are indicated in the image. The layout of the image is follows: Lane1, RNA 6000 ladder (Agilent); Lane 2, Control replicate 1; Lane 3, 4 ^{o}C cold-shock replicate 1; Lane 4, 10 ^{o}C acclimation replicate 1; Lane 5, Control replicate 2; Lane 6, 4 ^{o}C cold-shock replicate 2; Lane 7, 10 ^{o}C acclimation replicate 2; Lane 8, Control replicate 3; Lane 9, 4 ^{o}C cold-shock replicate 3; Lane 10, 10 ^{o}C acclimation replicate 3.



(i)



(ii)

Figure 7.3. FastQC analysis on quality of RNA-seq data. (i) Quality scores calculated across all bases for each position in the reads. The whiskers represent the 1st lower quartile, and the 4th upper quartile (not visible here because of the tight distribution at the top of the graph) and the yellow box is the range of the 2nd and 3rd quartile. The line in the middle is the mean. (ii) Distribution of the quality score over all sequences.

Across each position in a read the percentages of each of the four bases remains constant (Figure 7.4). Although the quality at the start of the read is high there were spikes in the percentage of the bases content across the first 10 positions in the read. This analysis also revealed that the *P. superbus* transcriptome is AT rich, 60% of the bases are either A or T bases compared to 40% C or G bases found in the reads.

7.2.4 Alignment of RNA-seq reads to the *P. superbus* reference transcriptome

A *P. superbus* transcriptome was used as a reference in this project. It was sequenced on a Roche 454 Titanium high-throughput sequencer from a paired end cDNA library containing an equal concentration mixture from nematodes that had been subjected to no stress (20 ^{o}C/ 24 hrs/ shaking), exposed to oxidative stress (38 µM paraquat /24hrs/ shaking/ 20 ^{o}C), heat stress (32 ^{o}C/ 24 hrs/ shaking), cold stress (4 ^{o}C/ 24 hrs/ shaking) or desiccation stress (98% 60 hrs/ 20 ^{o}C). The transcriptome assembly from the Roche 454 Titanium reads was generated using CAP3 to create a hybrid assembly of Newbler and MIRA assemblies. This CAP3 assembly was then further improved by putting it through the CLoBB assembly program and the CLoBB improved by putting it through the Phrap assembly program.

Indexing of the *P. superbus* reference transcriptome sequences and alignment of the RNA-seq reads was done using Bowtie (Langmead et al., 2009). Table 7.1 shows the number of reads that were obtained for each of the conditions from the sequencer. Of those reads Bowtie was able align approximately 80% to the reference *P. superbus* transcriptome. Other reads were removed during pre-processing steps before the alignment by Bowtie. The number of reads was also reduced by the removal of reads that mapped to multiple regions in the *P. superbus* reference transcriptome. PCR duplicates, were removed using the SAMtools program (Li et al., 2009).

7.2.5 Counting, normalisation and statistical analysis

The number of reads that mapped to a given annotated gene in the *P. superbus* transcriptome was counted by the program HTSeq-count. The table of counts from each replicate was combined into one master table that contained all of the counts, conditions and replicates. Normalisation and differential expression analysis were carried out using the R package DEseq (Anders and Huber, 2010). After normalisation multidimensional scaling (MDS) analysis was carried out using the

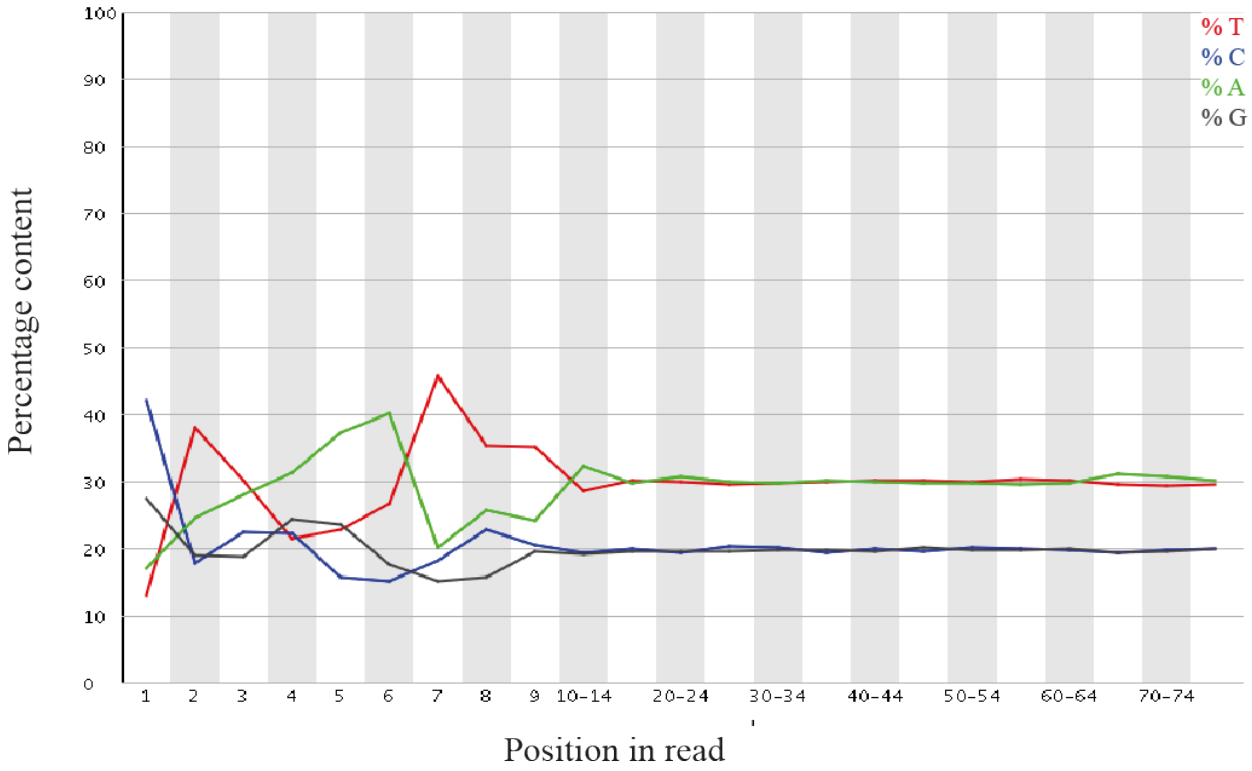


Figure 7.4. FastQC analysis on base composition in RNA-seq data. Bases are represented by the colours red (T), blue (C), green (A) and black (G).

Table 7.1. Number of *P. superbus* reads obtained for each experimental condition, the number of reads removed during the pre-processing and alignment steps and the final number of reads retained for statistical analysis.

Experimental condition	Replicate	Number of reads from sequencer	Reads removed by pre-processing (%)	Reads that did not align (%)	Reads that aligned multiple times (%)	Final number of reads
Control	1	11767545	0.012	20.360	0.168	9350395
	2	7321746	0.018	20.663	0.048	5804013
	3	11312015	1.081	19.570	0.038	8971605
4 ^{o}C Cold-Shock	1	12585530	0.035	21.590	0.027	9860459
	2	11275039	0.017	20.216	0.086	8984052
	3	6327444	1.069	18.325	0.051	5097051
10 ^{o}C Acclimation	1	9897958	0.034	19.535	0.021	7958914
	2	9392154	0.019	18.146	0.030	7683268
	3	5539187	0.019	16.622	0.030	4615705

edgeR package (Robinson et al., 2010) to give an overview of differences between treatments and replicates (Figure 7.5). This Figure shows that all the biological replicates for treatment 2 (4 ^{o}C cold shock) cluster together and are clearly distinct from all the other treatments. Replicates 2 and 3 for treatment 1 (control) and treatment 3 (10 ^{o}C acclimation) also form distinct treatment specific clusters. However, treatment 1 replicate 1 and treatment 3 replicate 1 appear to form a separate cluster. Because it was still possible to separate all the replicates in treatment 1 from all of the replicates comprising treatment 3 in the MDS plot, treatment 1 replicate 1 and treatment 3 replicate 1 were not treated as outliers and retained for statistical analysis. The data was also visualised as a heatmap showing the distances between samples as calculated from the variance-stabilising transformation of the count data (Figure 7.6). However these analyses show that treatment 3 replicate 1 (10 ^{o}C acclimation) is clustering within the treatment 1 (control) biological replicates. If time had permitted, the statistical analysis would have been repeated following removal of the treatment 3 replicate 1 dataset.

DEseq estimates a function for each condition and allows the variance to be predicated from the mean by calculating the mean of each treatment for each gene and the variance between each of the replicates then fitting a curve to this data. To check whether this fit was good and to visualise the level of noise in the data a squared coefficient of variance (SCV) plot was produced (Figure 7.7). The solid lines are the SCV for the raw variance (variance calculated from mean before being scaled up using the size factor). There is one calculated solid line per treatment. The difference between these lines is due to biological noise. The dotted lines show all conditions and their replicates. This is the calculated base variance (the full variance scaled up by size factors). The distances between each of the replicates is due to technical noise. The difference between these dotted lines and the solid lines is shot noise. It can be seen that the shot noise is greatest for the weakly expressed genes and for any genes with expression strengths above 100 the shot noise disappears, the raw variance is equal to the base variance. The biological noise is highest for the strongly expressed genes, but it would be expected that the difference between the control, cold-shock and acclimation read counts would be most apparent for those that are very strongly expressed. The solid black line is the density estimate of the base means. There is

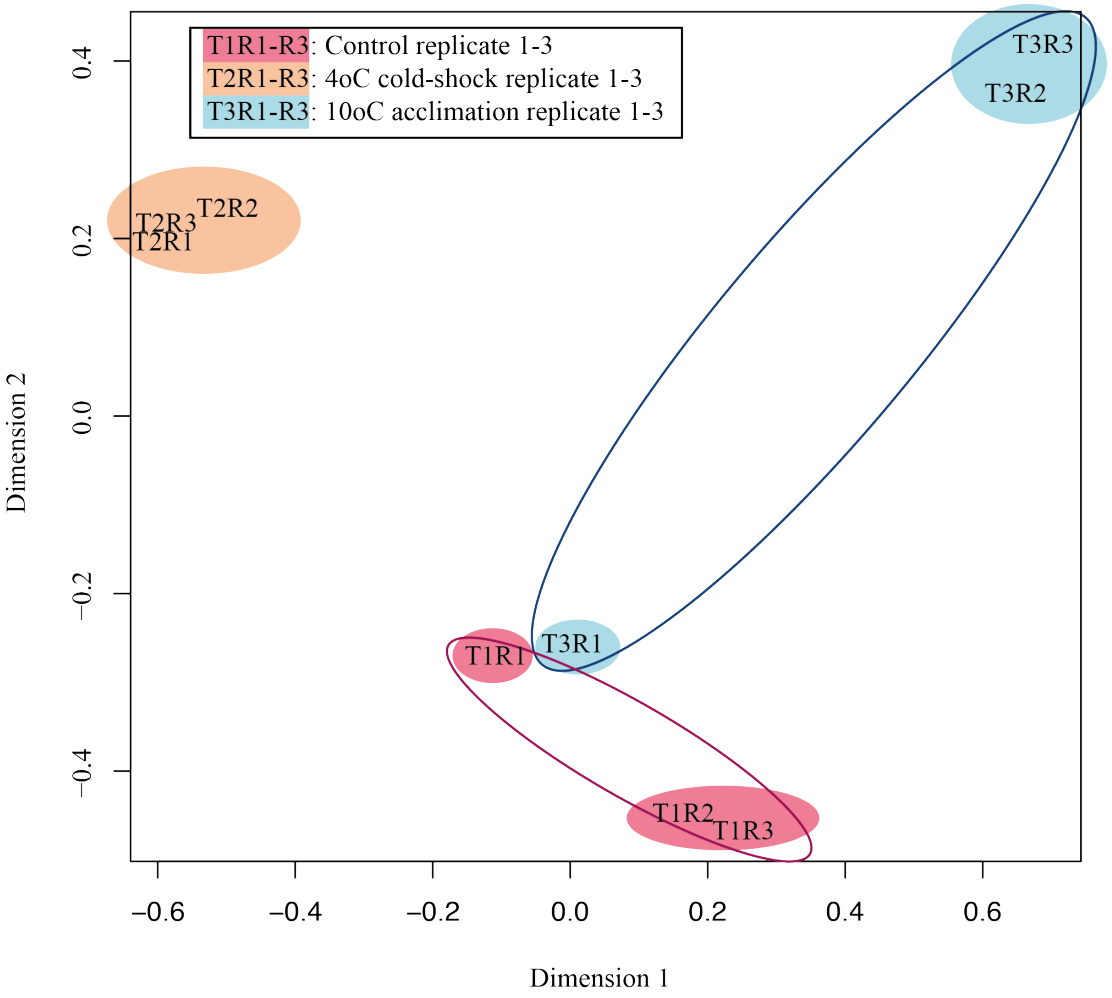


Figure 7.5. Multidimensional scaling (MDS) plot for the *P. superbus* RNA-seq count data showing the relations between the samples in two dimensions. The three treatments are indicated in pink (control), orange (4°C cold-shock for 24 h) and blue (10°C acclimation for 10 days).

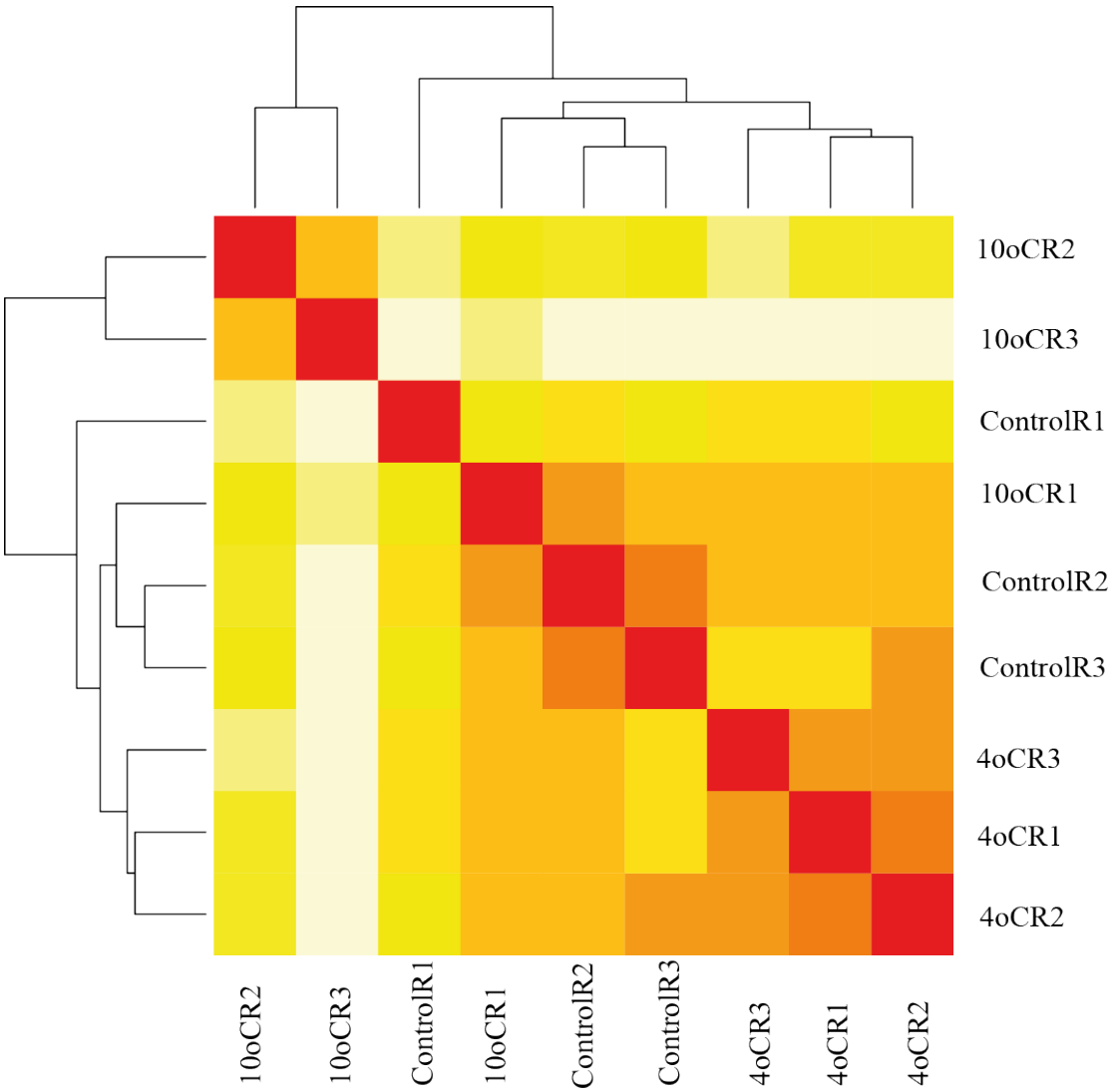


Figure 7.6. A heatmap showing the distances between *P. superbus* RNA-seq the samples as calculated from the variance-stabilising transformation of the count data. The heatmap shows a representation of the Euclidean distance matrix and the dendrogram represents a hierarchical clustering of the treatment replicates.

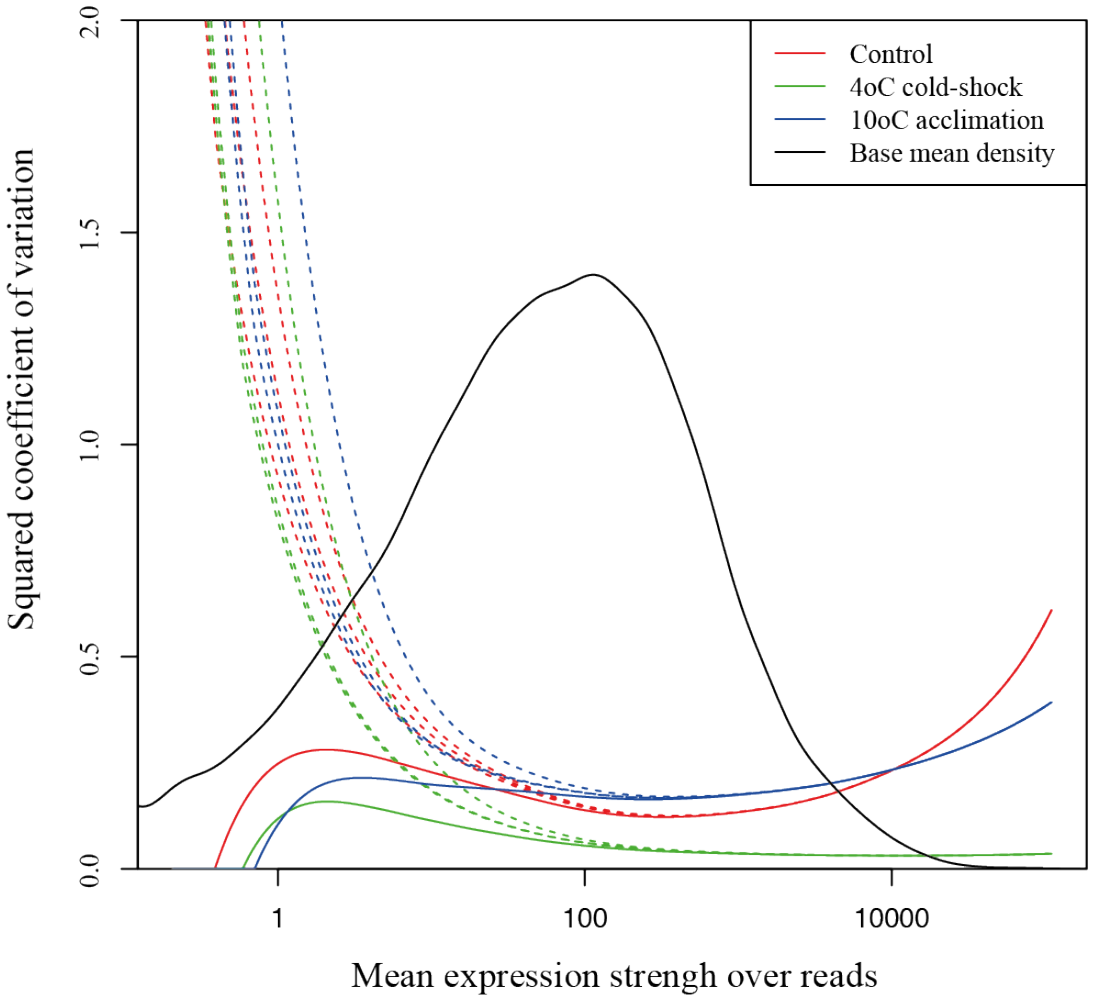


Figure 7.7. Plot showing the estimated variances for the *P. superbus* RNA-seq count data. The solid coloured lines are the squared coefficient of variation (SCV) of the raw variance for each treatment. The dotted line is the base variance, (full variance scaled down to base level by the size factors). The vertical distance between the solid and dotted lines is the shot noise.

sufficient density of counts in this dataset to expect that the estimation of the variance is accurate.

The relationship between the estimated base variance and the base mean for each graph was also visualized by plotting the values and drawing a line with the fit from the local regression (Figure 7.8). This showed that for the control, cold-shock and acclimation that base variance functions fit the count data well, the base mean is proportional to the base variance.

Following the verification of the variance-mean dependence of the RNA-seq counts for all treatments, the differential expression for each gene between the control and 4 ^{o}C cold-shock and control and 10 ^{o}C acclimation treatments was calculated using the negative binomial test in the DEseq package as described in Section 2.2.6.11. The proportion of genes differentially expressed was visualised by plotting the fold change against the mean expression level (Figure 7.9). The power to detect differential expression depends on overall counts. The results presented in Figure 7.9 show that in both treatments several genes are differentially expressed when compared to the control.

7.2.6 Differentially expressed genes

The list of the *P. superbus* transcripts and the corresponding statistical data were exported from DEseq into an Excel file. A stringent cut off of a 2-fold change in expression and a statistical adjusted *P*-value of no more than 0.01 was applied to the dataset to ensure a minimal level of false positives. There were a total of 913 genes differentially expressed in response to a 4 ^{o}C cold-shock and 815 genes differentially expressed in the 10 ^{o}C acclimation. The main focus of this study was the identification of genes up-regulated in response to these cold stresses. There were 589 genes significantly up-regulated in the cold-shock dataset and 724 genes in the acclimation dataset. These transcripts were identified from a database prepared by Georgina O'Mahony Zamora at NUIM, which contains the UniRef (www.uniprot.org) identities of the genes in the *P. superbus* reference transcriptome. A subset of these identified genes is listed in Table 7.2 and Table 7.3. Many of the up-regulated genes were novel and did not have a match in the UniRef database. There were 322 novels identified out of 589 up-regulated genes (54.66%) found in the cold-

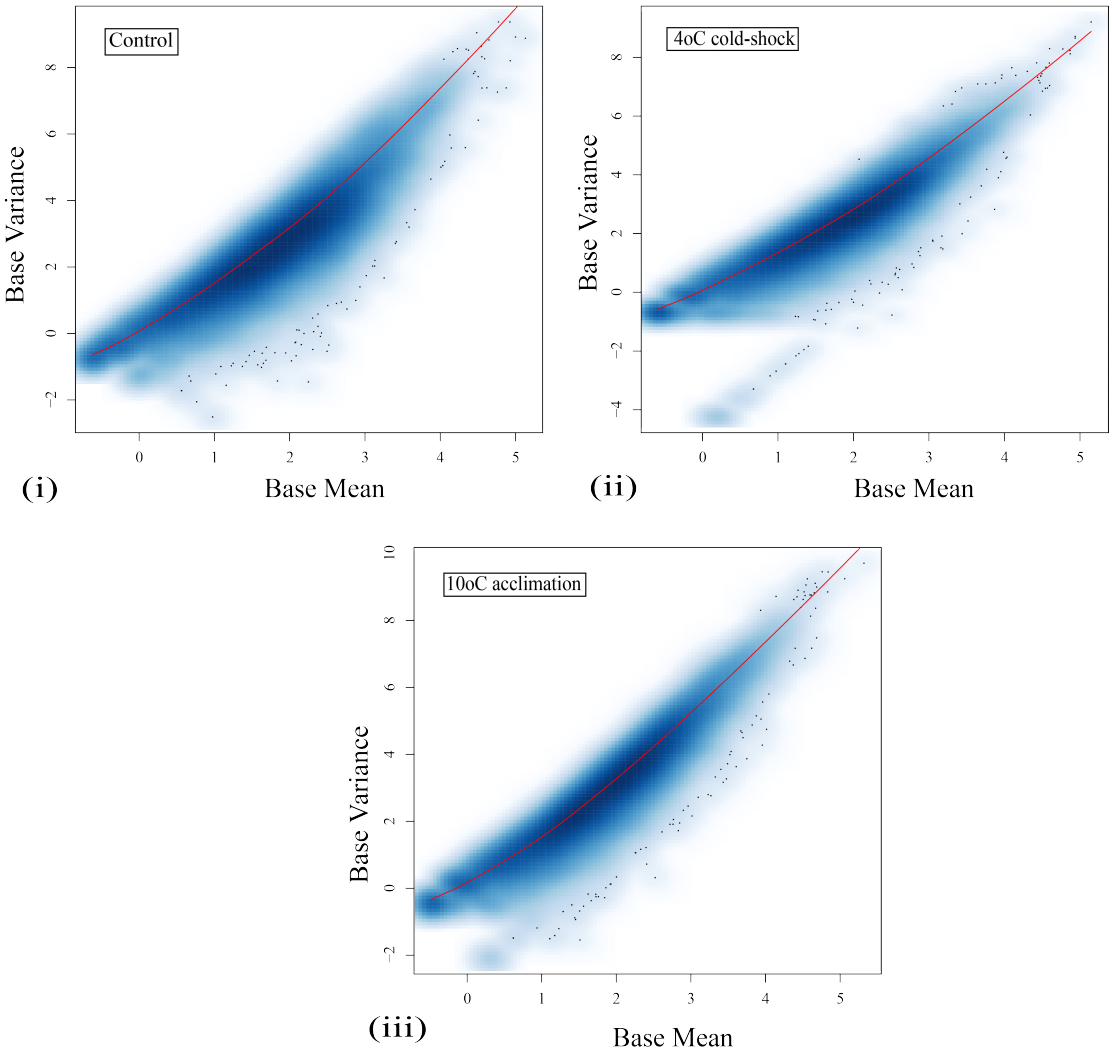
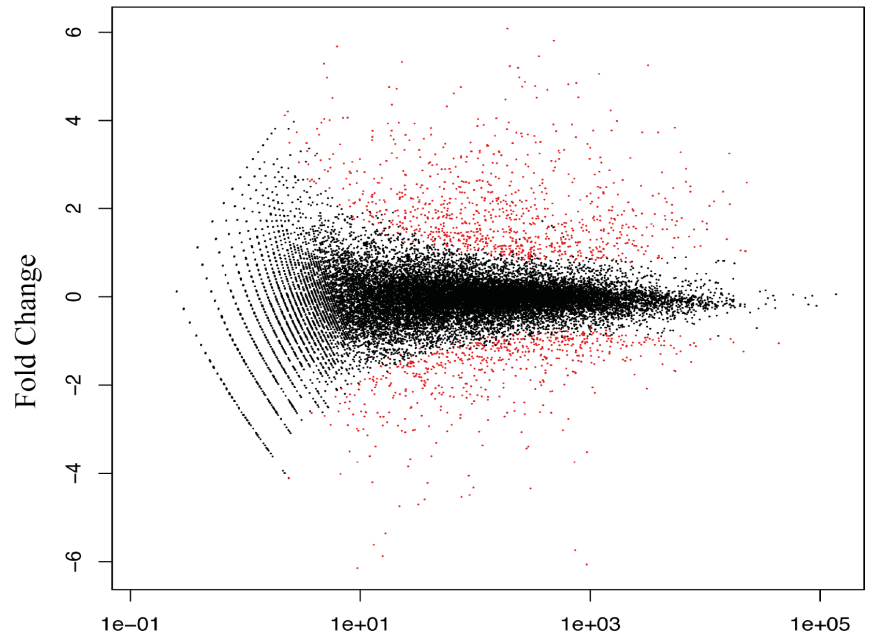
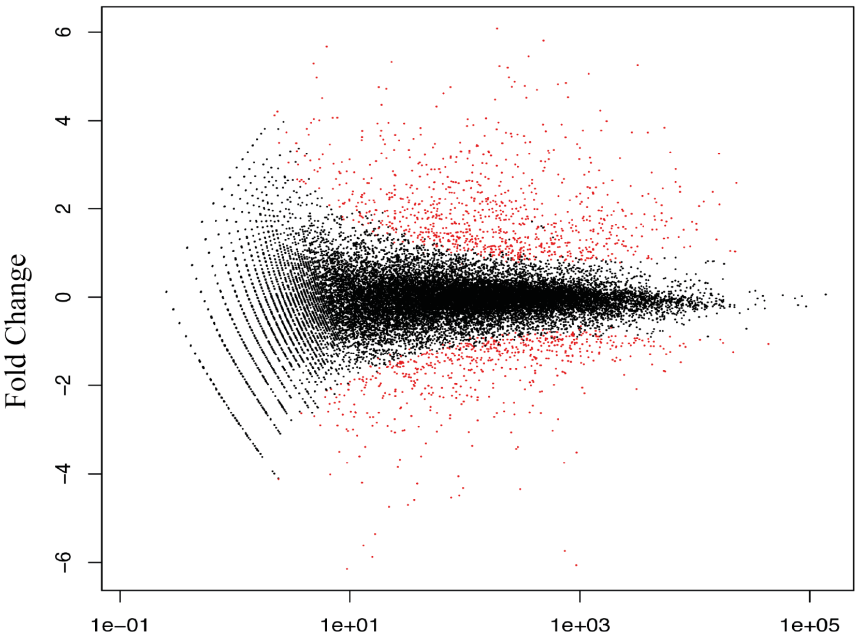


Figure 7.8. Diagnostic plot to investigate the dependence of the variance on the mean for each treatment in the *P. superbus* RNA-seq data in (i) control, (ii) 4 ^{o}C cold-shock for 24 h and (iii) and 10 ^{o}C acclimation for 10 days. The solid line shows the local regression fit for the data. The x and y-axis are in logarithmic scale (log10).



(i)



(ii)

Figure 7.9. Testing for differential expression between the control and 4 ^{o}C cold-shock for 24 h (i) and control and 10 ^{o}C acclimation for 10 days (ii) with a scatter plot of the fold change versus mean. The red colour marks genes detected as differentially expressed at a 10% false discovery rate, when Benjamini-Hochberg multiple testing adjustment is used. The x-axis scale is logarithmic and y-axis is the fold-change difference logged to the base 2.

Table 7.2. Identification of a selection of the up-regulated transcripts in the 4 ^{o}C cold-shock dataset.

Role	Transcript identification	UniRef ID number	*P. superbus* transcript ID	Fold-change	Adjusted *P*-value
Metabolism	Class 3 lipase protein	UniRef100_E0YDN7	PST08091	38.14	4.74×10^{-22}
	L-isoleucine-4-hydroxylase	UniRef100_E2GIN1	PST17852	14.94	4.64×10^{-14}
	Trehalase	UniRef100_F1KU03	PST22475	9.56	1.96×10^{-5}
	Proline dehydrogenase	UniRef100_O45228	PST15082	3.53	4.46×10^{-6}
	UTP-glucose-1-phosphate uridylyltransferase	UniRef100_E1FH65	PST01919	2.54	1.65×10^{-3}
	Phosphoglycerate mutase family protein	UniRef100_A8PAY9	PST03728	4.00	6.90×10^{-6}
Detoxification	Glutathione S-transferase	UniRef100_Q9NAW7	PST09179	9.86	6.89 x 10-^{10}
	Alcohol dehydrogenase	UniRef100_F1LB68	PST01983	9.29	4.85×10^{-12}
	Glutaredoxin	UniRef100_Q7YUB8	PST23995	8.70	2.52×10^{-14}
	Lysozyme protein	UniRef100_Q09975	PST04279	8.42	5.49×10^{-13}
	Cathepsin L	UniRef100_Q86MN0	PST22276	3.82	6.47×10^{-6}
	Cytochrome P450	UniRef100_Q27499	PST10598	5.95	4.59×10^{-3}
	Catalase	UniRef100_B2XQY0	PST20067	5.65	3.79×10^{-8}
	Late embryogenesis abundant protein	UniRef100_UPI0000221018	PST28289	3.19	2.65×10^{-3}
	Cadmium-inducible lysosomal protein	UniRef100_Q18831	PST00142	2.94	9.99×10^{-5}
Transporting	ABC transporter family protein	UniRef100_Q9TXV8	PST15703	9.96	1.43×10^{-10}
	Metal transporter	UniRef100_F1KV71	PST28166	10.44	3.69×10^{-15}

Table 7.2. Identification of a selection of the up-regulated transcripts in the 4 °C cold-shock dataset.

Role	Transcript identification	UniRef ID number	*P. superbus* transcript ID	Fold-change	Adjusted *P*-value
Signalling	Ribosomal-S kinase-5	UniRef100_E1FI35	PST17445	2.84	2.71×10^{-4}
	Extended synaptotagmin-2	UniRef100_O76835	PST27899	10.08	3.06×10^{-7}
	Mitogen-activated protein kinase kinase kinase	UniRef100_E0VIQ1	PST27196	6.17	1.23×10^{-7}
	Calmodulin-dependent proteins	UniRef100_E5S347	PST22740	4.45	1.55×10^{-8}
Cytoskeleton remodelling	Actin	UniRef100_F1A8F6	PST25435	4.46	3.36×10^{-6}
	Endo-1,3(4)-beta-glucanase	UniRef100_F2SUP1	PST26536	3.41	9.21×10^{-4}
Immunity	Surface-associated antigen	UniRef100_Q1W204	PST12834	29.39	2.20×10^{-11}
	Chitinase	UniRef100_F1L1B2	PST05051	2.77	4.93×10^{-4}
Transcription	Zinc finger protein	UniRef100_F1KXU1	PST22796	8.19	8.90×10^{-10}
	Ribosomal RNA methyltransferase	UniRef100_A8P9A3	PST07843	9.69	3.21×10^{-11}
Reproduction	S-adenosylmethionine decarboxylase	UniRef100_F1L3Z4	PST03343	3.32	9.69×10^{-6}
Lipid bilayer remodelling	Annexin	UniRef100_E1G226	PST28510	3.03	1.08×10^{-4}
Miscellaneous	Calexcitin	UniRef100_A8WUE6	PST15173	3.03	3.48×10^{-15}
	Osmotic avoidance abnormal protein 11	UniRef100_A8XNG4	PST17726	13.28	2.84×10^{-15}
	Glutamic acid-rich protein	UniRef100_P13816	PST20782	9.48	1.60×10^{-14}
	C-type lectin protein	UniRef100_E3N8J	PST00453	37.50	3.79×10^{-23}
	Transthyretin-like protein	UniRef100_A8P382	PST18048	3.57	2.68×10^{-5}
	Beta-lactamase domain-containing protein	UniRef100_F1L324	PST16626	6.14	8.42×10^{-7}

Table 7.3. Identification of a selection of the up-regulated transcripts in the 10 °C acclimation dataset.

Role	Transcript identification	UniRef ID number	*P. superbus* transcript ID	Fold-change	Adjusted *P*-value
Metabolism	Lipase	UniRef100_E0YDN6	PST26425	12.27	1.79×10^{-7}
	L-isoleucine-4-hydroxylase	UniRef100_E2GIN1	PST17852	12.57	5.08×10^{-8}
	Phosphoglycerate mutase family protein	UniRef100_A8PAY9	PST03728	5.35	1.96×10^{-4}
Detoxification	Glutathione S-transferase	UniRef100_E0VL93	PST05681	10.63	5.79×10^{-7}
	Alcohol dehydrogenase	UniRef100_Q9P6C8	PST17738	10.25	7.30×10^{-3}
	Glutaredoxin	UniRef100_A8Y1E0	PST23995	5.68	6.83×10^{-5}
	Lysozyme protein	UniRef100_Q19698	PST12030	4.64	1.42×10^{-3}
	Cathepsin L	UniRef100_E3TGS2	PST18644	27.98	1.43×10^{-10}
	Cytochrome P450	UniRef100_O16482	PST19109	9.94	1.56×10^{-4}
	Hydrolase	UniRef100_D3BU18	PST22268	24.21	3.30×10^{-13}
	Late embryogenesis abundant protein	UniRef100_UPI0000221018	PST28289	5.00	1.73×10^{-3}
	Cadmium-inducible lysosomal protein	UniRef100_A8Y3T4	PST03745	15.85	1.50×10^{-7}
Transporting	ABC transporter family protein	UniRef100_A8Q5J5	PST22785	8.33	4.87×10^{-4}
	Metal transporter	UniRef100_F1KV71	PST28166	4.43	2.20×10^{-3}
	Aquaporin	UniRef100_F1L4B4	PST19364	27.03	7.08×10^{-15}
Signalling	Extended synaptotagmin-2	UniRef100_F1KSQ5	PST27899	6.29	2.76×10^{-3}
	Mitogen-activated protein kinase kinase kinase	UniRef100_E0VIQ1	PST27196	13.03	3.87×10^{-9}

Table 7.3. Identification of a selection of the up-regulated transcripts in the 10 ^{o}C acclimation dataset.

Role	Transcript identification	UniRef ID number	*P. superbus* transcript ID	Fold-change	Adjusted *P*-value
Cytoskeleton remodelling	Actin	UniRef100_C4M3T9	PST02651	4.74	1.31×10^{-3}
	Endo-1,3(4)-beta-glucanase	UniRef100_F2S9W9	PST26536	27.43	1.39×10^{-4}
Immunity	Surface-associated antigen	UniRef100_Q1W204	PST12834	157.29	3.02×10^{-25}
	Chitinase	UniRef100_B7QD78	PST09417	10.17	8.74×10^{-8}
	Thaumatin	UniRef100_A7XZK5	PST13458	8.72	1.93×10^{-5}
	Interferon-inducible GTPase	UniRef100_UPI0000F2038B	PST09795	11.02	1.07×10^{-7}
Transcription	Zinc finger protein	UniRef100_Q9U239	PST07043	5.64	5.68×10^{-4}
	Ribosomal RNA methyltransferase	UniRef100_A8P9A3	PST07843	7.50	3.03×10^{-5}
DNA repair	Poly(ADP-ribose) polymerase	UniRef100_F1KUX9	PST00992	13.81	4.58×10^{-9}
Reproduction	Zonadhesin	UniRef100_F1KUR2	PST27363	20.88	3.01×10^{-6}
Lipid bilayer remodelling	Fatty acid desaturase	UniRef100_E3MST0	PST25293	5.64	7.66×10^{-3}
Miscellaneous	Calexcitin	UniRef100_Q10131	PST00301	4.91	1.39 a 10^{-3}
	Osmotic avoidance abnormal protein 11	UniRef100_E3MAM1	PST25352	6.10	3.78×10^{-4}
	Glutamic acid-rich protein	UniRef100_Q9GTW3	PST20782	3.78	9.82×10^{-3}
	C-type lectin protein	UniRef100_E3M6P3	PST25476	38.91	5.22×10^{-17}
	Transthyretin-like protein	UniRef100_F1LFY5	PST00710	17.34	1.66×10^{-10}
	Expansin	UniRef100_F1DTC5	PST01758	24.42	2.39×10^{-13}
	Neuropeptide-like protein protein	UniRef100_Q22739	PST07179	9.73	3.42×10^{-7}
	Guanine deaminase	UniRef100_F1LFY0	PST08813	10.90	5.75×10^{-4}
	Beta-lactamase domain-containing protein	UniRef100_A8WN12	PST16626	9.55	1.62×10^{-6}

shock dataset and 406 novels out of 724 up-regulated genes identified in the acclimation dataset (56.08%). The most highly up-regulated genes in both datasets tended to be novel sequences. Several of the highly up-regulated genes also gave hits to hypothetical proteins. Only the genes that were up-regulated and found to have an identity were further analysed (Appendix 9.1 and 2.2).

The results presented in Tables 7.2 and 7.3 show that cold shock and cold acclimation induce the expression of genes whose products have roles in a great diversity of pathways and processes. These include transcription, signalling, metabolism, transport, detoxification, DNA repair, lipid bilayer remodelling, cytoskeleton remodelling and immunity. Both datasets had up-regulated transcripts associated with the oxidative stress response. Catalase, glutathione peroxidase and glutathione-S transferase (GST) were up-regulated after a 4 ^{o}C cold-shock dataset, while peroxiredoxin, metallothioneins and superoxide dismutase (SOD) were up-regulated in the acclimation dataset only. Both datasets also contained a member of the multi-functional late embryogenesis abundant protein (LEA) family. ABC, glucose and metal transporters were up-regulated in both datasets and the 10 ^{o}C acclimation dataset also had up-regulation of ATPase and the integral membrane protein aquaporin. Only one transcript was identified as a fatty acid desaturase in the 10 ^{o}C acclimation dataset, these enzymes are essential enzymes for maintenance of membrane fluidity (Nishida and Murata, 1996). The datasets contained several immunity related genes including pathogenesis related proteins (PR) such as glucanases, chitinase and thaumatin. Other differentially expressed genes included zinc fingers, signalling molecules including calexcitin, calmodulin-dependent proteins or mitogen activated kinase (MAPK) kinases as well as enzymes involved in metabolism. The full list of significantly up-regulated genes may be found on the Appendix CD files 9.1 (4 ^{o}C cold-shock dataset) and 9.2 (10 ^{o}C acclimation dataset).

7.2.7 Statistically differentially expressed pathways

The *P. superbus* transcriptome was mapped to the Kyoto Encyclopaedia of Genes and Genomes (KEGG) identifications for *C. elegans*. Those transcripts that had a *C. elegans* KEGG identification were tested for the presence of overrepresented pathways using the R package, GOseq (Young et al., 2010). Five pathways were found to be differentially expressed in the 4 ^{o}C cold-shock dataset; fatty acid

metabolism, tryptophan metabolism, alanine, aspartate and glutamate metabolism, arginine and proline metabolism and nitrogen metabolism. Seven pathways were differentially expressed in the 10 ^{o}C acclimation dataset; metabolism of xenobiotics by cytochrome P450, drug metabolism by cytochrome P450, fatty acid metabolism, starch and sucrose metabolism, ubiquione and other terpenoid-quinone biosynthesis, fatty acid biosynthesis and hedgehog signalling pathways.

The GOseq analysis found that in the 4 ^{o}C cold-shock dataset the transcripts associated with the enzymes in nitrogen and alanine, asparate and glutamate metabolism were all down-regulated (nitrogen metabolism enzymes: glutaminase, nitrilase and carbonic anhydrase family member, alanine, asparate and glutamate metabolism: 4-aminobutyrate aminotransferase, glutaminase, amidophosphoribosyl transferase). In this dataset a transcript that is associated with one enzyme is down-regulated (nitrilase family member) and two are up-regulated (glutaminase and proline hydroxylase). The arginine and proline metabolism pathways had two enzymes down-regulated (glutaminase and proline hydroxylase) and two up-regulated (proline dehydrogenase and S-adenosylmethionine decarboxylase). Fatty acid metabolism pathway had three enzymes down-regulated (acyl-CoA oxidase, acyl-CoA dehydrogenase-3 and alcohol dehydrogenase) and two enzymes up-regulated (a fatty acid CoA synthase member and cytochrome P450).

In the 10 ^{o}C acclimation dataset fatty acid metabolism was also found to be affected. Similarly to the 4 ^{o}C dataset the transcript associated with cytochrome P450 is up-regulated, but in contrast a fatty acid CoA synthase family member is downreguated while alcohol dehydrogenase is up-regulated. Fatty acid biosythesis is also down-regulated as a major enzyme in the pathway, fatty acid synthase family member (*fasn-1*) is down-regulated. The drug metabolism and metabolism of xenobiotics pathways both have up-regulation of transcripts associated with glutathionine S-transferease, alcohol dehdrogenase and hydrolase. In the ubiquinone and other terpenoid-quinone biosynthesis pathway the enzyme 4-coumarate CoA ligase is down-regulated and ubiquinone biosythesis monooxygenase (coq6) is up-regulated. In the hedgehog signalling pathway the membrane protein megalin is down-regulated glycogen synthase kinase β is up-regulated.

7.2.8 QPCR to confirm RNA-seq

QPCR of four genes was performed on control cDNA, 4 ^{o}C (cold-shock) and 10 ^{o}C (acclimation) (Figure 7.10). For the 4 ^{o}C treatment the genes *ctl* (catalase), *rsk-5* (ribosomal protein s6 kinase), *nex* (annexin) and *smd* (S-adenosylmethionine decarboxylase) were selected. For the 10 ^{o}C sample the genes *lipase*, *gst* (glutathioine S-transferase), *thn* (thaumatin) and a *lec* (lectin) were selected. QPCR found all of these genes to be significantly up-regulated (ANOVA $P<0.001$) in agreement with the RNA-seq results (Figure 7.10). The fold-change in expression was not the same for both experiments. This is expected since the RNA sample used in this experiment was not the same as that used in the sequencing run.

7.3 Discussion

The development of the RNA-seq technique has revolutionised the field of transcriptomics. It has the power to detect the slightest differences in gene expression levels between different cell types, growth stages or environmental conditions. Its high sensitivity for the detection of differential splicing and single nucleotide polymorphisms is useful in understanding how altered splicing patterns contribute to development, cell differentiation and human disease (Ozsolak and Milos, 2011).

In was shown in Chapter 3 that *P. superbus* has increased ability to survive freezing following a period of cold acclimation. RNA-seq was selected to study the gene expression response that occurs during this time. It is possible that during exposure to a period of low non-lethal temperature before freezing these nematodes may produce proteins that improve their freezing capabilities. A pilot qPCR study showed that there was some level of up-regulation in *P. superbus* in response to a cold-shock and the amount of time required to give optimal expression of the genes tested was determined. The four genes used were *hsp-90*, *pyk*, *tps* and *lea-3*. These are all genes that have been shown to be up-regulated in response to stress. Heat shock proteins (Hsp) including Hsp-90 have been shown to be significantly up-regulated in response to cold stress as well as desiccation and anoxia (Parsell and Lindquist, 1993; Wu, 1995; Feder and Hofmann, 1999). Pyruvate kinase is expected to be up-regulated during a cold stress since it is essential in glycolysis. The non-reducing disaccharide, trehalose accumulates in other desiccation tolerant organisms and is thought to protect

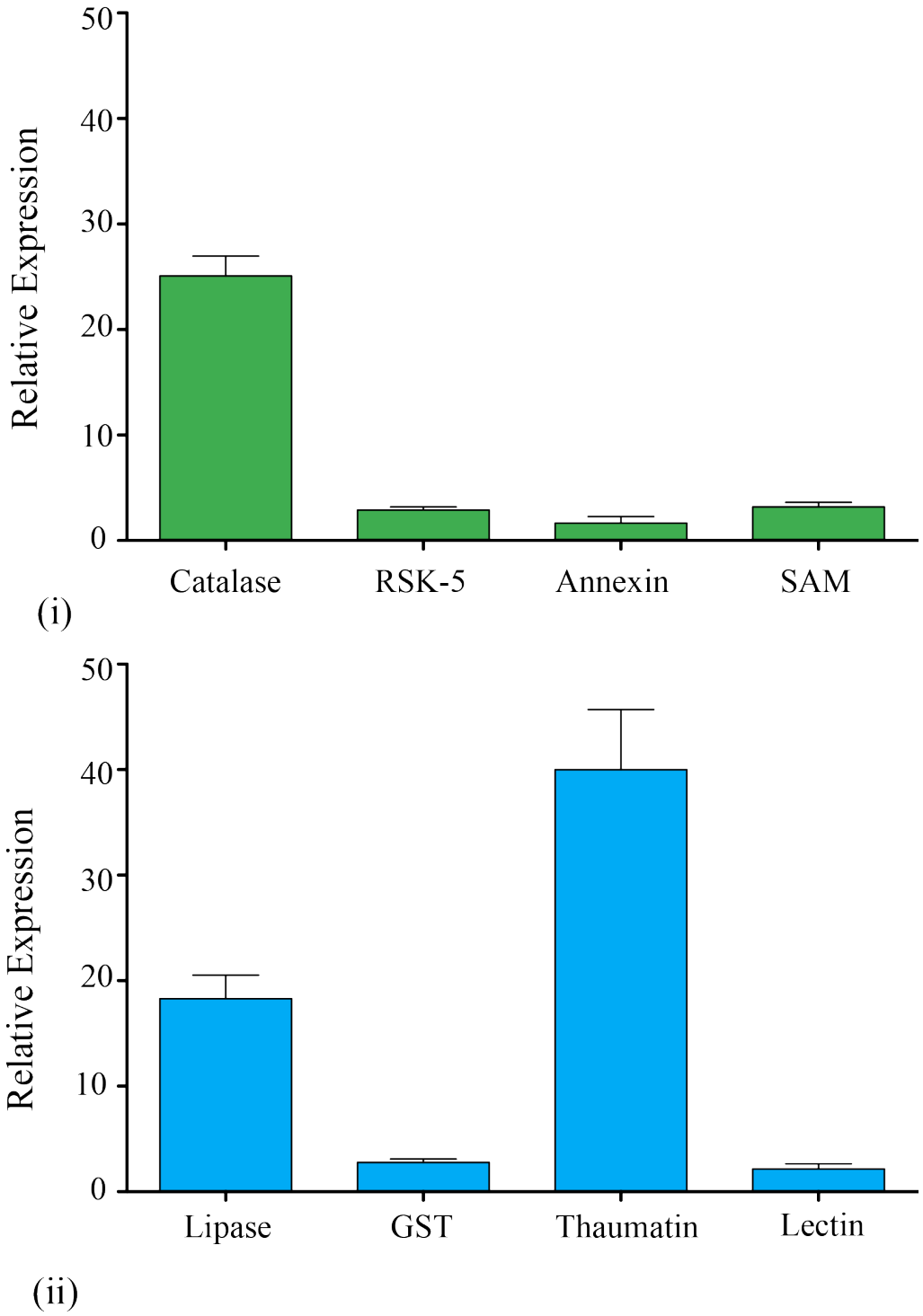


Figure 7.10. Relative expression as measured by qPCR of genes up-regulated in RNA-seq analysis of a 4 ^{o}C cold-shock (24 h) sample (i) and a 10 ^{o}C acclimation (10 days) sample (ii).

cell membranes by acting as a water replacement. Trehalose concentrations are also increased in several nematodes in response to cold (Wharton *et al.*, 2000; Jagdale and Grewal, 2003). Trehalose-6-phosphate synthase is an essential enzyme in the trehalose biosynthesis, so its up-regulation would also be expected. Late embryogenesis abundant (LEA) proteins expression is also found to be up-regulated in response to the water stress caused by desiccation, salt or low temperatures (Tunnacliffe and Wise, 2007). All four genes were up-regulated in 4 ^{o}C treated *P. superbus*, with the highest level after 24 h. In the genes *hsp-90*, *pyk* and *tps* gene expression appears to have increased at 48 h. The global expression levels of *P. superbus* acclimated for 10 days at 10 ^{o}C was also analysed by RNA-seq. This regime significantly improves freezing survival in many *Panagrolaimus* strains including *P. superbus* (see Section 3.2.2). The genes selected for the pilot qPCR were checked for up-regulation levels in the 4 ^{o}C RNA-seq dataset. The genes *hsp-90*, *pyk* and *tps* were found to be slightly up-regulated (1-1.25 fold-change). This level of up-regulation was not considered to be significant (*P adjusted* >0.01). The qPCR primers for *lea-3* gene were designed from a LEA sequence found in a *P. superbus* EST database (Tyson *et al.*, 2012). These sequences could not be found in the reference transcriptome sequence.

The RIN algorithm was used to calculate the integrity of the RNA samples that were used for RNA-seq. The RIN algorithm analyses the entire electrophoretic trace of an RNA sample, including the absence or presence of degradation products to determine sample integrity. This algorithm scores the RNA as having a RIN of between 1 and 10, where 10 is completely intact RNA. A high RIN value was essential to ensure that as many of the transcripts as possible were seen in the sequence data. All RNA samples used were had high RIN values as determined by the Bioanalyzer.

The RNA-seq raw data was found to be of excellent quality using the FastQC program (www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). The Phred quality scores were all high across the read. Phred quality scores are assessments of accuracy of the base-calling in the sequencing run. In the majority of our dataset the probability of calling an incorrect base was 1 in 10,000 reducing to 1 in 1000 towards the end of the read where the chemistry of the sequencer tends to start breaking down. This analysis also gave information on the base composition across the reads. The composition of

the bases is inconsistent for the first 10 bases as the run starts. This is common in Illumina sequencing and it probably represents a bias in the RNA fragmentation step during library prep, where fragmentation is occurring very slightly more often at a C or G than at a T or A. The base content across the remainder of the read it was constant. The GC content is 40%, in line with that found in several other nematodes (Mitreva et al., 2006). Approximately 80% of the reads could be aligned to the reference transcriptome, this is a relatively high number but it might have been improved by aligning to a *de novo* RNA-seq assembly from all of the reads (Grabherr et al., 2011). Preparing a *de novo* short read assembly from paired end Illumina reads is computationally intensive and was not attempted.

The R package DEseq allowed for the data to be normalised and to calculate the level of differential expression. This program compared the level of gene expression between the control and the cold-shock and acclimation treatments. Normalisation was performed on the datasets before any further analysis. This step is required since each sample will contain slightly different levels of cDNA and as a result will have different numbers of reads per gene. Following normalisation a MDS plot and a heat map demonstrated that the gene expression profiles for the three datasets were different. In both plots the cold-shock replicates all clustered together away from the others. Two of the acclimation replicates clustered together as did two of the control replicates. However the other replicate from the acclimation and cold-shock were clustered together. These replicates could have been removed but it was decided that they were still separated sufficiently for the calculation of differential expression. The estimated variance functions were shown to be related to the mean and to describe the data well, this allowed the negative binomial model to accurately estimate differential expression. This level of checking the variance mean relationship is required to as an inaccurate estimate of the relationship can lead to the calculation of false positive genes for differential expression.

The DEseq analysis found that there were 913 genes significantly differentially expressed in the cold-shock dataset and 815 in the acclimation dataset. It is possible that for both datasets the cut-off of a *P*-adjusted value of 0.01 may be too strict, a value of 0.05 may have been sufficient and shown a much high number of genes differentially expressed. The number of downregulated genes in the cold-shock

dataset corresponds to over 40% of the differentially expressed genes. Although only up-regulated genes were further analysed in this study, this finding suggests that down-regulation of particular processes may also have an importance in freezing tolerance. In both datasets novel genes were found (genes with no hits in any database), the cold-shock dataset had a total of 322 novels significantly up-regulated (54.66%) while the cold acclimated datset had 406 novels significantly up-regulated (56.08%). The majority of the top up-regulated genes were also novel sequences or had hits to hypothetical proteins. This very high proportion of novel sequences suggest that they may have very important roles in freezing tolerance and be the key genes that affect the freezing tolerance of *P. superbus*. Unfortunately due to the delay in receiving the RNA-seq data these novel sequences were not further investigated. If time had permitted the novel sequences would have been individually assessed for whether they could be possible ice-binding proteins. The amino acid sequence composition could have been compared to other antifreeze proteins, analysed for any potentially binding or secretion sequences and the hydrophobicity calculated. Potential novel ice-binding proteins would have been made recombinantly and tested using the ice-shaping activity assay (Section 2.2.5.8). In the following paragraphs those genes that were up-regulated and identified are discussed. The full list of significantly up-regulated genes may be found on the Appendix CD files 9.1 (4 ^{o}C cold-shock dataset) and 9.2 (10 ^{o}C acclimation dataset).

In this study, there were several genes found to be up-regulated in response to cold associated with the response to oxidative stress. Cold to freezing stress will lead to the accumulation of reactive oxygen species (ROS) (Abele and Puntarulo, 2004). ROS accumulation in cells leads to the oxidative modification of proteins, lipids, DNA and other macromolecules (Temple et al., 2005). During a period of cold-shock or acclimation *P. superbus* upregulates several genes that encode enzymes capable of quenching ROS. The *sod* (superoxide dismutase) transcripts up-regulated in the 10 ^{o}C acclimation dataset could encode a family of metalloenzymes responsible for quenching the potentially deleterious effects of superoxide radicals by converting them to hydrogen peroxide. Transcripts encoding the enzymes catalase, glutathione peroxidase and peroxiredoxin were also up-regulated. These enzymes deactivate hydrogen peroxide, another potent oxidising agent. Glutathione S-transferases (GSTs) are a diverse super-family of multifunctional proteins that play prominent roles in

detoxification metabolism in nematodes (Lindblom and Dodd, 2006). Transcripts encoding GST sequences were also strongly up-regulated in both cold treatment datasets. The lysosomes, together with the peroxisomes and the ubiquitin-proteasome system are responsible for the removal of oxidatively damaged proteins. The protease cathepsin L, which is primarily functional in the lysosome was also up-regulated in both datasets. *P. superbus* also upregulates a number of antioxidants in response to cold. The metallothioneins were up-regulated in the 10 ^{o}C acclimation dataset, these are small, cysteine-rich, metal binding proteins that can act as free radical scavengers (Kiningham and Kasarskis, 1998). In the 4 ^{o}C cold-shock dataset S-adenosylmethionine decarboxylase (*sam*) transcripts were up-regulated. SAM is an enzyme involved in the biosynthesis of spermine and spermidine, polyamines that may act as scavengers of free radicals and aldehydes. *C. elegans* also synthesises polyamines and requires them for normal growth and embryogenesis (MacRae et al., 1998). They are not known as being antioxidants in nematodes but the up-regulation of transcripts encoding this enzyme suggests a possible role.

An LEA protein was found to be up-regulated in the both of the datasets. LEA proteins were first associated with desiccation tolerance in plant seeds and resurrection plants (Tunnacliffe and Wise, 2007) but recently they have also been associated with desiccation tolerance in nematodes (Browne *et al.*, 2002; Browne *et al.*, 2004; Gal *et al.*, 2004; Adhikari *et al.*, 2009) and other invertebrates (Hand *et al.*, 2006; Kikawada *et al.*, 2006). In addition to desiccation these proteins have been found to increase in abundance in response to cold stress in the alga *Chlorella vulgaris* (Honjoh et al., 2000). LEA proteins are abundant in lysine and arginine residues that are readily oxidised and could potentially serve as a non-enzymatic antioxidant defence system in anhydrobiotic nematodes. Since the antioxidant enzymes are only functional where water is available they could potentially fill this role until the nematodes are rehydrated in the case of desiccation (Burnell and Tunnacliffe, 2011). These proteins have also been postulated to act as molecular shields, protecting proteins from negative interactions with other proteins. Other activities that they may be involved in include membrane protection, ion binding, hydration buffering and nucleic acid binding (Tunnacliffe and Wise, 2007). Since there is a strong association between freezing and desiccation tolerance it is likely that

during cold acclimation in *P. superbus* synthesis of this protein is up-regulated and may improve freezing survival.

During freezing, the osmotic and ionic balance set up by the exclusion of solutes from rapidly growing ice crystals requires a redistribution of water, ions, and cryoprotectants across cell membranes. In this study several up-regulated transcripts could encode proteins involved transport facilitation. Both datasets had ABC transporter proteins, glucose transporters and metal transporters. The 10 ^{o}C acclimation dataset also had up-regulation of aquaporin and ATP synthase. The up-regulation of ion and water transporters has been found to correlated with sensitivity to different types of stress in nematodes. Sixty ABC transporters have been identified and functionally characterised in *C. elegans* (Sheps et al., 2004). The members of this family are associated with resistance to heavy metals. The protein ATP synthase is likely up-regulated in an effort to maintain efficient homeostasis. The earthworm, *Dendrobaena octaedra* and the wood frog, *Rana sylvatica* both utilise glucose as their primary cryoprotectant (King *et al.*, 1995; Calderon *et al.*, 2009). In *R. sylvatica* there is an increased in numbers of glucose transporters in their membranes to deal with the rapid cryoprotectant movement (King et al., 1995). The up-regulation of glucose transporters may suggest a similar role in *P. superbus* or they may simply be required to increase the movement of glucose for use as an energy source. The aquaporins are integral membrane proteins that have been suggested to be involved in the cold-hardening response in insects (Clark and Worland, 2008). They control the flow of water and have additional functions in osmoregulation and metabolite transport, this would be very important for managing the flux changes that occur during freezing and melting.

When the cell membrane is below a critical temperature, known as the transition temperature, the membrane will become reorganised from the efficient, fluid and flexible liquid crystalline phase to the rigid solid gel phase. When membranes are in this state for long periods of time they undergo a loss of elasticity that leads to a loss in functional membrane proteins and membrane rupturing causing ion, water and metabolite leakage from the cells (Drobnis et al., 1993). Therefore during a period of cold acclimation it is essential to prevent for this phase transition by increasing the concentration of phospholipids with unsaturated fatty acids. Phospholipids with

unsaturated fatty acids are more flexible and have lower melting points. The desaturase enzymes play an essential role in this process and are expressed at low temperatures in fish (Tiku et al., 1996), bacteria (Sakamoto and Bryant, 1997), plants (Vega et al., 2004) and insects (Kayukawa et al., 2007). In the acclimation dataset this may be occurring, as the there was up-regulation of the transcripts encoding the fatty acid desaturase, FAT-6. Annexin was also found in the cold-shock dataset. Annexins are traditionally thought of as calcium-dependent phospholipid-binding proteins. The family has been linked with inhibition of phospholipase activity, exocytosis and endoctyosis, signal transduction, organisation of the extracellular matrix, resistance to reactive oxygen species and DNA replication (Braun et al., 1998). The role of annexins in response to low temperature is not clear but it is thought that they may in some way be involved in protecting the lipid bilayer from the damaging effects of freezing stress (Clark et al., 2010), so may improve *P. superbus* cold tolerance by modification of membranes.

Both datasets showed that C-type lectins were among the most highly expressed in response to low temperatures. The C-type lectins are a large protein family that is Ca^{2+} dependent glycan-binding protein that share primary and secondary structural homology in their carbohydrate-recognition domains (CRDs). They all have a C-type fold but not all of these proteins necessarily bind to carbohydrates. Many of these proteins function in mediating cell adhesion and signal transduction events following binding to carbohydrates (Dodd and Drickamer, 2001). Type II AFPs are homologs of these proteins and share the same characteristic fold (Ewart et al., 1992). Type II AFPs have adapted the carbohydrate binding ability of the lectins to bind to ice and prevents its further growth. It is possible that one of the C-type lectins up-regulated in response to cold is in fact an antifreeze protein. This would require further investigation of the protein sequences along with functional studies.

Both datasets also showed up-regulation of transcripts identified as pathogenesis related (PR) proteins such as glucanase, chitinase and thaumatin. These proteins are generally produced as a response to pathogen attack but also appear to be up-regulated in response to various stresses. An AFP from winter rye was found to have homology to PR proteins (Hon et al., 1995). It is possible that these proteins have dual roles in pathogen-induced and cold-induced stress responses. Therefore the PR genes

in both datasets may also be involved in cold-tolerance in addition to the pathogen response. Transcripts encoding lysozyme was found to be among the most highly up-regulated transcripts in both datasets. The main function of lysozymes appears to be defence against pathogens, although some of them have also been implicated in digestion. In the 4 ^{o}C dataset the lysozyme transcript had highest homology to *lys-8* from *C. elegans,* while in the 10 ^{o}C acclimation library the highest homology was to *C. elegans lys-5.* The *C. elegans* genome encodes 15 lysozyme genes (Schulenburg and Boehnisch, 2008). In *C. elegans lys-8* encodes a putative lysozyme required for normal longevity. LYS-8 expression is induced by infection with the pathogenic bacteria *Serratia marcescens* and *Pseudomonas aeruginosa* (Mallo et al., 2002).

Only a proportion of the transcripts identified as being differentially expressed are discussed here. There were also several others up-regulated in the datasets are likely important in the *P. superbus* cold response such as those genes involved in transcription, signalling, cytoskeleton remodelling, DNA repair, reproduction and metabolism. In both datasets there were a large number of novel sequences that had no hits in any database. There were also several genes that gave hits to hypothetical proteins. These proteins may be the key components of the freezing response in *P. superbus* and explain its superior freezing ability compared to some relatively closely related nematodes.

In Chapter 6 peptides recovered from enriched protein bands from *P. superbus* ice fractions were identified (Table 6.2). If these proteins were true ice-binding proteins and induced by cold stress it may be expected that they would be up-regulated in the RNA-seq datasets. The expression levels of the corresponding genes were checked in the 4 ^{o}C and 10 ^{o}C RNA-seq datasets but the did not show any significant levels of up-regulation (adjusted *P*-value <0.01 and fold-change >2). The majority of these genes were actually found to be down-regulated but not significantly.

The biochemical pathways that were found to be differentially expressed (up or down-regulated) in response to a period of acclimation were analysed using the R package GOseq (Young et al., 2010). This analysis was carried out on the *P. superbus* genes that could be mapped to the KEGG identities for *C. elegans* and revealed the pathways that were significantly differentially expressed (P<0.001). In the cold-shock

dataset five pathways were differentially expressed. Exposure to low temperatures is known to impair metabolic activity, in this dataset genes involved in nitrogen, alanine, asparate and glutamate metabolism were down-regulated. A transcript encoding an enzyme in the tryptophan biosynthesis pathway is down-regulated (nitrilase) while two other components of this pathway are up-regulated, catalase and cytochrome P450. These enzymes are likely to be up-regulated across many different metabolic pathways as a response to the oxidative stress and the toxins produced as a result of the cold-induced stress.

Many plants and microbial cells accumulate proline as a cryoprotectant (Takagi, 2008). In *P. superbus* this does not seem to occurred, the enzymes involved in proline production appear to be down-regulated (glutaminase and proline decarboxylase) while proline dehydrogenase is up-regulated. Proline dehydrogenase is involved in the rapid reduction of proline levels. The enzyme SAM decarboylase was also up-regulated, this may also be a general response to stress resulting in turn to increased polyamines spermine and spermidine. The pathway for fatty acid metabolism was significantly differentially expressed. A fatty acid CoA synthase enzyme (ASC-17) was up-regulated. This gene activates the breakdown of complex fatty acids. A subsequent genes in this pathway, acyl CoA oxidase and acyl CoA dehydrogenase were down-regulated so overall fatty acid metabolism may be down-regulated. Two other genes in this pathway were differentially expressed, alcohol dehydrogenase and cytochrome P450. Their expression is again likely a general stress response and may not be specific to fatty acid metabolism.

In the acclimation dataset the fatty acid metabolism pathway was also differentially expressed. However a transcript encoding the enzyme ASC-17 was down-regulated, therefore during acclimation fatty acid metabolism is decreased. Both the general stress enzymes alcohol dehydrogenase and cytochrome P450 are up-regulated. The fatty acid biosynthesis pathway is also down-regulated in the acclimation dataset, since transcripts encoding the enzyme FASN-1, a fatty acid synthase family member that plays a major role in fatty acid synthesis is down-regulated. The pathway for starch and sugar metabolism was differentially expressed. The enzyme trehalase was up-regulated. This enzyme catalyses the conversion of the sugar trehalose to glucose. Although this pathway was not significant in the cold-shock dataset this gene was

significantly up-regulated. In this pathway glycogen synthase was down-regulated. This enzyme is responsible for the conversion of glucose to glycogen. The regulation of the genes in this pathway and the presence of glucose transporters further suggest that *P. superbus* utilises glucose as an energy source during cold acclimation.

A key regulator of animal development, the hedgehog signalling pathway contained a up-regulated gene for an enzyme (glycogen synthase kinase β (GSK3β)) and a down-regulated gene for megalin. Mutational studies of megalin in *C. elegans* have shown it is involved in growth and moulting (Yochem et al., 1999), while GSK3β phosphorylates another member of the hedgehog family targeting it for proteolysis to an inactive form. Overall this expression pattern would likely lead to a down-regulation and reduced growth rate in *P. superbus*. This is unsurprising as cold stress would generally lead to a reduction in growth rate and was seen in the *P. superbus* cultures placed at 10 ^{o}C.

Carrasco *et al.* (2011) have identified 39 primary research papers containing data on differentially expressed proteins/gene products in response to low temperatures. This is a relatively low number of studies considering that the literature encompasses data from bacteria, yeast invertebrates and vertebrates and that low temperatures and freezing conditions are important factors that determining the activity and distribution of living organisms. A possible reason for the paucity of studies on low temperature transcriptomics and proteomics is that most cold tolerant organisms are not model organisms whose genomes have not been sequenced. The data presented in this Chapter show the power of high-throughput RNAseq protocols to identify differentially expressed transcripts in a whole transcriptome study of *P. superbus*, a non-model organism.

Chapter VIII-General Discussion

The majority of the Earth's biosphere is cold and is exposed to low and subzero temperatures throughout the year. There are vast areas of the soil ecosystem that are only unfrozen for a few weeks of the year, and 90% of the ocean volume is below 5 ^{o}C. Freezing temperatures cause significant injury to cells and tissues through physical damage of the subcellular architecture caused by ice-crystals. Freezing also causes water to be drawn out from the cell resulting in cellular dehydration. These injuries will result in the release of reactive oxygen species and eventually lead to the damage of proteins and nucleic acids, loss of cellular compartmentalisation, and electrolyte and solute leakage. Despite these harsh conditions many organisms have colonised these environments. Birds and insects may undertake a long distance migratory flight to avoid the lowest temperatures. Others seek some refuge from the most extreme temperatures by spending winter in thermally buffered microclimates such as underwater, underground or under a snowpack. Species that cannot escape and winter above the snow-pack may have to deal with prolonged periods of very low subzero temperatures. This requires the development of biochemical and physiological adaptations to survive. These may include the production of proteins such as AFPs or cryoprotectants, elimination of ice nucleators, undergoing cryoprotective dehydration or the altered expression of stress associated genes.

An understanding of the mechanisms these freezing tolerant organisms employ to survive low temperature stress has important biotechnological, agricultural and medical applications. Cold-tolerant microorganisms produce cold-active enzymes that have high efficiency at low temperatures. The major current application of these enzymes is in the detergent industry but they may also be useful in the pharmaceutical and food industries (Cavicchioli et al., 2002). Cold-tolerant microorganisms have significant benefits in environmental biotechnology in reducing the cost of waste and contamination removal in cold locations (Margesin et al., 2007). They may also be used in the extraction of specific valuable metals from their ores through the use of bioleaching bacteria such as *Acidothiobacillus ferroxidans* as several mines worldwide operate at average temperatures of 8-10 ^{o}C (Margesin et al., 2007). The identification of the genes and proteins involved in freezing tolerance have potential

in the production of transgenic plants that are cold tolerant. This would be particularly useful for economically important freezing sensitive strains and would expand the geographic regions where they may be grown. The identification of potent cryoprotectants has the potential to improve methods of cryopreservation of gametes, cells, and tissues. Studying the molecular mechanisms of freezing tolerance may give an insight into how valuable products such as vaccines, blood and organs could be engineered to maintain their structural stability during cryopreservation. In addition ice recrystallisation inhibition proteins have applications in the frozen food industry (Regand and Goff, 2006; Kontogiorgos *et al.*, 2007).

Nematodes provide excellent animal models for the study of freezing stress tolerance. The phylum Nematoda is one of the most diverse grouping of taxa in the animal kingdom, with up to 100 million species (Blaxter et al., 1998). They occupy a wide variety of terrestrial and marine habitats and the ability to survive extreme stress varies greatly within a single genus as demonstrated in Chapters 3 and 4. This variability in stress resistance both within and between nematode genera provides an excellent research opportunity, allowing the comparison of the protective mechanisms across a variety of nematode species. The focus of this study was to characterise the freezing ability of members of the genus *Panagrolaimus*. This nematode genus was selected for this study for two reasons. Firstly, the nematode is easily cultured in large volumes using standard culturing methods developed for *C. elegans* (Brenner, 1974). Secondly, initial work by Shannon (Shannon, 2007) showed that this genus had different levels of freezing survival. Extensive studies on the nematode *P. davidi* have shown that it is freezing tolerant and can survive intracellular ice formation, a trait that is rarely seen in animals (Wharton and Brown, 1991; Wharton and Ferns, 1995). The ability of some *Panagrolaimus sp.* to survive freezing stress was confirmed in this study (Figure 3.2). The genus showed a range of freezing ability from strains that show high survival upon direct exposure to – 80 ^{o}C (>50%) to those that were freezing sensitive. Acclimation has been shown to improve the cold tolerance of many plants and animals including nematodes (Thomashow, 1999; Wharton *et al.*, 2000). It is thought that a period of acclimation may allow an individual time to conduct the necessary biochemical and metabolic adjustments that are required to survive low to freezing temperatures. Acclimation also significantly improves the freezing survival

of *Panagrolaimus sp.* with the most freezing tolerant nematodes having over 80% survival upon exposure to -80^{o}C after a 10 day acclimation period at 10 ^{o}C.

Shannon *et al.* (2005) showed that several strains of *Panagrolaimus sp.* are desiccation tolerant and that the genus has a range of tolerance from fast dehydration strategists to slow dehydration strategists and desiccation sensitive strains. In addition to the mechanical damage caused by growing ice, freezing also results in cellular dehydration as ice crystals grow outside the cell causing an efflux of water from the cell (Gusta et al., 1975). In Chapter 3 it was shown that there appears to be a correlation between desiccation tolerance and freezing tolerance of *Panagrolaimus* strains from temperate, polar, subpolar and continental regions of the world. This would suggest that the adaptations needed to survive freezing overlap with those required to survive desiccation. The freezing ability of these *Panagrolaimus sp.* could possibly to due to excellent ability to survive the desiccation stress that occurs during freezing. However data from Chapter 4 show that many tropical strains of *Panagrolaimus* are desiccation tolerant but not freezing tolerant suggesting that there are some key specialised adaptations required to survive freezing. One such adaptation may be the evolution of ice binding or ice recrystallisation inhibition proteins.

Many microorganisms, plants and animals produce ice-binding proteins. These proteins can adsorb to a growing ice crystal and restrict the growth of the ice front. AFPs can cause a non-colligative, non-equilibrium lowering of the melting point without significantly altering the melting point, a phenomenon known as thermal hysteresis (Raymond and DeVries, 1977). Whole-cell protein extracts that contain ice-binding proteins can bind a single hexagonal ice crystal and inhibit the growth of ice along specific axes of the crystal, as illustrated in Figure 3.5 resulting typically in hexagonal bipyrimidal ice crystals. A selection of *Panagrolaimus sp.* extracts were analysed for their ability to shape ice (Figure 3.5). The freezing tolerant strains and species can cause faceting of ice crystals while the freezing sensitive strains produce a similar ice crystal to the control with no disruption to ice growth. The freezing tolerant strains of *Panagrolaimus sp.* do not have significant TH ability, unlike other organisms such as insects that may depress the freezing point below the melting point by up to 5.5 ^{o}C (Graham et al., 1997). This poor TH activity has previously been

shown in *P. davidi* by Wharton *et al.* (2005), they proposed that it produces a recrystallisation inhibition protein. These RIPs cannot stop the growth ice crystals but enhance freezing survival by the inhibiting the growth of larger ice crystals that cause the major mechanical damage to cells (Yu et al., 2010). Because of the ability of the freezing tolerant strains in our collection of *Panagrolaimus* to shape ice while having minimal TH activity we hypothesised that these strains may also produce a RIP. It may be possible that all of the strains have this protein but with differing affinity for ice, with the RIPs from the most tolerant strains having high affinity for ice.

In Chapter 6 an attempt was made to purify an ice binding protein from protein extracts from *P. superbus.* The diversity of IBPs is great and there is no reliable way of identifying an AFP from amino acid sequence, this makes purification and identification of IBPs a difficult and time consuming process. In this Chapter we used a purification method based on the only characteristic that defines IBPs, affinity for ice, to purify IBPs from *P. superbus* (Kuiper et al., 2003). This method followed by mass spectrometry analysis identified several proteins that are enriched in the ice fraction. However no single highly enriched candidate protein was detected. Because of the large number of potential candidate proteins it was impossible to determine which of these had IBP activity without the testing for ice shaping through the production of several recombinant proteins. These potential ice-binding proteins were not up-regulated in either of the RNA-seq dataset. Because of the time required to carry out those experiments and the high risk of negative results these experiments were not undertaken. It may also be possible that the proteins that shape the ice do not actually bind to the ice; they may be interfering with ice growth in some form of colligative manner. When this experiment was conducted all existing publications inferred that IBPs were proteins. Recently a non-protein TH-producing biomolecule has been identified and isolated from the freezing-tolerant beetle, *Upis cerambsoides* (Walters et al., 2009b). Further work by Walters *et al.* (2011) isolated a xylomannan-based antifreeze glycolipid from a plant and 6 additional insect species. Since an IBP has not been successfully found in *P. davidi* through both ice-affinity and biochemical purification methods (Wharton, Personal Communication) and no clear IBPs were found in this study it may be possible that freezing tolerant nematodes do not produce proteinaceous antifreeze molecules but some form of glycolipid. Further work could use the method employed by Walters *et al.* (2009, 2011) to see whether *P. superbus*

contains xylomannan-based glycolipids and whether they have the ability to shape ice.

Previous publications have shown that the desiccation tolerant *Panagrolaimus* strains and species form a single phylogenetic lineage, which is separate from the desiccation sensitive strains, based on the D3 expansion region from the 28S rRNA and the ITS regions (Shannon et al., 2005). In this study a more extensive phylogenetic analysis was conducted with more strains and species and using the complete DNA sequence of 18S and the D3 rRNA genes. Separate and concatenated phylogenetic trees were constructed to see whether these freezing and desiccation tolerant nematodes would form a single lineage. This was not found, two sensitive strains were found amongst two separate groups of tolerant nematodes (Figure 3.8). The 18S and D3 are evolutionary conserved and slowly evolving molecules and although they are generally considered excellent for estimating the phylogenetic relationships, they may be too conservative to give sufficient signal to show an accurate representation of the phylogenetic relationship of closely related nematode strains. The best way of truly seeing the relationships between closely related species is through the creation of super trees based on complete genome sequences (Rannala and Yang, 2008). The super tree approach has been shown to give improved the resolution of the phylogeny of closely related individuals such as bacterial strains than that obtained using the 16S rRNA-sequence and sequences from a small number of housekeeping genes housekeeping genes (Haggerty et al., 2009). Currently such an approach is not possible for the genus *Panagrolaimus* but with advent of next generation sequencing non-model species can be sequenced with relative ease and speed so this may be a real possibility in the future. Genome sequencing is currently progressing for *Panagrolaimus sp.* 1159 and *P. superbus* and has been proposed for *Panagrolaimus sp.* JU765 (959 Nematode Genomes Project).

Following the completion of the stress survival experiments of Chapter 3, we acquired 10 new strains of *Panagrolaimus* from Dr. Félix (Institut Jacques Monod, Paris, France) (Chapter 4). These were isolated from more tropical regions that would have a range from dry, arid to humid, monsoon climates at different times throughout the year. Since these nematodes would not encounter freezing stress to the same degree as some of the nematodes analysed in Chapter 3 that were isolated from Iceland (*P.*

superbus) or Antarctica (*P. davidi*) it was interesting to measure their tolerance to desiccation and freezing stress. Nine out of ten of these strains can tolerate some level of freezing that is significantly improved by acclimation (Figure 4.2) but they are poor at freezing compared to the tolerant strains of Chapter 3. Interestingly nine of the ten *Panagrolaimus sp.* Félix collection nematodes are excellent slow desiccation strategists. Therefore amongst these nematodes there is no correlation between desiccation tolerance and freezing tolerance. This suggests that in nematodes freezing tolerance is not an ability that is a fortunate consequence of desiccation tolerance, the two stresses are distinct. It maybe that very few modifications are required for freezing tolerance to evolve from desiccation tolerance but these strains would not have been subjected to the required selection pressure. A phylogeny was also constructed for the Félix collection to see whether there was any link between phylogeny, stress tolerance and biogeography. Although there did seem to be some loose correlation it was not definite, similar to the Chapter 3 phylogenies, the freezing and desiccation sensitive strain JU765 was found amongst the more tolerant strains. A phylogeny was also constructed will all of the species and strains studied in Chapter 3 and 4 (Figure 4.7). This showed that apart from JU765 the Félix collection of strains that are slow dehydration strategists are divergent from the other strains and species except AS01. However AS01 has a similar level of freezing a desiccation survival as many of these strains. This phylogeny has demonstrated that these nematodes do have a general relationship between their phylogeny and stress tolerance but seem quite unusual in that sensitive strains may be found amongst the tolerant. It may be that these sensitive strains once had the ability to survive stresses but due to their biogeography or microhabitat they may no longer require this adaptations so it has been lost.

The sister species of the Antarctic nematode *P. davidi* has been shown to be PS1579 (Lewis et al., 2009). This was relationship was supported in this study in the D3 region phylogeny (Figure 3.8). *P. superbus* was isolated from an Icelandic island and its sister species is AF36. Since PS1579 was found in California and AF36 from Pennsylvania, there does not seem to be a relationship between biogeography and freezing tolerance amongst these nematodes. It is apparent since *P. superbus* was isolated from a gull's nest with a freezing and desiccation sensitive strain *P. detritophagus* (Boström, 1988) that it did not evolve on Surtsey island but was

transported there in a desiccated state by an Arctic bird. However all described Antarctic nematode species are considered to be endemic to the region (Andrássy, 1998; Maslen and Convey, 2006). In Chapter 5 the divergence times for these *Panagrolaimus sp.* were estimated using the relaxed molecular clock approach. The calculation of these dates estimated age of origin of these *Panagrolaimus sp*. and give insights into the time of *P. davidi* dispersal to Antarctica. This analysis found that AS01 diverged from the other strains approximately 70.12 MYA (41.16-108.5) in the Upper Cretaceous period, with further diversification of the freezing and desiccation tolerant Panagrolaimids occurring during the Eocene and Miocene epochs of the Cenozoic Paleogene period. *P. davidi* and PS1579 diverged from *P. superbus* and AF36 in the Eocene epoch at 47.77 MYA. During this time Australia was separating from Antarctica and drifting northwards leading to isolation of the Antarctica continent (Kennett, 1977). As the Cenozoic era progressed Antarctica changed from been a warm and mainly ice-free to the extreme polar continent seen today. We found *P. davidi* and *P. superbus* diverged from their temperate sister species in the Miocene epoch. This was an epoch where high rates of cooling occurred in Antarctica. It is unknown whether *P. davidi* evolved in the Antarctica as it cooled during this period or whether it may have been transported there more recently. Future analysis could try to narrow the range of divergence times between *P. davidi* and PS1579, perhaps by adding in more calibration dates (or adding the dates found in this analysis as calibration points) particularly within the nematodes.

Following the results that were found in Chapters 3, 4 and 6 it was decided to look at the types of genes that tend to up-regulated upon cold stress and whether a period of acclimation upregulates genes that may potentially improve freezing survival (Chapter 7). The NGS method RNA-seq was used in this study. Several genes were found significantly differentially expressed in *P. superbus* following a cold-shock at 4 ^{o}C for 24 h and an acclimation at 10 ^{o}C for 10 days. In general the cold-shock dataset had less differentially expressed genes than the acclimated dataset. The cold-shock may not have been severe enough to cause a large difference in the gene expression whereas in the acclimated dataset the nematodes had a long period of time to adjust and up-regulated a larger number of response genes. The datasets all showed a large proportion of novel sequences were up-regulated in response to cold stress. The genes with the highest levels of up-regulation were novel sequences. These highly up-

regulated novel sequences are likely important in freezing tolerance. Unfortunately these were not investigated. Future work would include further analysis of the sequences. These novel genes may encode ice-binding proteins; this could be tested with ice-activity assays of recombinantly expressed highly expressed novels. In both datasets the genes that were significantly up regulated and identified were involved in the oxidative stress response, transporting, membrane modification, metabolism, signalling and cytoskeleton remodelling. Since the majority of the up-regulated genes appear to relate to a more general response to stress it is possible that the novel genes distinguish the cold tolerance phenotype seen in *P. superbus* from related sensitive strains. This could be investigated by performing RNA-seq on an acclimated freezing sensitive *Panagrolaimus sp.* and comparing the gene expression response between both strains.

In addition to RNA-seq of cold-shocked and acclimated *P. superbus* another researcher from our laboratory is investigating differential gene expression in *P. superbus* in response to preconditioning to mild desiccation. It will be very interesting to see the overlap in the genes that are up and downregulated between the desiccated and cold datasets. This would give another level of understanding to the cross-adaptation that exists between freezing and desiccation tolerances.

This project has shown that *Panagrolaimus sp.* has the potential to be an excellent model system for the investigation of molecular aspects of freezing survival as well as desiccation survival. This genus contains closely related strains and species, which possess a diverse range of survival abilities, which is ideal for comparative analysis of the stress response. This project has only touched on the surface of the stress tolerance and phylogeny of the *Panagrolaimus* sp. It has shown that this genus represents a highly complex and plastic group of nematodes that can be used to greatly advance our knowledge into the molecular mechanisms that underlie anhydrobiosis and cryobiosis.

Bibliography

ABELE, D. & PUNTARULO, S. (2004) Formation of reactive species and induction of antioxidant defence systems in polar and temperate marine invertebrates and fish. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology,* 138**,** 405-15.

ADHIKARI, B. N., WALL, D. H. & ADAMS, B. J. (2009) Desiccation survival in an Antarctic nematode: molecular analysis using expressed sequenced tags. *BMC Genomics,* 10**,** 69.

ADHIKARI, B. N., WALL, D. H. & ADAMS, B. J. (2010) Effect of slow desiccation and freezing on gene transcription and stress survival of an Antarctic nematode. *Journal of Experimental Biology,* 213**,** 1803-12.

AL-FAGEEH, M. B. & SMALES, C. M. (2006) Control and regulation of the cellular responses to cold shock: the responses in yeast and mammalian systems. *Biochemical Journal,* 397**,** 247-59.

AL-FAGEEH, M. B. & SMALES, C. M. (2009) Cold-inducible RNA binding protein (CIRP) expression is modulated by alternative mRNAs. *RNA,* 15**,** 1164-76.

ALIA, MOHANTY, P. & MATYSIK, J. (2001) Effect of proline on the production of singlet oxygen. *Amino Acids,* 21**,** 195-200.

ALOIA, R. C., PENGELLEY, E. T., BOLEN, J. L. & ROUSER, G. (1974) Changes in phospholipid composition in hibernating ground squirrel, *Citellus lateralis,* and their relationship to membrane function at reduced temperature. *Lipid,* 9**,** 993-999.

AMORNWITTAWAT, N., WANG, S., DUMAN, J. G. & WEN, X. (2008) Polycarboxylates enhance beetle antifreeze protein activity. *Biochimica et Biophysica Acta,* 1784**,** 1942-8.

AMORNWITTAWAT, N., WANG, S., BANATLAO, J., CHUNG, M., VELASCO, E., DUMAN, J. G. & WEN, X. (2009) Effects of polyhydroxy compounds on beetle antifreeze protein activity. *Biochemica et Biophysica Acta,* 1794**,** 341-6.

ANCHORDOGUY, T. J., RUDOLPH, A. S., CARPENTER, J. F. & CROWE, J. H. (1987) Models of interaction of cryoprotectants with membrane phospholipids during freezing. *Cryobiology,* 24**,** 324-331.

ANDERS, S. & HUBER, W. (2010) Differential expression analysis for sequence count data. *Genome Biology,* 11**,** R106.

ANDERSON, J. B., WARNY, S., ASKIN, R. A., WELLNER, J. S., BOHATY, S. M., KIRSHNER, A. E., LIVSEY, D. N., SIMMS, A. R., SMITH, T. R., EHRMANN, W., LAWVER, L. A., BARBEAU, D., WISE, S. W., KULHENEK, D. K., WEAVER, F. M. & MAJEWSKI, W. (2011) Progressive

Cenozoic cooling and the demise of Antarctica's last refugium. *Proceedings of the National Academy of Sciences of United States of America,* 108**,** 11356-60.

ANDRÁSSY, I. (1998) Nematodes in the sixth continent. *Journal of nematode morphology and systematics,* 1**,** 107-186.

ANTIKAINEN, M. & GRIFFITH, M. (1997) Antifreeze protein accumulation in freezing tolerant cereals. *Physiologia Plantarum,* 99**,** 423-432.

ARAV, A., RUBINSKY, B., FLETCHER, G. & SEREN, E. (1993) Cryogenic protection of oocytes with antifreeze proteins. *Molecular Reproduction and Development,* 36**,** 488-93.

ARNY, D. C., LINDOW, S. E. & UPPER, C. D. (1976) Frost sensitivity of corn increased by application of *Pseudomonas syringae*. *Nature,* 262**,** 282-284.

AROIAN, R. V., CARTA, L., KALOSHIAN, I. & STERNBERG, P. W. (1993) *Panagrolaimus* sp. from Armenia can survive in anhydrobiosis for 8.7 years. *Journal of Nematology,* 25**,** 500-502.

ASHRAF, M. & FOOLAD, M. R. (2007) Roles of glycine betaine and proline in improving plant abiotic stress resistance. *Environmental and Experimental Botany,* 59**,** 206-216.

ATICI, O. & NALBANTOGLU, B. (1999a) Apoplastic proteins associated with the cold acclimation process in leaves. *Bio-Science Research Bulletin,* 15**,** 55-60.

ATICI, O. & NALBANTOGLU, B. (1999b) Effect of apoplastic proteins on freezing tolerance in leaves. *Phytochemistry,* 50**,** 755-761.

ATLUNG, T. & HANSEN, F. G. (1999) Low-temperature-induced DnaA protein synthesis does not change initiation mass in *Escherichia coli* K-12. *Journal of Bacteriolgy,* 181**,** 5557-62.

AVONCE, N., MENDOZA-VARGAS, A., MORETT, E. & ITURRIAGA, G. (2006) Insights on the evolution of trehalose biosynthesis. *BMC Evolutionary Biology,* 6**,** 109.

BAARDSNES, J., KONDEJEWSKI, L. H., HODGES, R. S., CHAO, H., KAY, C. & DAVIES, P. L. (1999) New ice-binding face for type I antifreeze protein. *FEBS Letters,* 463**,** 87-91.

BAARDSNES, J. & DAVIES, P. L. (2001) Sialic acid synthase: the origin of fish type III antifreeze protein? *Trends in Biochemical Sciences,* 26**,** 468-9.

BAKER, S. S., WIHELM, K. S. & THOMASHOW, M. F. (1994) the 5'-region of *Arabidopsis thaliana cor15a* cis-acting elements that confer cold-,drought- and ABA-regulated gene expression. *Plant Molecular Biology,* 24**,** 701-713.

BALDURSSON, S. & INGADÓTTIR, Á. (2007) Nomination of Surtsey for the UNESCO World Heritage List. (Icelandic Institute of Natural History, Reykjavík).

BARRETT, J. (2001) Thermal hysteresis proteins. *International Journal of Biochemistry & Cell Biology,* 33**,** 105-117.

BENAROUDJ, N., LEE, D. H. & GOLDBERG, A. L. (2001) Trehalose accumulation during cellular stress protects cells and cellular proteins from damage by oxygen radicals. *Journal of Biological Chemistry,* 276**,** 24261-7.

BENTLEY, D. R., BALASUBRAMANIAN, S., SWERDLOW, H. P., SMITH, G. P., MILTON, J., BROWN, C. G., HALL, K. P., EVERS, D. J., BARNES, C. L., BIGNELL, H. R., BOUTELL, J. M., BRYANT, J., CARTER, R. J., KEIRA CHEETHAM, R., COX, A. J., ELLIS, D. J., FLATBUSH, M. R., GORMLEY, N. A., HUMPHRAY, S. J., IRVING, L. J., KARBELASHVILI, M. S., KIRK, S. M., LI, H., LIU, X., MAISINGER, K. S., MURRAY, L. J., OBRADOVIC, B., OST, T., PARKINSON, M. L., PRATT, M. R., RASOLONJATOVO, I. M., REED, M. T., RIGATTI, R., RODIGHIERO, C., ROSS, M. T., SABOT, A., SANKAR, S. V., SCALLY, A., SCHROTH, G. P., SMITH, M. E., SMITH, V. P., SPIRIDOU, A., TORRANCE, P. E., TZONEV, S. S., VERMAAS, E. H., WALTER, K., WU, X., ZHANG, L., ALAM, M. D., ANASTASI, C., ANIEBO, I. C., BAILEY, D. M., BANCARZ, I. R., BANERJEE, S., BARBOUR, S. G., BAYBAYAN, P. A., BENOIT, V. A., BENSON, K. F., BEVIS, C., BLACK, P. J., BOODHUN, A., BRENNAN, J. S., BRIDGHAM, J. A., BROWN, R. C., BROWN, A. A., BUERMANN, D. H., BUNDU, A. A., BURROWS, J. C., CARTER, N. P., CASTILLO, N., CHIARA, E. C. M., CHANG, S., NEIL COOLEY, R., CRAKE, N. R., DADA, O. O., DIAKOUMAKOS, K. D., DOMINGUEZ-FERNANDEZ, B., EARNSHAW, D. J., EGBUJOR, U. C., ELMORE, D. W., ETCHIN, S. S., EWAN, M. R., FEDURCO, M., FRASER, L. J., FUENTES FAJARDO, K. V., SCOTT FUREY, W., GEORGE, D., GIETZEN, K. J., GODDARD, C. P., GOLDA, G. S., GRANIERI, P. A., GREEN, D. E., GUSTAFSON, D. L., HANSEN, N. F., HARNISH, K., HAUDENSCHILD, C. D., HEYER, N. I., HIMS, M. M., HO, J. T., HORGAN, A. M., et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature,* 456**,** 53-9.

BENTON, M. J. & DONOGHUE, P. C. (2007) Paleontological evidence to date the tree of life. *Molecular Biology and Evolution,* 24**,** 26-53.

BENTON, M. J., DONOGHUE, P. C. & ASHER, R. J. (2009) Calibrating and constraining molecular clocks. IN HEDGES, S. B. & KUMAR, S. (Eds.) *The Timetree of Life.* Oxford University Press

BHADURY, P., AUSTEN, M. C., BILTON, D. T., LAMBSHEAD, P. J., ROGERS, A. D. & SMERDON, G. R. (2006) Development and evaluation of a DNA-barcoding approach for the rapid identification of nematodes. *Marine Ecology Progress Series,* 320**,** 1-9.

BLAXTER, M. (1998) *Caenorhabditis elegans* is a nematode. *Science,* 282**,** 2041-6.

BLAXTER, M. L., DE LEY, P., GAREY, J. R., LIU, L. X., SCHELDEMAN, P., VIERSTRAETE, A., VANFLETEREN, J. R., MACKEY, L. Y., DORRIS,

M., FRISSE, L. M., VIDA, J. T. & THOMAS, W. K. (1998) A molecular evolutionary framework for the Phylum Nematoda. *Nature,* 392**,** 71-5.

BLAXTER, M., ELSWORTH, B. & DAUB, J. (2004) DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. *Proceedings of the Royal Society B: Biological Sciences,* 271 Suppl 4**,** S189-92.

BLAXTER, M., MANN, J., CHAPMAN, T., THOMAS, F., WHITTON, C., FLOYD, R. & ABEBE, E. (2005) Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 360**,** 1935-43.

BLOCK, W. (2003) Water or ice?--the challenge for invertebrate cold survival. *Sci Prog,* 86**,** 77-101.

BODNER, M. & LARCHER, W. (1987) Chilling susceptibility of different organs and tissues of *Saintpaulia ionantha* and *Coffea arabica*. *Journal of Applied Botany,* 61**,** 225-242.

BORGONIE, G., GARCIA-MOYANO, A., LITTHAUER, D., BERT, W., BESTER, A., VAN HEERDEN, E., MOLLER, C., ERASMUS, M. & ONSTOTT, T. C. (2011) Nematoda from the terrestrial deep subsurface of South Africa. *Nature,* 474**,** 79-82.

BOSTRÖM, S. (1988) Descriptions and morphological variability of three populations of *Panagrolaimus* Fuchs, 1930 (Nematoda, Panagrolaimidae). *Nematologica,* 34**,** 144-155.

BRADFORD, J. R., HEY, Y., YATES, T., LI, Y., PEPPER, S. D. & MILLER, C. J. (2010) A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. *BMC Genomics,* 11**,** 282.

BRAUN, E. L., KANG, S., NELSON, M. A. & NATVIG, D. O. (1998) Identification of the first fungal annexin: analysis of annexin gene duplications and implications for eukaryotic evolution. *Journal of Molecular Evolution,* 47**,** 531-43.

BRENNER, S. (1974) The genetics of Caenorhabditis elegans. *Genetics,* 77**,** 71-94.

BRENNER, S., JOHNSON, M., BRIDGHAM, J., GOLDA, G., LLOYD, D. H., JOHNSON, D., LUO, S., MCCURDY, S., FOY, M., EWAN, M., ROTH, R., GEORGE, D., ELETR, S., ALBRECHT, G., VERMAAS, E., WILLIAMS, S. R., MOON, K., BURCHAM, T., PALLAS, M., DUBRIDGE, R. B., KIRCHNER, J., FEARON, K., MAO, J. & CORCORAN, K. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology,* 18**,** 630-4.

BROWN, I. M., WHARTON, D. A. & MILLAR, R. B. (2004) The influence of temperature on the life history of the Antarctic nematode *Panagrolaimus davidi*. *Nematology,* 6**,** 883-890.

BROWNE, J. A., TUNNACLIFFE, A. & BURNELL, A. M. (2002) Plant desiccation gene found in a nematode. *Nature,* 416.

BROWNE, J. A., DOLAN, K. M., TYSON, T., GOYAL, K., TUNNACLIFFE, A. & BURNELL, A. M. (2004) Dehydration-specific induction of hydrophilic protein genes in the anhydrobiotic nematode *Aphelenchus avenae*. *Eukaryotic Cell,* 3**,** 966-75.

BULLARD, J. H., PURDOM, E., HANSEN, K. D. & DUDOIT, S. (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics,* 11**,** 94.

BURNELL, A. M. & TUNNACLIFFE, A. (2011) Gene induction and desiccation stress in nematodes. IN PERRY, R. N. & WHARTON, D. A. (Eds.) *Molecular and Physiological Basis of Nematode Survival.* CABI. Wallingford Oxfordshire UK

BURROWS, M. & WHEELER, D. (1994) A block sorting lossless data compression algorithm. *Technical Report 124,* DEC, Digital Systems Reseach Center, Palo Alto, California.

BUTLER, M. H., WALL, S. M., LUEHRSEN, K. R., FOX, G. E. & HECHT, R. M. (1981) Molecular relationships between closely related strains and species of nematodes. *Journal of Molecular Evolution,* 18**,** 18-23.

BUTTNER, M. P. & AMY, P. S. (1989) Survival of ice nucletaing active and genetically engineered non-ice nucleating *Pseudomonas syringae* strains after freezing *Applied Environmental Microbiology,* 55**,** 1690-1694.

BYRANT, G., KOSTER, K. L. & WOLFE, J. (2001) Membrane behaviour in seeds and other systems at low water content: the various effects of solutes. *Seed Science Research,* 11**,** 17-25.

CALDAS, T., LAALAMI, S. & RICHARME, G. (2000) Chaperone properties of bacterial elongation factor EF-G and initiation factor IF2. *Journal of Biological Chemistry,* 275**,** 855-60.

CALDERON, S., HOLMSTRUP, M., WESTH, P. & OVERGAARD, J. (2009) Dual roles of glucose in the freeze-tolerant earthworm *Dendrobaena octaedra*: cryoprotection and fuel for metabolism. *Journal of Experimental Biology,* 212**,** 859-66.

CARPENTER, J. F. & CROWE, J. H. (1988) The mechanism of cryoprotection of proteins by solutes. *Cryobiology,* 25**,** 224-225.

CARPENTER, J. F., CROWE, L. M. & CROWE, J. H. (1987) Stabilization of phosphofructokinase with sugars during freeze-drying: characterization of enhanced protection in the presence of divalent cations. *Biochimica et Biophysica Acta (BBA) - General Subjects,* 923**,** 109-115.

CARPENTER, J. F. (1993) Stabilization of proteins during freezing and dehydration:application of lessons from nature. *Cryobiology,* 30**,** 220-221.

CARRASCO, M. A., TAN, J. C. & DUMAN, J. G. (2011) A cross-species compendium of proteins/gene products related to cold stress identified by bioinformatic approaches. *Journal of Insect Physiology,* 57**,** 1127-35.

CAVICCHIOLI, R., SIDDIQUI, K. S., ANDREWS, D. & SOWERS, K. R. (2002) Low-temperature extremophiles and their applications. *Current Opinion in Biotechnology,* 13**,** 253-61.

CHAKRABARTTY, A., HEW, C. L., SHEARS, M. & FLETCHER, G. L. (1988) Primary structures of the alanine rich antifreeze polypeptides from grubby sculpin. *Canadian Journal of Zoology,* 66**,** 403-408.

CHAO, H., HODGES, R. S., KAY, C. M., GAUTHIER, S. Y. & DAVIES, P. L. (1996) A natural variant of type I antifreeze protein with four ice-binding repeats is a particularly potent antifreeze. *Protein Science,* 5**,** 1150-6.

CHAO, H., SONNICHSEN, F. D., DELUCA, C. I., SYKES, B. D. & DAVIES, P. L. (1994) Structure-function relationship in the globular type-III antifreeze protein: identification of a cluster of surface residues required for binding to ice. *Protein Science,* 3**,** 1760-1769.

CHAO, H., HOUSTON, M. E., JR., HODGES, R. S., KAY, C. M., SYKES, B. D., LOEWEN, M. C., DAVIES, P. L. & SONNICHSEN, F. D. (1997) A diminished role for hydrogen bonds in antifreeze protein binding to ice. *Biochemistry,* 36**,** 14652-60.

CHEN, A., GOLLOP, N. & GLAZER, I. (2005) Cross-stress tolerance and expression of stress-realted proteins in osmotically desiccated entomopathogenic *Steinernema feltiae* IS-6. *Parasitology,* 131**,** 695-703.

CHEN, L., DEVRIES, A. L. & CHENG, C. H. (1997a) Convergent evolution of antifreeze glycoprotein in Antarctic notothenioid fish and Arctic cod. *Proceedings of the National Academy of Sciences of United States of America,* 94**,** 3817-3822.

CHEN, L., DEVRIES, A. L. & CHENG, C. H. (1997b) Evolution of antifreeze glycoprotein gene from a trypsonogen gene in Antarctic notothenoid fish. *Proceedings of the National Academy of Sciences of United States of America,* 94**,** 3811-3816.

CHEN, W. H., SUN, L. T., TSAI, C. L., SONG, Y. L. & CHANG, C. F. (2002) Cold-stress induced the modulation of catecholamines, cortisol, immunoglobulin M, and leukocyte phagocytosis in tilapia. *General and Comparative Endocrinology,* 126**,** 90-100.

CHENG, C. H. & DEVRIES, A. L. (1989) Structures of antifreeze peptides from the antarctic eel pout, *Austrolycicthys brachycephalus*. *Biochimica et Biophysica Acta,* 997**,** 55-64.

CHITWOOD, B. G. (1933) The characters of a protonematode. *Journal of Parasitology,* 20**,** 130.

CLARK, G. T., KONOPKA-POSTUPOLSKA, D., HENNIG, J. & ROUX, S. (2010) Is annexin 1 a multifuntional protein during stress responses? *Plant Signaling and Behavior,* 5**,** 303-307.

CLARK, M. S. & WORLAND, M. R. (2008) How insects survive the cold: molecular mechanisms-a review. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology,* 178**,** 917-933.

CLEGG, J. S. (1967) Metabolic studies of cryptobiosis in encysted embryos of *Artemia salina*. *Comparative Biochemistry and Physiology,* 20**,** 801-809.

CLEGG, J. S. (1965) The origin of trehalose and its significance during the emergence of encysted dormant embryos of *Artemia salina*. *Comparative Biochemistry and Physiology,* 14**,** 135-143.

COLLMER, A. & KEEN, N. T. (1986) The Role of Pectic Enzymes in Plant Pathogenesis. *Annual review of Phytopathology,* 24**,** 383-409.

CONVEY, P., GIBSON, J. A., HILLENBRAND, C. D., HODGSON, D. A., PUGH, P. J., SMELLIE, J. L. & STEVENS, M. I. (2008) Antarctic terrestrial life--challenging the history of the frozen continent? *Biological Reviews of the Cambridge Philosophical Society,* 83**,** 103-17.

COSTA, V., ANGELINI, C., DE FEIS, I. & CICCODICOLA, A. (2010a) Uncovering the complexity of transcriptomes with RNA-Seq. *Journal of Biomedicine and Biotechnology,* 2010**,** 853916.

COSTA, V., ANGELINI, C., DE FEIS, I. & CICCODICOLA, A. (2010b) Uncovering the complexity of transcriptomes with RNA-Seq. *J Biomed Biotechnol,* 2010**,** 853916.

COSTANZO, J. P., GRENOT, C. & LEE, R. E., JR. (1995) Supercooling, ice inoculation and freeze tolerance in the European common lizard, *Lacerta vivipara*. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology,* 165**,** 238-44.

COUGHLAN, A. & WOLFE, K. H. (2002) Fourfold Faster Rate of Genome Rearrangement in Nematodes Than in *Drosophila*. *Genome Research,* 12**,** 857-867.

COUGHLAN, S. J. & HERBER, U. (1982) The role of glycine betaine in the protection of spinach thylakoids against freezing stress. *Planta,* 156**,** 62-69.

COWAN, K. J. & STOREY, K. B. (2003) Mitogen-activated protein kinases: new signaling pathways functioning in cellular responses to environmental stress. *Journal of Experimental Biology,* 206**,** 1107-15.

COX, J. C., INGERSOLL, J. E. & ROSS, S. A. (1985) A Theory of the Term Structure of Interest Rates. *Econometrica,* 53**,** 385-407.

CRAWFORD, J. E., GUELBEOGO, W. M., SANOU, A., TRAORE, A., VERNICK, K. D., SAGNON, N. & LAZZARO, B. P. (2010) De novo transcriptome

sequencing in Anopheles funestus using Illumina RNA-seq technology. *PLoS One,* 5**,** e14202.

CRERAR, M. M., DAVID, E. S. & STOREY, K. B. (1988) Electrophoretic analysis of liver glycogen phosphorylase activation in the freeze-tolerant wood frog. *Biochimica et Biophysica Acta,* 971**,** 72-84.

CROWE, J. H., JACKSON, S. & CROWE, L. M. (1983) Nonfreezable water in anhydriobiotic nematodes. *Molecular Physiology,* 3**,** 99- 105.

CROWE, J. H., CROWE, L. M. & CHAPMAN, D. (1984) Preservation of membranes in anhydrobiotic organisms. *Science,* 223**,** 701-703.

CROWE, J. H. (2007) *Trehalose As a" Chemical Chaperone": Fact and Fantasy*, New York, Plenum Press

CROWE, L. M. & CROWE, J. H. (1992) Anhydrobiosis: a strategy for survival. *Advances in Space Research,* 12**,** 239-47.

CROWE, L. M., REID, D. S. & CROWE, J. H. (1996) Is trehalose special for preserving dry biomaterials? *Biophysical Journal,* 71**,** 2087-2093.

CUTTER, A. D. (2008) Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Molecular Biology and Evolution,* 25**,** 778-86.

DAMMEL, C. S. & NOLLER, H. F. (1995) Suppression of a cold-sensitive mutation in 16S rRNA by overexpression of a novel ribosome-binding factor, RbfA. *Genes Dev,* 9**,** 626-37.

DANKS, H. V. (2000) Dehydration in dormant insects. *Journal of Insect Physiology,* 46**,** 837-852.

DARWIN, C. (1859) *The Origin of Species*, John Murray, London

DAVIES, P. L. & HEW, C. L. (1990) Biochemistry of fish antifreeze proteins. *Faseb Journal,* 4**,** 2460-2468.

DAVIES, P. L., ROACH, A. H. & HEW, C. L. (1982) DNA sequence coding for an antifreeze protein precursor from winter flounder. *Proceedings of the National Academy of Sciences of United States of America,* 79**,** 335-9.

DAVIES, P. L. & SYKES, B. D. (1997) Antifreeze proteins. *Current Opinion in Structural Biology,* 7**,** 828-834.

DAVIES, P. L., BAARDSNES, J., KUIPER, M. J. & WALKER, V. K. (2002) Structure and function of antifreeze proteins. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences,* 357**,** 927-933.

DE GARCIA, V., BRIZZIO, S., LIBKIND, D., BUZZINI, P. & VAN BROOCK, M. (2007) Biodiversity of cold-adapted yeasts from glacial meltwater rivers in Patagonia, Argentina. *FEMS Microbiology Ecology,* 59**,** 331-41.

DE LEY, P., FELIX, M. A., FRISSE, L. M., NADLER, S. A., STERNBER, P. W. & THOMAS, W. K. (1999) Molecular and morphological characterisation of two reproductively isolated species with mirror-image anatomy (Nematoda: Cephalobidae). *Nematology,* 1**,** 591-612.

DE LEY, P., DE LEY, I. T., MORRIS, K., ABEBE, E., MUNDO-OCAMPO, M., YODER, M., HERAS, J., WAUMANN, D., ROCHA-OLIVARES, A., JAY BURR, A. H., BALDWIN, J. G. & THOMAS, W. K. (2005) An integrated approach to fast and informative morphological vouchering of nematodes for applications in molecular barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 360**,** 1945-58.

DEMEURE, Y. & FRECKMAN, D. W. (1981) Recent advances in the study of anhydrobiotic nematodes. IN ZUCKERMAN, B. M. & ROHDE, R. A. (Eds.) *Plant Parasitic nematodes. Volume III.* Academic Press, New York

DENG, G., ANDREWS, D. W. & LAURSEN, R. A. (1997) Amino acid sequence of a new type of antifreeze protein, from the longhorn sculpin *Myoxocephalus octodecimspinosis*. *FEBS Letters,* 402**,** 17-20.

DENG, G. & LAURSEN, R. A. (1998) Isolation and characterization of an antifreeze protein from the longhorn sculpin, *Myoxocephalus octodecimspinosis*. *Biochimica et Biophysica Acta,* 1388**,** 305-14.

DERSCH, P., KNEIP, S. & BREMER, E. (1994) The nucleoid-associated DNA-binding protein H-NS is required for the efficient adaptation of *Escherichia coli* K-12 to a cold environment. *Molecular and General Genetics,* 245**,** 255-259.

DEVRIES, A. L. & WOHLSCHLAG, D. E. (1969) Freezing resistance in some Antarctic fishes. *Science,* 163**,** 1073-5.

DEVRIES, A. L., KOMATSU, S. K. & FEENEY, R. E. (1970) Chemical and physical properties of freezing point-depressing glycoproteins from Antarctic fishes. *Journal of Biological Chemistry,* 245**,** 2901-8.

DEVRIES, A. L., VANDENHEEDE, J. & FEENEY, R. E. (1971) Primary structure of freezing point-depressing glycoproteins. *Journal of Biological Chemistry,* 246**,** 305-8.

DEVRIES, A. L. (1983) Antifreeze peptides and glycopeptides in cold-water fishes. *Annual Review of Physiology,* 45**,** 245-260.

DEVRIES, A. L. (1984) Role of glycopeptides and peptides in inhibition of crystallization of water in polar fishes. *Philosophical Transactions of the Royal Society of London B,* 304**,** 574-588.

DINGLE, H. (1996) *Migration : the biology of life on the move,* New York ; Oxford, Oxford University Press

DODD, R. B. & DRICKAMER, K. (2001) Lectin-like proteins in model organisms: implications for evolution of carbohydrate-binding activity. *Glycobiology,* 11**,** 71R-9R.

DOHM, J. C., LOTTAZ, C., BORODINA, T. & HIMMELBAUER, H. (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research,* 36**,** e105.

DORRIS, M., DE LEY, P. & BLAXTER, M. L. (1999) Molecular analysis of nematode diversity and the evolution of parasitism. *Parasitology Today,* 15**,** 188-93.

DOUCET, C. J., BYASS, L., ELIAS, L., WORRALL, D., SMALLWOOD, M. & BOWLES, D. J. (2000) Distribution and characterization of recrystallization inhibitor activity in plant and lichen species from the UK and maritime Antarctic. *Cryobiology,* 40**,** 218-27.

DOUCET, D., WALKER, V. K. & QIN, W. (2009) The bugs that came in from the cold: molecular adaptations to low temperatures in insects. *Cellular and Molecular Life Sciences,* 66**,** 1404-1418.

DRICKAMER, K. & DODD, R. B. (1999) C-Type lectin-like domains in *Caenorhabditis elegans:* predictions from the complete genome sequence. *Glycobiology,* 9**,** 1357-69.

DROBNIS, E. Z., CROWE, L. M., BERGER, T., ANCHORDOGUY, T. J., OVERSTREET, J. W. & CROWE, J. H. (1993) Cold shock damage is due to lipid phase transition in cell membranes- a demonstration using sperm as a model. *Journal in experimental zoology,* 265**,** 432-437.

DRUMMOND, A. J., HO, S. Y., PHILLIPS, M. J. & RAMBAUT, A. (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol,* 4**,** e88.

DUMAN, J. G., MORRIS, J. P. & CASTELLINO, F. J. (1984) Purification and composition of an ice nucleating protein from queens of the hornet, *Vespula maculate*. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology,* 154**,** 83-89.

DUMAN, J. G., WU, D. W., XU, L., TURSMAN, D. & OLSEN, T. M. (1991) Adaptations of insects to subzero temperatures. *The Quarterly Review of Biology,* 66**,** 387-410.

DUMAN, J. G. & OLSEN, T. M. (1993) Thermal hysteresis protein activity in bacteria, fungi, and phylogenetically diverse plants. *Cryobiology,* 30**,** 322-328

DUMAN, J. G., LI, N., VERLEYE, D., GOETZ, F. W., WU, D. W., ANDORFER, C. A., BENJAMIN, T. & PARMELEE, D. C. (1998) Molecular characterization and sequencing of antifreeze proteins from larvae of the beetle *Dendroides canadensis*. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology,* 168**,** 225-32.

DUMAN, J. G. (2001) Antifreeze and ice nucleator proteins in terrestrial arthropods. *Annual Review of Physiology,* 63**,** 327-57.

DUMAN, J. G. & SERIANNI, A. S. (2002) The role of endogenous antifreeze protein enhancers in the hemolymph thermal hysteresis activity of the beetle *Dendroides canadensis*. *Journal of Insect Physiology,* 48**,** 103-111.

DUMAN, J. G., BENNETT, V., SFORMO, T., HOCHSTRASSER, R. & BARNES, B. M. (2004) Antifreeze proteins in Alaskan insects and spiders. *Journal of Insect Physiology,* 50**,** 259-66.

EDGAR, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* 32**,** 1792-7.

EDWARDS, A. W. F. (1972) Likelihood. Cambridge: Cambridge University Press

ELBEIN, A. D. (1974) The metabolism of α, α-trehalose. *Advances in Carbohydrate Chemistry and Biochemistry,* 30**,** 227-256.

EVANS, R. P. & FLETCHER, G. L. (2001) Isolation and characterization of type I antifreeze proteins from Atlantic snailfish (*Liparis atlanticus*) and dusky snailfish (*Liparis gibbus*) *Biochimica et Biophysica Acta,* 1547**,** 235-244.

EVANS, R. P. & FLETCHER, G. L. (2004) Isolation and purification of antifreeze proteins from skin tissues of snailfish, cunner and sea raven. *Biochimica et Biophysica Acta,* 1700**,** 209-217.

EWART, K. V., RUBINSKY, B. & FLETCHER, G. L. (1992) Structural and functional similarity between fish antifreeze proteins and calcium-dependent lectins. *Biochemical and Biophysical Research Communications,* 185**,** 335-40.

EWART, K. V. & FLETCHER, G. L. (1993) Herring antifreeze protein: primary structure and evidence for a C-type lectin evolutionary origin. *Molecular Marine Biology and Biotechnology,* 2**,** 20-7.

EWART, K. V., YANG, D. S., ANANTHANARAYANAN, V. S., FLETCHER, G. L. & HEW, C. L. (1996) Ca2+-dependent antifreeze proteins. Modulation of conformation and activity by divalent metal ions. *Journal of Biological Chemistry,* 271**,** 16627-32.

EWART, K. V., LIN, Q. & HEW, C. L. (1999) Structure, function and evolution of antifreeze proteins. *Cellular and Molecular Life Sciences,* 55**,** 271-283.

EWING, B., HILLIER, L., WENDL, M. C. & GREEN, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research,* 8**,** 175-85.

FEDER, M. E. & HOFMANN, G. E. (1999) Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annual Review of Physiology,* 61**,** 243-82.

FEENEY, R. E. (1974) Biological antifreeze. *American Scientist,* 62**,** 712-719.

FEENEY, R. E. & YEH, Y. (1978) Antifreeze proteins from fish bloods. *Advances in Protein Chemistry,* 32**,** 191-282.

FELSENTEIN, J. (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology,* 27**,** 401-410.

FELSENSTEIN, J. (1985) Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution,* 39**,** 783-791.

FENN, J. B., MANN, M., MENG, C. K., WONG, S. F. & WHITEHOUSE, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science,* 246**,** 64-71.

FITCH, W. M. (1970) Toward defining the course of evolution: minimal change for a specific tree topology. *Systematic Zoology,* 20**,** 406-416.

FLETCHER, G. L., HEW, C. L. & DAVIES, P. L. (2001) Antifreeze proteins of teleost fishes. *Annual Review of Physiology,* 63**,** 359-390.

FLOWER, B. P. & KENNETT, J. P. (1993) Middle Miocene ocean-climate transition: High-resolution oxygen and carbon isotopic records from Deep Sea Drilling Project Site 588A, southwest Pacific. *Paleoceanography,* 8**,** 811-843.

FLOYD, R., ABEBE, E., PAPERT, A. & BLAXTER, M. (2002) Molecular barcodes for soil nematode identification. *Molecular Ecology,* 11**,** 839-50.

FORGE, T. A. & MACGUIDWIN, A. E. (1990) Cold hardening of *Meloidogyne hapla* second-stage juveniles. *Journal of Nematology,* 22**,** 101-105.

FRECKMAN, D. W., DEMEURE, Y., MUNNECKE, D. & VAN GUNDY, S. D. (1980) Resistance of Anhydrobiotic *Aphelenchus avenae* to Methyl Bromide Fumigation. *Journal of Nematology,* 12**,** 19-22.

GAL, T. Z., GLAZER, I. & KOLTAI, H. (2004) An LEA group 3 family member is involved in survival of *C. elegans* during exposure to stress. *FEBS Letters,* 577**,** 21-6.

GARNHAM, C. P., GILBERT, J. A., HARTMAN, C. P., CAMPBELL, R. L., LAYBOURN-PARRY, J. & DAVIES, P. L. (2008) A Ca2+-dependent bacterial antifreeze protein domain has a novel beta-helical ice-binding fold. *Biochemical Journal,* 411**,** 171-80.

GARNHAM, C. P., CAMPBELL, R. L. & DAVIES, P. L. (2011) Anchored clathrate water bind antifreeze proteins to ice. *Proceedings of the National Academy of Sciences of United States of America,* 108**,** 7363-7367.

GAUTHIER, S. Y., MARSHALL, C. B., FLETCHER, G. L. & DAVIES, P. L. (2005) Hyperactive antifreeze protein in flounder species. The sole freeze protectant in American plaice. *FEBS Journal,* 272**,** 4439-49.

GAUTHIER, S. Y., SCOTTER, A. J., LIN, F. H., BAARDSNES, J., FLETCHER, G. L. & DAVIES, P. L. (2008) A re-evaluation of the role of type IV antifreeze protein. *Cryobiology,* 57**,** 292-6.

GHEDIN, E., WANG, S., SPIRO, D., CALER, E., ZHAO, Q., CRABTREE, J., ALLEN, J. E., DELCHER, A. L., GUILIANO, D. B., MIRANDA-SAAVEDRA, D., ANGIUOLI, S. V., CREASY, T., AMEDEO, P., HAAS, B., EL-SAYED, N. M., WORTMAN, J. R., FELDBLYUM, T., TALLON, L., SCHATZ, M., SHUMWAY, M., KOO, H., SALZBERG, S. L., SCHOBEL, S., PERTEA, M., POP, M., WHITE, O., BARTON, G. J., CARLOW, C. K., CRAWFORD, M. J., DAUB, J., DIMMIC, M. W., ESTES, C. F., FOSTER, J. M., GANATRA, M., GREGORY, W. F., JOHNSON, N. M., JIN, J., KOMUNIECKI, R., KORF, I., KUMAR, S., LANEY, S., LI, B. W., LI, W., LINDBLOM, T. H., LUSTIGMAN, S., MA, D., MAINA, C. V., MARTIN, D. M., MCCARTER, J. P., MCREYNOLDS, L., MITREVA, M., NUTMAN, T. B., PARKINSON, J., PEREGRIN-ALVAREZ, J. M., POOLE, C., REN, Q., SAUNDERS, L., SLUDER, A. E., SMITH, K., STANKE, M., UNNASCH, T. R., WARE, J., WEI, A. D., WEIL, G., WILLIAMS, D. J., ZHANG, Y., WILLIAMS, S. A., FRASER-LIGGETT, C., SLATKO, B., BLAXTER, M. L. & SCOTT, A. L. (2007) Draft genome of the filarial nematode parasite *Brugia malayi*. *Science,* 317**,** 1756-60.

GIANGROSSI, M., GIULIODORI, A. M., GUALERZI, C. O. & PON, C. L. (2002) Selective expression of the beat-subunit of nucleoid-associated protein HU during cold shock in *Escherichia coli*. *Molecular Microbiology,* 44**,** 205-216.

GILBERT, J. A., HILL, P. J., DODD, C. E. & LAYBOURN-PARRY, J. (2004) Demonstration of antifreeze protein activity in Antarctic lake bacteria. *Microbiology,* 150**,** 171-80.

GILBERT, J. A., DAVIES, P. L. & LAYBOURN-PARRY, J. (2005) A hyperactive, Ca2+-dependent antifreeze protein in an Antarctic bacterium. *FEMS Microbiology Letters,* 245**,** 67-72.

GILMOUR, S. J., SEBOLT, A. M., SALAZAR, M. P., EVERARD, J. D. & THOMASHOW, M. F. (2000) Overexpression of the Arabidopsis CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. *Plant Physiology,* 124**,** 1854-65.

GOLDSTEIN, J., POLLITT, N. S. & INOUYE, M. (1990) Major cold shock protein of *Escherichia coli*. *Proceedings of the National Academy of Sciences of United States of America,* 87**,** 283-287.

GRABHERR, M. G., HAAS, B. J., YASSOUR, M., LEVIN, J. Z., THOMPSON, D. A., AMIT, I., ADICONIS, X., FAN, L., RAYCHOWDHURY, R., ZENG, Q., CHEN, Z., MAUCELI, E., HACOHEN, N., GNIRKE, A., RHIND, N., DI PALMA, F., BIRREN, B. W., NUSBAUM, C., LINDBLAD-TOH, K., FRIEDMAN, N. & REGEV, A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology,* 29**,** 644-52.

GRAETHER, S. P., KUIPER, M. J., GAGNE, S. M., WALKER, V. K., JIA, Z. C., SYKES, B. D. & DAVIES, P. L. (2000) Beta-helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature,* 406**,** 325-328.

GRAHAM, L. A., LIOU, Y. C., WALKER, V. K. & DAVIES, P. L. (1997) Hyperactive antifreeze protein from beetles. *Nature,* 388**,** 727-728.

GRAHAM, L. A., WALKER, V. K. & DAVIES, P. L. (2000) Developmental and environmental regulation of antifreeze proteins in the mealworm beetle *Tenebrio molitor*. *European Journal of Biochemistry,* 267**,** 6452-6458.

GRAHAM, L. A. & DAVIES, P. L. (2005) Glycine-rich antifreeze proteins from snow fleas. *Science,* 310**,** 461-461.

GRAHAM, L. A., QIN, W., LOUGHEED, S. C., DAVIES, P. L. & WALKER, V. K. (2007) Evolution of hyperactive, repetitive antifreeze proteins in beetles. *Journal of Molecular Evolution,* 64**,** 387-98.

GRAHAM, L. A., LOUGHEED, S. C., EWART, K. V. & DAVIES, P. L. (2008a) Lateral transfer of a lectin-like antifreeze protein gene in fishes. *PLoS One,* 3**,** e2616.

GRAHAM, L. A., MARSHALL, C. B., LIN, F. H., CAMPBELL, R. L. & DAVIES, P. L. (2008b) Hyperactive antifreeze protein from fish contains multiple ice-binding sites. *Biochemistry,* 47**,** 2051-63.

GREENWAY, S. C. & STOREY, K. B. (2000) Activation of mitogen-activated protein kinases during natural freezing and thawing in the wood frog. *Molecular and Cellular Biochemistry,* 209**,** 29-37.

GRIFFITH, M., ALA, P., YANG, D. S., HON, W. C. & MOFFATT, B. A. (1992) Antifreeze protein produced endogenously in winter rye leaves. *Plant Physiology,* 100**,** 593-6.

GRIFFITH, M., ANTIKAINEN, M., HON, W.-C., PIHAKASKI-MAUNSBACH, K., YU, X.-M., CHUN, Y. U. & YANG, S. C. (1997) Antifreeze proteins in winter rye. *Physiologia Plantarum,* 100**,** 327-332.

GRIFFITH, M. & YAISH, M. W. (2004) Antifreeze proteins in overwintering plants: a tale of two activities. *Trends Plant Science,* 9**,** 399-405.

GRIMSTONE, A. V., MULLINGER, A. M. & RAMSAY, J. A. (1968) Further studies on the rectal complex of the mealworm *Tenebrio molitor*, L(Coleoptera, Tenebrionidae). *Philosophical Transactions of the Royal Society of London A,* 253**,** 334-382.

GRONWALD, W., LOEWEN, M. C., LIX, B., DAUGULIS, A. J., SONNICHSEN, F. D., DAVIES, P. L. & SYKES, B. D. (1998) The solution structure of type II antifreeze protein reveals a new member of the lectin family. *Biochemistry,* 37**,** 4712-21.

GUALERZI, C. O., GIULIODORI, A. M. & PON, C. L. (2003) Trancriptional and post-transcriptional control of cold-shock genes. *Journal of Molecular Biology,* 331**,** 527-539.

GUBANOV, N. M. (1951) Giant nematoda from the placenta of Cetacea; *Placentonema gigantissima* nov. gen. sp. *Proceedings of the USSR Academy of Sciences,* 77**,** 1123-1125

GUSTA, L. V., BURKE, M. J. & KAPOOR, A. (1975) Determination of unfrozen water in winter cereals at subfreezing temperatures. *Plant Physiology,* 56**,** 707-709.

HAGGERTY, L. S., MARTIN, F. J., FITZPATRICK, D. A. & MCINERNEY, J. O. (2009) Gene and genome trees conflict at many levels. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 364**,** 2209-19.

HAMILTON, H. C., STRICKLAND, M. S., WICKINGS, K., BRADFORD, M. A. & FIERER, N. (2009) Surveying soil faunal communities using a direct molecular approach. *Soil Biology and Biochemistry,* 41**,** 1311-1314.

HAN, E., -N. & BAUCE, E. (1993) Physiological changes and cold hardiness of spruce budworm larvae, *Choristoneura fumiferana,* during pre-diapause and diapause development under laboratory conditions. *The Canadian Entomologist,* 125**,** 1043-1053.

HAND, S. C., JONES, D., MENZE, M. W. & WITT, T. L. (2006) Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. *Biochemical and Biophysical Research Communications,* 348**,** 56-61.

HARA, M. (2010) The multifunctionality of dehydrins: An overview. *Plant Signaling & Behavior,* 5.

HARDING, M. M., WARD, L. G. & HAYMET, A. D. J. (1999) Type I 'antifreeze' proteins - Structure-activity studies and mechanisms of ice growth inhibition. *European Journal of Biochemistry,* 264**,** 653-665.

HAYAKAWA, Y. (1985) Activation of insect fat-body phosphorylase by cold-phosphorylase-kinase, phosphatase and ATP level. *Biochemistry,* 15**,** 123-128.

HAYASHI, M. & WHARTON, D. A. (2011) The oatmeal nematode *Panagrellus redivivus* survives moderately low temperatures by freezing tolerance and cryoprotective dehydration. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology,* 181**,** 335-42.

HAYMET, A. D., WARD, L. G., HARDING, M. M. & KNIGHT, C. A. (1998) Valine substituted winter flounder 'antifreeze': preservation of ice growth hysteresis. *FEBS Letters,* 430.

HAYWARD, S. A., RINEHART, J. P., SANDRO, L. H., LEE, R. E., JR. & DENLINGER, D. L. (2007) Slow dehydration promotes desiccation and freeze tolerance in the Antarctic midge *Belgica antarctica*. *Journal of Experimental Biology,* 210**,** 836-44.

HERDEIRO, R. S., PEREIRA, M. D., PANEK, A. D. & ELEUTHERIO, E. C. (2006) Trehalose protects *Saccharomyces cerevisiae* from lipid peroxidation during oxidative stress. *Biochimica et Biophysica Acta,* 1760**,** 340-6.

HESCHL, M. F. & BAILLIE, D. L. (1990) Functional elements and domains inferred from sequence comparisons of a heat shock gene in two nematodes. *Journal of Molecular Evolution,* 31**,** 3-9.

HEW, C. L., JOSHI, S., WANG, N. C., KAO, M. H. & ANANTHANARAYANAN, V. S. (1985) Structures of shorthorn sculpin antifreeze polypeptides. *European Journal of Biochemistry,* 151**,** 167-72.

HEW, C. L., WANG, N. C., JOSHI, S., FLETCHER, G. L., SCOTT, G. K., HAYES, P. H., BUETTNER, B. & DAVIES, P. L. (1988) Multiple genes provide the basis for antifreeze protein diversity and dosage in the ocean pout, *Macrozoarces americanus*. *Journal of Biological Chemistry,* 263**,** 12049-55.

HOFMANN, G. E. (2005) Patterns of Hsp gene expression in exothermic marine organisms on small to large biogeographic scales. *Integrative and Comparative Biology,* 51**,** 631-640.

HOLDER, M. & LEWIS, P. O. (2003) Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics,* 4**,** 275-84.

HOLMSTRUP, M. & WESTH, P. (1994) Dehydration of earthworm cocoons exposed to cold; a novel cold hardiness mechanism. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology,* 164**,** 312-315.

HOLMSTRUP, M., BAYLEY, M. & RAMLOV, H. (2002) Supercool or dehydrate? An experimental analysis of overwintering strategies in small permeable arctic invertebrates. *Proceedings of the National Academy of Sciences of United States of America,* 99**,** 5716-20.

HOLTERMAN, M., VAN DER WURFF, A., VAN DEN ELSEN, S., VAN MEGEN, H., BONGERS, T., HOLOVACHOV, O., BAKKER, J. & HELDER, J. (2006) Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown Clades. *Molecular Biology and Evolution,* 23**,** 1792-800.

HOLTERMAN, M., HOLOVACHOV, O., VAN DEN ELSEN, S., VAN MEGEN, H., BONGERS, T., BAKKER, J. & HELDER, J. (2008a) Small subunit ribosomal DNA-based phylogeny of basal Chromadoria (Nematoda) suggests that transitions from marine to terrestrial habitats (and vice versa) require relatively simple adaptations. *Molecular Phylogenetics and Evolution,* 48**,** 758-63.

HOLTERMAN, M., RYBARCZYK, K., S, V. D. E., H, V. A. N. M., MOOYMAN, P., SANTIAGO, R. P., BONGERS, T., BAKKER, J. & HELDER, J. (2008b) A ribosomal DNA-based framework for the detection and quantification of

stress-sensitive nematode families in terrestrial habitats. *Molecular Ecology Resources,* 8**,** 23-34.

HON, W. C., GRIFFITH, M., CHONG, P. & YANG, D. (1994) Extraction and Isolation of Antifreeze Proteins from Winter Rye (*Secale cereale* L.) Leaves. *Plant Physiology,* 104**,** 971-980.

HON, W. C., GRIFFITH, M., MLYNARZ, A., KWOK, Y. C. & YANG, D. S. (1995) Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. *Plant Physiology,* 109**,** 879-89.

HONJOH, K. I., MATSUMOTO, H., SHIMIZU, H., OOYAMA, K., TANAKA, K., ODA, Y., TAKATA, R., JOH, T., SUGA, K., MIYAMOTO, T., IIO, M. & HATANO, S. (2000) Cryoprotective activities of group 3 late embryogenesis abundant proteins from *Chlorella vulgaris* C-27. *Bioscience, Biotechnology, and Biochemistry,* 64**,** 1656-63.

HORIKAWA, D. D., IWATA, K., KAWAI, K., KOSEKI, S., OKUDA, T. & YAMAMOTO, K. (2009) High hydrostatic pressure tolerance of four different anhydrobiotic animal species. *Zoological Science,* 26**,** 238-42.

HOSHINO, T., KIRIAKI, M., OHGIYA, S., FUJIWARA, M., KONDO, H., NISHIMIYA, Y., YUMOTO, I. & TSUDA, S. (2003) Antifreeze proteins from snow mold fungi. *Canadian Journal of Botany,* 81**,** 1175-1181.

HSIEH, T. H., LEE, J. T., YANG, P. T., CHIU, L. H., CHARNG, Y. Y., WANG, Y. C. & CHAN, M. T. (2002) Heterology expression of the *Arabidopsis* C-repeat/dehydration response element binding factor 1 gene confers elevated tolerance to chilling and oxidative stresses in transgenic tomato. *Plant Physiology,* 129**,** 1086-94.

HUANG, T. & DUMAN, J. G. (2002) Cloning and characterization of a thermal hysteresis (antifreeze) protein with DNA-binding activity from winter bittersweet nightshade, *Solanum dulcamara*. *Plant Molecular Biology,* 48**,** 339-350.

HUELSENBECK, J. P. & RONQUIST, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics,* 17**,** 754-5.

HUELSENBECK, J. & RANNALA, B. (2004) Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology,* 53**,** 904-13.

HUGOT, J. P., BAUJARD, P. & MORAND, S. (2001) Biodiversity in helminths and nematodes as a field of study: an overview. *Nematology,* 3**,** 199-208.

HUYBRECHTS, P. (2002) Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles. *Quaternary Science Reviews,* 21**,** 203-231.

JAGDALE, G. B. & GREWAL, P. S. (2003) Acclimation of entopathogenic nematodes to novel temperatures: trehalose accumulation and the acquisition of thermotolerance. *International Journal for Parasitology,* 33**,** 145-152.

JAGLO, K. R., KLEFF, S., AMUNDSEN, K. L., ZHANG, X., HAAKE, V., ZHANG, J. Z., DEITS, T. & THOMASHOW, M. F. (2001) Components of the *Arabidopsis* C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in *Brassica napus* and other plant species. *Plant Physiology,* 127**,** 910-7.

JAGLO-OTTOSEN, K. R., GILMOUR, S. J., ZARKA, D. G., SCHABENBERGER, O. & THOMASHOW, M. F. (1998) Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. *Science,* 280**,** 104-6.

JANECH, M. G., KRELL, A. K., MOCK, T., KANG, J.-S. & RAYMOND, J. A. (2006) Ice-binding proteins from sea ice diatoms (bacillariophyceae). *Journal of Phycology,* 42**,** 410-416.

JIA, Z. C., DELUCA, C. I., CHAO, H. M. & DAVIES, P. L. (1996) Structural basis for the binding of a globular antifreeze protein to ice. *Nature,* 384**,** 285-288.

JOANISSE, D. R. & STOREY, K. B. (1994) Enzyme activity profiles in an overwintering population of freeze-tolerant larvae of the gall fly, *Eurosta solidaginis*. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology,* 164**,** 247.

JOANISSE, D. R. & STOREY, K. B. (1995) Temperature acclimation and seasonal responses by enzymes in cold-hardy gall insects. *Archives of Insect Biochemistry and Physiology,* 28**,** 339-349.

JONES, P. G., VANBOGELEN, R. A. & NEIDHART, F. C. (1987) Induction of proteins at low temperature in *Escherichia coli*. *Journal of Bacteriolgy,* 169**,** 2092-2095.

JONES, P. G., KRAH, R., TAFURI, S. R. & WOLFFE, A. P. (1992) DNA gyrase, CS7.4, and the cold shock response in *Escherichia coli*. *Journal of Bacteriology,* 174**,** 5798-802.

JONES, P. G. & INOUYE, M. (1994) The cold-shock response- a hot topic. *Molecular Microbiology,* 11**,** 811-8.

JONES, P. G. & INOUYE, M. (1996) RbfA, a 30S ribosomal binding factor, is a cold-shock protein whose absence triggers the cold-shock response. *Molecular Microbiology,* 6**,** 1207-1218.

JONSSON, K. I. (2003) Causes and consequences of excess resistance in cryptobiotic metazoans. *Physiological and Biochemical Zoology,* 76**,** 429-35.

JOYCE, S. A., REID, A., DRIVER, F. & CURRAN, J. (1994) Application of polymerase chain reaction (PCR) methods to the identification of entomopathogenic nematodes. IN BURNELL, A. M., EHLERS, R.-U. & MASSON, J. E. (Eds.) *COST 812 Biotechnology: Genetics of*

entomopathogenic nematodes-bacterium complexes. Proceedings of symposium and workshop, St. Patrick's College, Maynooth, Co. Kildare, Ireland, Luxembourg, European Commission, DGXII.

KANDROR, O. & GOLDBERG, A. L. (1997) Trigger factor is induced upon cold shock and enhances viability of *Escherichia coli* at low temperatures. *Proceedings of the National Academy of Sciences of United States of America,* 94**,** 4978-4881.

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. *Proceedings of the National Academy of Sciences of United States of America,* 99**,** 9727-9732.

KARNER, M. B., DELONG, E. F. & KARL, D. M. (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature,* 409**,** 507-10.

KASSAHN, K. S., CROZIER, R. H., PORTNER, H. O. & CALEY, M. J. (2009) Animal performace and stress: responses and tolerance limits at different levels of biological organisation. *Biological Reviews,* 84**,** 227-292.

KASUGA, M., LIU, Q., MIURA, S., YAMAGUCHI-SHINOZAKI, K. & SHINOZAKI, K. (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nature Biotechnology,* 17**,** 287-91.

KAUL, S., SHARMA, S. S. & MEHTA, I. K. (2008) Free radical scavenging potential of L-proline:evidence from in vitro assays. *Amino Acids,* 34**,** 315-320.

KAYUKAWA, T., CHEN, B., HOSHIZAKI, S. & ISHIKAWA, Y. (2007) Up-regulation of a desaturase is associated with enhancement of cold hardiness in the onion maggot *Delia antiqua*. *Insect Biochemistry and Molecular Biology,* 37**,** 1160-1167.

KELLER, G., SAHNI, A. & BAJPAI, S. (2009) Deccan volcanism, the KT mass extinction and dinosaurs. *Journal of Biosciences,* 34**,** 709-28.

KENNEDY, B. P., AAMODT, E. J., ALLEN, F. L., CHUNG, M. A., HESCHL, M. F. & MCGHEE, J. D. (1993) The gut esterase gene (ges-1) from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *Journal of Molecular Biology,* 229**,** 890-908.

KENNETT, J. P. (1977) Cenozoic Evolution of Antarctic Glaciation, the Circum-Antarctic Ocean, and Their Impact on Global Paleoceanography. *Journal of Geophysical Research,* 82**,** 3843-3860.

KIKAWADA, T., NAKAHARA, Y., KANAMORI, Y., IWATA, K., WATANABE, M., MCGEE, B., TUNNACLIFFE, A. & OKUDA, T. (2006) Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. *Biochemical Biophysical Research Communications,* 348**,** 56-61.

KIM, J., NUEDA, A., MENG, Y. H., DYNAN, W. S. & MIVECHI, N. F. (1997) Analysis of the phosphorylation of human heat shock transcription factor-1 by MAP kinase family members. *Journal of Cellular Biochemistry,* 67**,** 43-54.

KIM, J. H. (1996) General inconsistency conditions for maximum parsimony: effects of branch lenghts and increasing numbers of taxa. *Systematic Biology,* 45**,** 363-374.

KING, P. A., ROSHOLT, M. N. & STOREY, K. B. (1993) Adaptations of plasma membrane glucose transport facilitate cryoprotectant distribution in freeze-tolerant frogs. *American Journal of Physiology,* 265**,** R1036-42.

KING, P. A., ROSHOLT, M. N. & STOREY, K., M. (1995) Seasonal changes in plasma membrane glucose transport in freeze-tolerant wood frogs. *Canadian Journal of Zoology,* 73**,** 1-9.

KININGHAM, K. & KASARSKIS, E. (1998) Antioxidant function of metallothioneins. *Journal of Trace Elements in Experimental Medicine,* 11**,** 2-3.

KISHINO, H., THORNE, J. L. & BRUNO, W. J. (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution,* 18**,** 352-61.

KLINE, M. P. & MORIMOTO, R. I. (1997) Repression of the heat shock factor 1 transcriptional activation domain is modulated by constitutive phosphorylation. *Molecular and Cellular Biology,* 17**,** 2107-15.

KNIGHT, C. A. (1967) *The freezing of supercooled liquids,* Princeton, NJ: D, Van Nostrand Company Inc.

KNIGHT, C. A., DEVRIES, A. L. & OOLMAN, L. D. (1984) Fish antifreeze protein and the freezing and recrystallization of ice. *Nature,* 308**,** 295-6.

KNIGHT, C. A., CHENG, C. C. & DEVRIES, A. L. (1991) Adsorption of alpha-helical antifreeze peptides on specific ice crystal surface planes. *Biophysical Journal,* 59**,** 409-418.

KNIGHT, H., ZARKA, D. G., OKAMOTO, H., THOMASHOW, M. F. & KNIGHT, M. R. (2004) Abscisic acid induces CBF gene transcription and subsequent induction of cold-regulated genes via the CRT promoter element. *Plant Physiology,* 135**,** 1710-7.

KNOWLTON, N. & WEIGT, L. A. (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London,* 265**,** 1157-2263.

KOBE, B. & DEISENHOFER, J. (1995) Proteins with leucine-rich repeats. *Current Opinion in Structural Biology,* 5**,** 409-16.

KONTOGIORGOS, V., REGAND, A., YADA, R. Y. & GOFF, H. D. (2007) Isolation and characterization of ice structuring proteins from cold-acclimated

winter wheat grass extract for recrystallization inhibition in frozen foods *Journal of Food Biochemistry,* 31**,** 139-160.

KRELL, A., BESZTERI, B., DIECKMANN, G., GLOCKNER, G., VALENTIN, K. & MOCK, T. (2008) A new class of ice-binding proteins discovered in a salt-stress-induced cDNA library of the psychrophilic diatom *Fragilariopis cylindrus*. *European Journal of Phycology,* 43**,** 423-433.

KUIPER, M. J., DAVIES, P. L. & WALKER, V. K. (2001) A theoretical model of a plant antifreeze protein from *Lolium perenne*. *Biophysical Journal,* 81**,** 3560-3565.

KUIPER, M. J., LANKIN, C., GAUTHIER, S. Y., WALKER, V. K. & DAVIES, P. L. (2003) Purification of antifreeze proteins by adsorption to ice. *Biochemical and Biophysical Research Communications,* 300**,** 645-648.

LAEMMILI, U. K. (1970) Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature,* 227**,** 680-685.

LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology,* 10**,** R25.

LARTILLOT, N., LEPAGE, T. & BLANQUART, S. (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics,* 25**,** 2286-8.

LEE, Y. H., HUANG, X. Y., HIRSH, D., FOX, G. E. & HECHT, R. M. (1992) Conservation of gene organization and trans-splicing in the glyceraldehyde-3-phosphate dehydrogenase-encoding genes of Caenorhabditis briggsae. *Gene,* 121**,** 227-35.

LEE, J. K., PARK, K. S., PARK, S., PARK, H., SONG, Y. H., KANG, S. H. & KIM, H. J. (2010) An extracellular ice-binding glycoprotein from an Arctic psychrophilic yeast. *Cryobiology,* 60**,** 222-8.

LEINALA, E. K., DAVIES, P. L., DOUCET, D., TYSHENKO, M. G., WALKER, V. K. & JIA, Z. (2002) A beta-helical antifreeze protein isoform with increased activity. Structural and functional insights. *Journal of Biological Chemistry,* 277**,** 33349-52.

LELIVELT, M. J. & KAWULA, T. H. (1995) Hsc66, an Hsp70 homolog in Escherichia coli, is induced by cold shock but not by heat shock. *Journal of Bacteriolgy,* 177**,** 4900-4907.

LEPAGE, T., BRYANT, D., PHILIPPE, H. & LARTILLOT, N. (2007) A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution,* 24**,** 2669-80.

LETSCH, H. O., KUCK, P., STOCSITS, R. R. & MISOF, B. (2010) The impact of rRNA secondary structure consideration in alignment and tree reconstruction:

simulated data and a case study on the phylogeny of hexapods. *Molecular Biology and Evolution,* 27**,** 2507-21.

LEVITT, J. (1980) *Responses of plants to environmental stresses*, Vol 1, 2nd Edition, Academic, New York

LEWIS, S. C., DYAL, L. A., HILBURN, C. F., WEITZ, S., LIAU, W. S., LAMUNYON, C. W. & DENVER, D. R. (2009) Molecular evolution in *Panagrolaimus* nematodes: origins of parthenogenesis, hermaphroditism and the Antarctic species P. davidi. *BMC Evolutionary Biology,* 9**,** 15.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G. & DURBIN, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, N., ANDORFER, C. A. & DUMAN, J. G. (1998a) Enhancement of insect antifreeze protein activity by solutes of low molecular mass. *Journal of Experimental Biology,* 201**,** 2243-51.

LI, N., CHIBBER, B. A., CASTELLINO, F. J. & DUMAN, J. G. (1998b) Mapping of disulfide bridges in antifreeze proteins from overwintering larvae of the beetle *Dendroides canadensis*. *Biochemistry,* 37**,** 6343-50.

LIN, D. S., TABB, D. L. & YATES III, J. R. (2003) Large-scale protein identification using mass spectrometry. *Biochemica et Biophysica Acta,* 1646**,** 1-10.

LIN, F. H., GRAHAM, L. A., CAMPBELL, R. L. & DAVIES, P. L. (2007) Structural modeling of snow flea antifreeze protein. *Biophysical Journal,* 92**,** 1717-23.

LINDBLOM, T. H. & DODD, A. K. (2006) Xenobiotic detoxification in the nematode *Caenorhabditis elegans*. *Journal of Experimental Zoology Part A: Comparative Experimental Biology,* 305**,** 720-30.

LIOU, Y. C., THIBAULT, P., WALKER, V. K., DAVIES, P. L. & GRAHAM, L. A. (1999) A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle *Tenebrio molitor*. *Biochemistry,* 38**,** 11415-24.

LIOU, Y. C., TOCILJ, A., DAVIES, P. L. & JIA, Z. C. (2000) Mimicry of ice structure by surface hydroxyls and water of a beta-helix antifreeze protein. *Nature,* 406**,** 322-324.

LITVAITIS, M. K., BATES, J. W., HOPE, W. D. & MOENS, T. (2000) Inferring a classification of the Adenophorea (Nematoda) from nucleotide sequences of the D3 expansion segment (26/28S rDNA). *Canadian Journal of Zoology,* 78**,** 911-922.

LOCKHART, D. J., DONG, H., BYRNE, M. C., FOLLETTIE, M. T., GALLO, M. V., CHEE, M. S., MITTMANN, M., WANG, C., KOBAYASHI, M., HORTON, H. & BROWN, E. L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology,* 14**,** 1675-80.

LOGSDON, J. M. & DOOLITTLE, W. F. (1997) Origin of antifreeze protein genes: A cool tale in molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 3485-3487.

LOW, W. K., MIAO, M., EWART, K. V., YANG, D. S., FLETCHER, G. L. & HEW, C. L. (1998) Skin-type antifreeze protein from the shorthorn sculpin, *Myoxocephalus scorpius*. Expression and characterization of a Mr 9, 700 recombinant protein. *Journal of Biological Chemistry,* 273**,** 23098-103.

LOW, W. K., LIN, Q., STATHAKIS, C., MIAO, M., FLETCHER, G. L. & HEW, C. L. (2001) Isolation and characterization of skin-type, type I antifreeze polypeptides from the longhorn sculpin, *Myoxocephalus octodecemspinosus*. *Journal of Biological Chemistry,* 276**,** 11582-9.

MACRAE, M., KRAMER, D. L. & COFFINO, P. (1998) Developmental effect of polyamine depletion in *Caenorhabditis elegans*. *Biochemical Journal,* 333 (Pt 2)**,** 309-15.

MAKI, L. R., GALYAN, E. L., CHANF-CHIEN, M.-M. & CALDWELL, D. R. (1974) Ice nucleation induced by *Pseudomonas syringae*. *Applied Microbiology,* 28**,** 456-459.

MALIK, M. K., SLOVIN, J. P., HWANG, C. H. & ZIMMERMAN, J. L. (1999) Modified expression of a carrot small heat shock protein gene, hsp17. 7, results in increased or decreased thermotolerancedouble dagger. *Plant Journal,* 20**,** 89-99.

MALLO, G. V., KURZ, C. L., COUILLAULT, C., PUJOL, N., GRANJEAUD, S., KOHARA, Y. & EWBANK, J. J. (2002) Inducible antibacterial defense system in *C. elegans*. *Current Biology,* 12**,** 1209-14.

MARENTEZ, E., GRIFFITH, M., MLYNARZ, A. & BRUSH, R. A. (1993) Proteins accumulate in the apoplast of winter rye leaves during cold acclimation. *Physiologia Plantarum,* 87**,** 499-507.

MARGESIN, R., NEUNER, G. & STOREY, K. B. (2007) Cold-loving microbes, plants, and animals--fundamental and applied aspects. *Naturwissenschaften,* 94**,** 77-99.

MARGUERAT, S. & BAHLER, J. (2010) RNA-seq: from technology to biology. *Cellular and Molecular Life Sciences,* 67**,** 569-79.

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y. J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L., JARVIE, T. P., JIRAGE, K. B., KIM, J. B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W., NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J.,

SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. & ROTHBERG, J. M. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-80.

MARIONI, J. C., MASON, C. E., MANE, S. M., STEPHENS, M. & GILAD, Y. (2008a) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research,* 18**,** 1509-17.

MARIONI, J. C., MASON, C. E., MANE, S. M., STEPHENS, M. & GILAD, Y. (2008b) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res,* 18**,** 1509-17.

MARSHALL, C. B., DALEY, M. E., GRAHAM, L. A., SYKES, B. D. & DAVIES, P. L. (2002) Identification of the ice-binding face of antifreeze protein from *Tenebrio molitor*. *FEBS Letters,* 529**,** 261-267.

MARSHALL, C. B., FLETCHER, G. L. & DAVIES, P. L. (2004) Hyperactive antifreeze protein in a fish. *Nature,* 429**,** 153.

MASLEN, N. R. & CONVEY, P. (2006) Nematode diversity and distribution in the southern maritime Antarctic - clues to history? *Soil Biology and Biochemistry,* 38**,** 3141-3151.

MCELWEE, J. J., SCHUSTER, E., BLANC, E., THOMAS, J. H. & GEMS, D. (2004) Shared transcriptional signature in *Caenorhabditis elegans* Dauer larvae and long-lived daf-2 mutants implicates detoxification system in longevity assurance. *Journal of Biological Chemistry,* 279**,** 44533-43.

MELDAL, B. H., DEBENHAM, N. J., DE LEY, P., DE LEY, I. T., VANFLETEREN, J. R., VIERSTRAETE, A. R., BERT, W., BORGONIE, G., MOENS, T., TYLER, P. A., AUSTEN, M. C., BLAXTER, M. L., ROGERS, A. D. & LAMBSHEAD, P. J. (2007) An improved molecular phylogeny of the Nematoda with special emphasis on marine taxa. *Molecular Phylogenetics and Evolution,* 42**,** 622-36.

METZKER, M. L. (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics,* 11**,** 31-46.

MEYER, K., KEIL, M. & NALDRETT, M. J. (1999) A leucine-rich repeat protein of carrot that exhibits antifreeze activity *FEBS Letters,* 447**,** 171-178.

MIDDLETON, A. J., BROWN, A. M., DAVIES, P. L. & WALKER, V. K. (2009) Identification of the ice-binding face of a plant antifreeze protein. *FEBS Letters,* 583**,** 815-819.

MIGUEL, M., JAMES, D., DOONER, H. & BROWSE, J. (1993) *Arabidopsis* requires polyunsaturated lipids for low-temperature survival. *Proceedings of the National Academy of Sciences of United States of America,* 90**,** 6208-6212.

MITREVA, M., WENDL, M. C., MARTIN, J., WYLIE, T., YIN, Y., LARSON, A., PARKINSON, J., WATERSTON, R. H. & MCCARTER, J. P. (2006) Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. *Genome Biology,* 7**,** R75.

MOERMAN, D. G. & FIRE, A. (1977) Muscle: structure, function and development. IN RIDDLE, D. L., BLUMENTHAL, T., MEYER, B.J. AND PREISS, J.R. (Ed.) *C. elegans II.* Cold Spring Harbor Laboratory Press New York

MORITA, R. Y. (1975) Psychrophilic bacteria. *Bacteriological Reviews,* 39**,** 114-167.

MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods,* 5**,** 621-8.

MUGNANO, J., LEE, R. & TAYLOR, R. (1996) Fat body cells and calcium phosphate spherules induce ice nucleation in the freeze-tolerant larvae of the gall fly *Eurosta solidaginis* (Diptera, Tephritidae). *Journal of Experimental Biology,* 199**,** 465-71.

MURATA, N., ISHIZAKINISHIZAWA, Q., HIGASHI, S., TASAKA, Y. & NISHIDA, I. (1992) Genetically engineered alternation in the chilling sensitivity of plants. *Nature,* 356**,** 710-713.

NAGALAKSHMI, U., WANG, Z., WAERN, K., SHOU, C., RAHA, D., GERSTEIN, M. & SNYDER, M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science,* 320**,** 1344-9.

NAKAGAWA, T., IKEHATA, R., UCHINO, M., MIYAJI, T., TAKANO, K. & TOMIZUKA, N. (2006) Cold-active acid beta-galactosidase activity of isolated psychrophilic-basidiomycetous yeast *Guehomyces pullulans*. *Research in Microbiology,* 161**,** 75-9.

NAKAMURA, A., YASUDA, K., ADACHI, H., SAKURAI, Y., ISHLII, N. & GOTO, S. (1999) Vitellogenin-6 is a major carbonylated protein in aged nematode, *Caenorhabditis elegans*. *Biochemical and Biophysical Research Communications,* 264**,** 580-583.

NAKASHIMA, K., KANAMARU, K., MIZUNO, T. & HORIKOSHI, K. (1996a) A novel member of the cspA family of genes that is induced by cold shock in *Escherichia coli*. *Journal of Bacteriolgy,* 178**,** 2994-2997.

NAKASHIMA, K., KANAMARU, K., MIZUNO, T. & HORIKOSHI, K. (1996b) Ribosome-associated protein that inhibits translation at the aminoacyl-tRNA binding stage. *The EMBO Journal,* 2**,** 399-402.

NG, N. F. L. & HEW, C. L. (1992) Structure of an antifreeze polypeptide from sea raven: disulphide bonds and similarity to lectin-binding proteins. *Journal of Biological Chemistry,* 267**,** 16069-16075.

NISHIDA, I. & MURATA, N. (1996) Chilling sensitivity in plants and cyanobacteria: The Crucial Contribution of Membrane Lipids. *Annual Review of Plant Physiology and Plant Molecular Biology,* 47**,** 541-568.

NISHIMIYA, Y., KONDO, H., TAKAMICHI, M., SUGIMOTO, H., SUZUKI, M., MIURA, A. & TSUDA, S. (2008) Crystal structure and mutational analysis of Ca2+-independent type II antifreeze protein from longsnout poacher, *Brachyopsis rostratus*. *Journal of Molecular Biology,* 382**,** 734-46.

O'GRADY, S. M., SCHRAG, J. D., RAYMOND, J. A. & DEVRIES, A. L. (1982) Comparison of antifreeze glycopeptides from arctic and antarctic fishes. *Journal of Experimental Zoology,* 224**,** 177-185.

OKONIEWSKI, M. J. & MILLER, C. J. (2006) Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics,* 7**,** 276.

OSUGA, D. T. & FEENEY, R. E. (1978) Antifreeze glycoproteins from Arctic fish. *Journal of Biological Chemistry,* 253**,** 5338-43.

OVERHOFF, A., FRECKMAN, D. W. & VIRGINIA, R. A. (1971) Life-cycle of the microbivorous Antarctic dry Valley nematode *Scottnema lindsayae* (Timm 1971). *Polar Biology,* 13**,** 151-156.

OZSOLAK, F. & MILOS, P. M. (2011) RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics,* 12**,** 87-98.

PARK, J. I., GRANT, C. M., DAVIES, M. J. & DAWES, I. W. (1998) The cytoplasmic Cu,Zn superoxide dismutase of saccharomyces cerevisiae is required for resistance to freeze-thaw stress. Generation of free radicals during freezing and thawing. *Journal of Biological Chemistry,* 273**,** 22921-8.

PARSELL, D. A. & LINDQUIST, S. (1993) The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annual Reviews Genetics,* 27**,** 437-96.

PELTIER, R., BRIMBLE, M. A., WOJNAR, J. M., WILLIAMS, D. E., EVANS, C. W. & DEVRIES, A. L. (2010) Synthesis and antifreeze activity of fish antifreeze glycoproteins and their analogues. *Chemical Science,* 1**,** 538-551.

PENTELUTE, B. L., GATES, Z. P., TERESHKO, V., DASHNAU, J. L., VANDERKOOI, J. M., KOSSIAKOFF, A. A. & KENT, S. B. (2008) X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. *Journal of the American Chemical Society,* 130**,** 9695-701.

PERRY, R. N. & WHARTON, D. A. (1985) Cold tolerance of hatched and unhatched second stage juveniles of the potato cyst-nematode *Globodera rostochiensis*. *International Journal for Parasitology,* 15**,** 441-445.

PERRY, R. N. (1999) Desiccation survival of parasitic nematodes. *Parasitology,* 119 Suppl**,** S19-30.

PHADTARE, S. & INOUYE, M. (1999) Sequence-selective interactions with RNA by CspB, CspC and CspE, members of the CspA family of Escherichia coli. *Molecular Microbiology,* 33**,** 1004-14.

PHADTARE, S. & SEVERINOV, K. (2010) RNA remodeling and gene regulation by cold shock proteins. *RNA Biology,* 7**,** 788-95.

PLACE, S. P. & HOFMANN, G. E. (2004) Constitutive expression of a stress-inducible heat shock protein gene, hsp70, in phylogenetically distant Antarctic fish. *Polar Biology,* 28**,** 261-267.

POINAR, G. (2003) Trends in the evolution of insect parasitism by nematodes as inferred from fossil evidence. *Journal of Nematology,* 35**,** 129-32.

POINAR, G., JR. & BOUCOT, A. J. (2006) Evidence of intestinal parasites of dinosaurs. *Parasitology,* 133**,** 245-9.

POINAR, G., KERP, H. & HAGEN, H. (2008) *Palaeonema phyticum* gen. n., sp. n. (Nematoda: Palaeonematidae fam. n.), a Devonian nematode associated with early land plants. *Nematology,* 10**,** 9-14.

POWERS, T. O., TODD, T. C., BURNELL, A. M., MURRAY, P. C. B., FLEMMING, C. C., SZALANSKI, A. L., ADAMS, B. A. & HARRIS, T. S. (1997) The rDNA internal transcribed spacer region as a taxonomic marker for nematodes. *Journal of Nematology,* 29**,** 441-450.

POWERS, T. O., NEHER, D. A., MULLIN, P., ESQUIVEL, A., GIBLIN-DAVIS, R. M., KANZAKI, N., STOCK, S. P., MORA, M. M. & URIBE-LORIO, L. (2009) Tropical nematode diversity: vertical stratification of nematode communities in a Costa Rican humid lowland rainforest. *Molecular Ecology,* 18**,** 985-96.

PRASAD, S. S. & BAILLIE, D. L. (1989) Evolutionarily conserved coding sequences in the dpy-20-unc-22 region of *Caenorhabditis elegans*. *Genomics,* 5**,** 185-98.

PRESTRELSKI, S. J., TEDESCHI, N., ARAKAWA, T. & CARPENTER, J. F. (1993) Dehydration-induced conformational transitions in proteins and their inhibition by stabilizers. *Biophysical Journal,* 65**,** 661-71.

PRESTRELSKI, S. J., PIKAL, K. A. & ARAKAWA, T. (1995) Optimization of lyophilization conditions for recombinant human interleukin-2 by dried-state conformational analysis using Fourier-transform infrared spectroscopy. *Pharmaceutical Research,* 12**,** 1250-9.

PUDNEY, P. D., BUCKLEY, S. L., SIDEBOTTOM, C. M., TWIGG, S. N., SEVILLA, M. P., HOLT, C. B., ROPER, D., TELFORD, J. H., MCARTHUR, A. J. & LILLFORD, P. J. (2003) The physico-chemical characterization of a boiling stable antifreeze protein from a perennial grass (*Lolium perenne*). *Archives of Biochemistry and Biophysics,* 410**,** 238-45.

RAISON, J. K. & ORR, G. R. (1990) Proposal for better understanding of the molecular basis og chilling injury. *Chilling Injury of Horticultural Crops.* CRC Press, Boca Raton

RAMSAY, J. A. (1964) The rectal complex of the mealworm *Tenebrio molitor*, L (Coleoptera, Tenebrionidae). *Philosophical Transactions of the Royal Society of London B,* 248**,** 279-314.

RANNALA, B. & YANG, Z. (2008) Phylogenetic inference using whole genomes. *Annual Review of Genomics and Human Genetics,* 9**,** 217-31.

RAYMOND, J. A. & DEVRIES, A. L. (1977) Adsorption inhibition as a mechanism of freezing resistance in polar fishes. *Proceedings of the National Academy of Sciences of United States of America,* 74**,** 2589-93.

RAYMOND, J. A. (1992) Glycerol is a colligative antifreeze in some northern fishes. *Journal in experimental zoology,* 262**,** 347-352.

RAYMOND, J. (1995) Glycerol synthesis in the rainbow smelt *Osmerus mordax*. *Journal of Experimental Biology,* 198**,** 2569-73.

RAYMOND, J. A. & FRITSEN, C. H. (2000) Ice-active substances associated with Antarctic freshwater and terrestrial photosynthetic organisms. *Antarctic science,* 12**,** 418-424.

RAYMOND, J. A. & FRITSEN, C. H. (2001) Semipurification and ice recrystallization inhibition activity of ice-active substances associated with Antarctic photosynthetic organisms. *Cryobiology,* 43**,** 63-70.

RAYMOND, J. A. & KNIGHT, C. A. (2003) Ice binding, recrystallization inhibition, and cryoprotective properties of ice-active substances associated with Antarctic sea ice diatoms. *Cryobiology,* 46**,** 174-181.

RAYMOND, J. A., FRITSEN, C. & SHEN, K. (2007) An ice-binding protein from an Antarctic sea ice bacterium. *FEMS Microbiology Ecology,* 61**,** 214-221.

RAYMOND, J. A., CHRISTNER, B. C. & SCHUSTER, S. C. (2008) A bacterial ice-binding protein from the Vostok ice core. *Extremophiles,* 12**,** 713-7.

RAYMOND, J. A. (2009) Novel ice-binding proteins from a psychrophilic Antarctic alga (Chlamydomonadaceae, Chlorophyceae). *Journal of Phycology*.

REGAND, A. & GOFF, H. D. (2006) Ice recrystallization inhibition in ice cream as affected by ice structuring proteins from winter wheat grass *Journal of Dairy Science,* 89.

RIDDLE, D. L. & ALBERT, P. S. (1997) Genetic and environmental regulation of dauer larva development. IN RIDDLE, D. L., BLUMENTHAL, T., MEYER, B. J. & PRIESS, J. R. (Eds.) *C. elegans II.* Cold Spring Harbor, New York.739-786

RINEHART, J. P., YOCUM, G. D. & DENLINGER, D. L. (2000) Developmental up-regulation of inducible hsp70 transcripts, but not the cognate form, during pupal diapause in the flesh fly, *Sarcophaga crassipalpis*. *Insect Biochemistry and Molecular Biology,* 30**,** 515-521.

RINEHART, J. P., LI, A., YOCUM, G. D., ROBICH, R. M., HAYWARD, S. A. L. & DENLINGER, D. L. (2007) Up-regulation of heat shock proteins is essential for cold survival during insect diapause. *Proceedings of the National Academy of Sciences of United States of America,* 104**,** 11130-11137.

RING, R. A. & DANKS, H. V. (1994) Desiccation and cryoprotection- overlapping adaptations. *Cryo Letters,* 15**,** 181-190.

ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26**,** 139-40.

RODES, D. & HANSON, A. D. (1993) Quaternary ammonium and teriary sulfonium compunds in higher plants. *Annual Review of Plant Physiology and Plant Molecular Biology,* 55**,** 357-384.

ROGRIGUES DA SILVA, N. R., CRISTINA DA SILVA, M., FONSECA GENEVOIS, V., ESTEVES, A. M., DE LEY, P., DECRAEMER, W., RIEGER, T. T. & DOS SANTOS CORREIA, M. T. (2010) Marine nematode taxonomy in the age of DNA: the present and future of molecular tools to assess their biodiversity. *Journal of Nematology,* 12**,** 661-672.

RONQUIST, F. & HUELSENBECK, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19**,** 1572-4.

RUBINSKY, B., ARAV, A., MATTIOLI, M. & DEVRIES, A. L. (1990a) The effect of antifreeze glycopeptides on membrane potential changes at hypothermic temperatures. *Biochemical and Biophysical Research Communications,* 173**,** 1369-74.

RUBINSKY, B., LEE, C. Y., BASTACKY, J. & ONIK, G. (1990b) The process of freezing and the mechanism of damage during hepatic cryosurgery. *Cryobiology,* 27**,** 85-97.

RUDOLPH, A. S. & CROWE, J. H. (1985) Membrane stabilization during freezing: the role of two natural cryoprotectants, trehalose and proline. *Cryobiology,* 22**,** 367-77.

RZHETSKY, A. & NEI, M. (1992) A simple method for estimating and testing minimum-evolution trees. *Molecular Biology and Evolution,* 9**,** 945-967.

SABEHAT, A., LURIE, S. & WEISS, D. (1998) Expression of small heat-shock proteins at low temperatures. *Plant Physiology,* 117**,** 651-658.

SAITOU, N. & NEI, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution,* 4**,** 406-25.

SAKAI, A. (1987) Frost survival of plants:Responses and adaptation to freezing stress. IN LARCHER, W. (Ed.) *Ecological Studies.*, Springer, Berlin Hidelberg, New York

SAKAMOTO, T. & BRYANT, D. A. (1997) Temperature-regulated mRNA accumulation and stabilization for fatty acid desaturase genes in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Molecular Microbiology,* 23**,** 1281-92.

SCHAGGER, H., VON JAGOW, G. (1987) Tricine sodium dodecyl-sulfate polyacrylamide-gel electrophoresis for the separation of proteins in the range from 1-kda to 100-kda. *Analytical Biochemistry,* 166**,** 368-379.

SCHNELL, R. C. & VALI, G. (1972) Atmospheric ice nuclei from decomposing vegetation. *Nature,* 236**,** 163-165.

SCHRAG, J. D., CHENG, C. H., PANICO, M., MORRIS, H. R. & DEVRIES, A. L. (1987) Primary and secondary structure of antifreeze peptides from arctic and antarctic zoarcid fishes. *Biochimica et Biophysica Acta,* 915**,** 357-70.

SCHROEDER, A., MUELLER, O., STOCKER, S., SALOWSKY, R., LEIBER, M., GASSMANN, M., LIGHTFOOT, S., MENZEL, W., GRANZOW, M. & RAGG, T. (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol,* 7**,** 3.

SCHULENBURG, H. & BOEHNISCH, C. (2008) Diversification and adaptive sequence evolution of *Caenorhabditis* lysozymes (Nematoda: Rhabditidae). *BMC Evolutionary Biology,* 8**,** 114.

SCOTT, G. K., FLETCHER, G. L. & DAVIES, P. L. (1986) Fish antifreeze proteins: recent gene evolution. *Canadian Journal of Fisheries and Aquatic Sciences,* 43**,** 1028-1034.

SCOTT, G. K., DAVIES, P. L., SHEARS, M. A. & FLETCHER, G. L. (1987) Structural variation in the alanine-rich antifreeze proteins of the Pleuronectidae. *European Journal of Biochemistry,* 168**,** 629-633.

SCOTT, G. K., HAYES, P. H., FLETCHER, G. L. & DAVIES, P. L. (1988) Wolffish antifreeze protein genes are primarily organized as tandem repeats that each contain two genes in inverted orientation. *Molecular and Cellular Biology,* 8**,** 3670-5.

SCOTTER, A. J., MARSHALL, C. B., GRAHAM, L. A., GILBERT, J. A., GARNHAM, C. P. & DAVIES, P. L. (2006) The basis for hyperactivity of antifreeze proteins. *Cryobiology,* 53**,** 229-239.

SHANNON, A. J., BROWNE, J. A., BOYD, J., FITZPATRICK, D. A. & BURNELL, A. M. (2005) The anhydrobiotic potential and molecular phylogenetics of species and strains of *Panagrolaimus* (Nematoda, Panagrolaimidae). *Journal of Experimental Biology,* 208**,** 2433-45.

SHANNON, A. J. (2007) Cryptobiosis in the genus *Panagrolaimus*. PhD thesis.

SHEPS, J. A., RALPH, S., ZHAO, Z., BAILLIE, D. L. & LING, V. (2004) The ABC transporter gene family of *Caenorhabditis elegans* has implications for the evolutionary dynamics of multidrug resistance in eukaryotes. *Genome Biology,* 5**,** R15.

SHEVCHENKO, A., TOMAS, H., HAVLIS, J., OLSEN, J. V. & MANN, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols,* 1**,** 2856-60.

SHIER, W. T., LIN, Y. & DE VRIES, A. L. (1972) Structure and mode of action of glycoproteins from an antarctic fish. *Biochimica et Biophysica Acta,* 263**,** 406-13.

SHIRAKI, T., KONDO, S., KATAYAMA, S., WAKI, K., KASUKAWA, T., KAWAJI, H., KODZIUS, R., WATAHIKI, A., NAKAMURA, M., ARAKAWA, T., FUKUDA, S., SASAKI, D., PODHAJSKA, A., HARBERS, M., KAWAI, J., CARNINCI, P. & HAYASHIZAKI, Y. (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proceedings of the National Academy of Sciences of United States of America,* 100**,** 15776-81.

SHIVAJI, S., BHADRA, B., RAO, R. S. & PRADHAN, S. (2008) Rhodotorula himalayensis sp. nov., a novel psychrophilic yeast isolated from Roopkund Lake of the Himalayan mountain ranges, India. *Extremophiles,* 12**,** 375-81.

SIDEBOTTOM, C., BUCKLEY, S., PUDNEY, P., TWIGG, S., JARMAN, C., HOLT, C., TELFORD, J., MCARTHUR, A., WORRALL, D., HUBBARD, R. & LILLFORD, P. (2000) Phytochemistry - Heat-stable antifreeze protein from grass. *Nature,* 406**,** 256-256.

SIMMONS, M. P., PICKETT, K. M. & MIYA, M. (2004) How Meaningful Are Bayesian Support Values? *Molecular Biology and Evolution,* 21**,** 188-199.

SIMPSON, D. J., SMALLWOOD, M., TWIGG, S., DOUCET, C. J., ROSS, J. & BOWLES, D. J. (2005) Purification and characterisation of an antifreeze protein from *Forsythia suspensa* (L.). *Cryobiology,* 51**,** 230-4.

SIMPSON, G. G. (1944) *Tempo and Mode in Evolution*, Columbia University Press, New York

SINCLAIR, B. J., ADDO-BEDIAKO, A. & CHOWN, S. L. (2003) Climatic variability and the evolution of insect freeze tolerance. *Biological Reviews of the Cambridge Philosophical Society,* 78**,** 181-95.

SINENSKY, M. (1974) Homeoviscous adaptation- A homeostatic process that regulates the viscosity of membrane lipids in *Escherichia coli*. *Proceedings of the National Academy of Sciences of United States of America,* 71**,** 522-525.

SINGER, M. A. & LINDQUIST, S. (1998) Multiple effects of trehalose on protein folding in vitro and in vivo. *Molecular Cell,* 1**,** 639-48.

SMIRNOFF, N. & CUMBES, Q. J. (1989) Hydroxyl radical scavenging activity of compatible solutes. *Phytochemistry,* 28**,** 1057-1060.

SMITH, T., WHARTON, D. A. & MARSHALL, C. J. (2008) Cold tolerance of an Antarctic nematode that survives intracellular freezing: comparisons with other nematode species. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology,* 178**,** 93-100.

SOHLENIUS, B. (1988) Interactions between two species of *Panagrolaimus* in agar cultures. *Nematologica,* 34**,** 208-217.

SOHLENIUS, B., BOSTROM, S. & HIRSCHFELDER, A. (1996) Distribution patterns of microfauna (nematodes, rotifers and tardigrades) on nunataks in Dronning Maud Land, East Antarctica. *Polar Biology,* 16**,** 191-200.

SONNICHSEN, F. D., SYKES, B. D., CHAO, H. & DAVIES, P. L. (1993) The nonhelical structure of antifreeze protein type III. *Science,* 259**,** 1154-7.

SONNICHSEN, F. D., SYKES, B. D. & DAVIES, P. L. (1995) Comparative modeling of the three-dimensional structure of type II antifreeze protein. *Protein Science,* 4**,** 460-71.

STEEN, H. & MANN, M. (2004) The ABC's (and XYZ's) of peptide sequencing. *Nature Reviews,* 5**,** 699-711.

STEIN, L. D., BAO, Z., BLASIAR, D., BLUMENTHAL, T., BRENT, M. R., CHEN, N., CHINWALLA, A., CLARKE, L., CLEE, C., COGHLAN, A., COULSON, A., D'EUSTACHIO, P., FITCH, D. H., FULTON, L. A., FULTON, R. E., GRIFFITHS-JONES, S., HARRIS, T. W., HILLIER, L. W., KAMATH, R., KUWABARA, P. E., MARDIS, E. R., MARRA, M. A., MINER, T. L., MINX, P., MULLIKIN, J. C., PLUMB, R. W., ROGERS, J., SCHEIN, J. E., SOHRMANN, M., SPIETH, J., STAJICH, J. E., WEI, C., WILLEY, D., WILSON, R. K., DURBIN, R. & WATERSTON, R. H. (2003) The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol,* 1**,** E45.

STOCSITS, R. R., LETSCH, H., HERTEL, J., MISOF, B. & STADLER, P. F. (2009a) Accuarte and efficient reconstruction of deep phylogenies from structured RNAs. *Nucleic Acids Research,* 37**,** 6184-6193.

STOCSITS, R. R., LETSCH, H., HERTEL, J., MISOF, B. & STADLER, P. F. (2009b) Accurate and efficient reconstruction of deep phylogenies from structured RNAs. *Nucleic Acids Res,* 37**,** 6184-93.

STOREY, J., M. & STOREY, K., M. (1985) Triggering of cryoprotectant synthesis by the initiation of ice nucleation in the freeze tolerant frog, *Rana sylvatica. Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology,* 156**,** 191-95.

STOREY, K., M. & STOREY, J., M. (1991) Biochemistry of cryoprotectants. *Insects at Low Temerature.* Chapman & Halll.64-93

STOREY, K. B. & STOREY, J., M. (1996) Natural freezing survival in animals. *Annual Review of Ecology and Systematics,* 27**,** 365-386.

STOREY, K. B. (1997) Organic solutes in freezing tolerance. *Comparative Biochemistry and Physiology,* 117**,** 319-26.

SU, T. P. & HAYASHI, T. (2003) Understanding the molecular mechanism of sigma-1 receptors: towards a hypothesis that sigma-1 receptors are intracellular amplifiers for signal transduction. *Current Medicinal Chemistry,* 10**,** 2073-80.

SWART, A. & HARRIS, J. M. (1996) *Panagrolaimus magnivulvatus* Bostrom, 1995 in nest material, algae and soils from inland nunataks in western Dronning Maud Land, Antarctica. *South African Journal of Zoology,* 31**,** 15-22.

TAKAGI, H. (2008) Proline as a stress protectant in yeast: physiological functions, metabolic regulations, and biotechnological applications. *Applied Microbiology and Biotechnology,* 81**,** 211-23.

TAMURA, K., PETERSON, D., PETERSON, N., STECHER, G., NEI, M. & KUMAR, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*.

TEMPLE, M. D., PERRONE, G. G. & DAWES, I. W. (2005) Complex cellular responses to reactive oxygen species. *Trends in Cell Biology,* 15**,** 319-26.

THE 1000 GENOMES CONSORTIUM (2010) A map of human genome variation from population-scale sequencing. *Nature,* 467**,** 1061-73.

THE *C. ELEGANS* SEQUENCING CONSORTIUM (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science,* 282**,** 2012-2018.

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

THOMASHOW, M. (1999) Plant Cold Acclimation: Freezing Tolerance Genes and Regulatory Mechanisms. *Annual Review of Plant Physiology & Plant Molecular Biology,* 50**,** 571-599.

TIKU, P. E., GRACEY, A. Y., MACARTNEY, A. I., BEYNON, R. J. & COSSINS, A. R. (1996) Cold-induced expression of delta 9-desaturase in carp by transcriptional and posttranslational mechanisms. *Science,* 271**,** 815-8.

TIMASHEFF, S. N. (1992) Water as ligand: Preferential binding and exclusion of denaturants in protein unfolding. *Biochemistry,* 31**,** 9857-9864.

TOMCZAK, M. M., MARSHALL, C. B., GILBERT, J. A. & DAVIES, P. L. (2003) A facile method for determining ice recrystallization inhibition by antifreeze proteins. *Biochemical and Biophysical Research Communications,* 311**,** 1041-1046.

TOONE, W. M., RUDD, K. E. & FRIESEN, J. D. (1991) deaD, a new *Escherichia coli* gene encoding a presumed ATP-dependent RNA helicase, can suppress a mutation in rpsB, the gene encoding ribosomal protein S2. *Journal of Bacteriolgy,* 11**,** 3291-3302.

TRAPNELL, C. & SALZBERG, S. L. (2009) How to map billions of short reads onto genomes. *Nature Biotechnology,* 27**,** 455-457.

TRAPNELL, C., WILLIAMS, B. A., PERTEA, G., MORTAZAVI, A., KWAN, G., VAN BAREN, M. J., SALZBERG, S. L., WOLD, B. J. & PACHTER, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology,* 28**,** 511-5.

TRIPATI, A., BACKMAN, J., ELDERFIELD, H. & FERRETTI, P. (2005) Eocene bipolar glaciation associated with global carbon cycle changes. *Nature,* 436**,** 341-6.

TUNNACLIFFE, A. & LAPINSKI, J. (2003) Resurrecting Van Leeuwenhoek's rotifers: a reappraisal of the role of disaccharides in anhydrobiosis. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 358**,** 1755-71.

TUNNACLIFFE, A. & WISE, M. J. (2007) The continuing conundrum of the LEA proteins. *Naturwissenschaften,* 94**,** 791-812.

TURCHETTI, B., BUZZINI, P., GORETTI, M., BRANDA, E., DIOLAIUTI, G., D'AGATA, C., SMIRAGLIA, C. & VAUGHAN-MARTINI, A. (2008) Psychrophilic yeasts in glacial environments of Alpine glaciers. *FEMS Microbiology Ecology,* 63**,** 73-83.

TYSHENKO, M. G., DOUCET, D., DAVIES, P. L. & WALKER, V. K. (1997) The antifreeze potential of the spruce budworm thermal hysteresis protein. *Nature Biotechnology,* 15**,** 887-890.

TYSON, T., O'MAHONY ZAMORA, G., WONG, S., SKELTON, M., DALY, B., JONES, J. T., MULVIHILL, E. D., ELSWORTH, B., PHILLIPS, M., BLAXTER, M. & BURNELL, A. M. (2012) A molecular analysis of desiccation tolerance mechanisms in the anhydrobiotic nematode Panagrolaimus superbus using expressed sequenced tags. *BMC Res Notes,* 5**,** 68.

URRUTIA, M. E., DUMAN, J. G. & KNIGHT, C. A. (1992) Plant thermal hysteresis proteins. *Biochimica et Biophysica Acta,* 1121**,** 199-206.

VAN MEGEN, H., VAN DEN ELSEN, S., HOLTERMAN, M., KARSSEN, G., MOOYMAN, P., BONGERS, T., HOLOVACHOV, O., BAKKER, J. & HELDER, J. (2009) A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology,* 11**,** 927-950.

VANFLETEREN, J. R., VAN DE PEER, Y., BLAXTER, M. L., TWEEDIE, S. A., TROTMAN, C., LU, L., VAN HAUWAERT, M. L. & MOENS, L. (1994) Molecular genealogy of some nematode taxa as based on cytochrome c and globin amino acid sequences. *Molecular Phylogenetics and Evolution,* 3**,** 92-101.

VEGA, S. E., DEL RIO, A. H., BAMBERG, J. B. & PALTA, J. P. (2004) Evidence for the up-regulation of stearoyl-ACP (A9) desaturase gene expression during cold acclimation. *American Journal of Potato Research,* 81**,** 125-135.

VELCULESCU, V. E., ZHANG, L., VOGELSTEIN, B. & KINZLER, K. W. (1995) Serial analysis of gene expression. *Science,* 270**,** 484-7.

VENKETESH, S. & DAYANANDA, C. (2008) Properties, potentials, and prospects of antifreeze proteins. *Critical Reviews in Biotechnology,* 28**,** 57-82.

VRAIN, T. C., WAKARCHUK, D. A., LEVESQUE, A. C. & HAMILTON, R. I. (1992) Intraspecific rDNA restriction fragment length polymorphisms in the *Xiphinema americanum*
group. *Fundamental and Applied Nematology,* 15**,** 563-573.

WALTERS, K. R., JR., PAN, Q., SERIANNI, A. S. & DUMAN, J. G. (2009a) Cryoprotectant biosynthesis and the selective accumulation of threitol in the freeze-tolerant Alaskan beetle, *Upis ceramboides*. *Journal of Biological Chemistry,* 284**,** 16822-31.

WALTERS, K. R., JR., SERIANNI, A. S., SFORMO, T., BARNES, B. M. & DUMAN, J. G. (2009b) A nonprotein thermal hysteresis-producing xylomannan antifreeze in the freeze-tolerant Alaskan beetle *Upis ceramboides*. *Proceedings of the National Academy of Sciences of United States of America,* 106**,** 20210-5.

WALTERS, K. R., JR., SERIANNI, A. S., VOITURON, Y., SFORMO, T., BARNES, B. M. & DUMAN, J. G. (2011) A thermal hysteresis-producing xylomannan glycolipid antifreeze associated with cold tolerance is found in diverse taxa. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology,* 181**,** 631-40.

WANG, L. & DUMAN, J. G. (2006) A thaumatin-like protein from larvae of the beetle *Dendroides canadensis* enhances the activity of antifreeze proteins. *Biochemistry,* 45**,** 1278-84.

WANG, N., YAMANAKA, K. & INOUYE, M. (1999) CspI, the ninth member of the CspA family of *Escherichia coli*, is induced upon cold shock. *Journal of Bacteriolgy,* 5**,** 1603-1609.

WANG, S., AMORNWITTAWAT, N., JUWITA, V., KAO, Y., DUMAN, J. G., PASCAL, T. A., GODDARD, W. A. & WEN, X. (2009) Arginine, a key residue for the enhancing ability of an antifreeze protein of the beetle *Dendroides canadensis*. *Biochemistry,* 48**,** 9696-703.

WANG, W., VINOCUR, B. & ALTMAN, A. (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta,* 218**,** 1-14.

WANG, W. & WEI, L. (2003) Purification of boiling-soluble antifreeze protein from the legume *Ammopiptanthus mongolicus*. *Preparative Biochemistry and Biotechnology,* 33**,** 67-80.

WANG, X., DEVRIES, A. L. & CHENG, C. H. (1995) Antifreeze peptide heterogeneity in an antarctic eel pout includes an unusually large major variant comprised of two 7 kDa type III AFPs linked in tandem. *Biochimica et Biophysica Acta,* 1247**,** 163-72.

WASMUTH, J. D. & BLAXTER, M. L. (2004) prot4EST: translating expressed sequence tags from neglected genomes. *BMC Bioinformatics,* 5**,** 187.

WELCH, W. J. & BROWN, C. R. (1996) Influence of molecular and chemical chaperones on protein folding. *Cell Stress Chaperones,* 1**,** 109-15.

WEN, D. & LAURSEN, R. A. (1992) A model for binding of an antifreeze polypeptide to ice. *Biophysical Journal,* 63**,** 1659-1662.

WHARTON, D. A. & BROWN, I. M. (1991) Cold tolerance mechanisms of the Antarctic nematode *Panagrolaimus davidi Journal of Experimental Biology,* 155**,** 629-641.

vWHARTON, D. A. & BARCLAY, S. (1993) Anhydrobiosis in the free-living Antarctic nematode *Panagrolaimus davidi. Fundamental and Applied Nematology,* 16**,** 17-22.

WHARTON, D. A. & FERNS, D. J. (1995) Survival of intracellular freezing by the antarctic nematode *Panagrolaimus davidi*. *Journal of Experimental Biology,* 198**,** 1381-1387.

WHARTON, D. A. (1997) Survival of low temperatures by the Antarctic nematode *Panagrolamius davidi* IN LYONS, W. B., HOWARD-WILLIAMS, C. & HAWES, I. (Eds.) *Ecosystem processes in Antarctic ice-free landscapes* Balkema, Rotterdam.57-60

WHARTON, D. A., JUDGE, K. F. & WORLAND, M. R. (2000) Cold acclimation and cryoprotectants in a freeze-tolerant Antarctic nematode, *Panagrolaimus davidi*. *Journal of Comparative Physiology B-Biochemical Systemic and Environmental Physiology,* 170**,** 321-327.

WHARTON, D. A. (2003) The environmental physiology of Antarctic terrestrial nematodes: a review. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology,* 173**,** 621-8.

WHARTON, D. A., GOODALL, G. & MARSHALL, C. J. (2003) Freezing survival and cryoprotective dehydration as cold tolerance mechanisms in the Antarctic nematode Panagrolaimus davidi. *Journal of Experimental Biology,* 206**,** 215-221.

WHARTON, D. A., BARRETT, J., GOODALL, G., MARSHALL, C. J. & RAMLOV, H. (2005) Ice-active proteins from the Antarctic nematode *Panagrolaimus davidi*. *Cryobiology,* 51**,** 198-207.

WIEGMANN, B. M., TRAUTWEIN, M. D., KIM, J. W., CASSEL, B. K., BERTONE, M. A., WINTERTON, S. L. & YEATES, D. K. (2009) Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biology,* 7**,** 34.

WILKE, T., SCHULTHEIB, R. & ALBRECHT, C. (2009) As time goes by: A simple fool's guide to molecular clock approaches in invertebrates. *American Malacological Bulletin,* 27**,** 25-45.

WOLTER, F. P., SCHMIDT, R. & HEINZ, E. (1992) Chilling sensitivity of *Arabidopsis thaliana* with genetically engineered membrane lipids. *EMBO Journal,* 11**,** 4685-92.

WOMERSLEY, C. Z. (1987) A reevaluation of strategies employed by nematode anhydrobiotes in relation to their natural environment. IN VEECH, J. & DICKSON, D. W. (Eds.) *Vistas on Nematology* Hyattsville: Society of Nematologists.165-173

WORRALL, D., ELIAS, L., ASHFORD, D., SMALLWOOD, M., SIDEBOTTOM, C., LILLFORD, P., TELFORD, J., HOLT, C. & BOWLES, D. (1998) A carrot leucine-rich-repeat protein that inhibits ice recrystallization. *Science,* 282**,** 115-117.

WU, C. (1995) Heat shock transcription factors: Structure and regulation. *Annual review of Cell and Developmental Biology,* 11**,** 441-469.

WU, D. W., DUMAN, J. G. & XU, L. (1991) Enhancement of insect antifreeze protein activity by antibodies. *Biochimica et Biophysica Acta,* 1076**,** 416-20.

XIAO, N., SUZUKI, K., NISHIMIYA, Y., KONDO, H., MIURA, A., TSUDA, S. & HOSHINO, T. (2010) Comparison of functional properties of two fungal antifreeze proteins from *Antarctomyces psychrotrophicus* and *Typhula ishikariensis*. *FEBS Journal,* 277**,** 394-403.

YAMAGUCHI-SHINOZAKI, K. & SHINOZAKI, K. (1994) A novel cis-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *Plant Cell,* 6**,** 251-264.

YAMANAKA, K. & INOUYE, M. (2001) Selective mRNA degradation by polynucleotide phosphorylase in cold shock adaptation in *Escherichia coli*. *Journal of Bacteriology,* 183**,** 2808-16.

YAMASHITA, Y., MIURA, R., TAKEMOTO, Y., TSUDA, S., KAWAHARA, H. & OBATA, H. (2003) Type II antifreeze protein from a mid-latitude freshwater fish, Japanese smelt (*Hypomesus nipponensis*). *Bioscience, Biotechnology, and Biochemistry,* 67**,** 461-6.

YAMASHITA, Y., NAKAMURA, N., OMIYA, K., NISHIKAWA, J., KAWAHARA, H. & OBATA, H. (2002) Identification of an antifreeze lipoprotein from *Moraxella* sp. of Antarctic origin. *Bioscience, Biotechnology, and Biochemistry,* 66**,** 239-47.

YANCEY, P. H. (2005) Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *Journal of Experimental Biology,* 208**,** 2819-30.

YANG, Z. & RANNALA, B. (1997) Bayesian phylogenetic inference using DNA sequences: a Markov Chain Carlo Method. *Molecular Biology and Evolution,* 14**,** 717-724.

YANKOFSKY, S. A., LEVIN, Z., BERTOLD, T. & SANDLERMAN, N. (1981) Some basic characterisitics of bacterial freezing nuclei. *Journal of Applied Meteorology,* 20**,** 1013-1019.

YOCHEM, J., TUCK, S., GREENWALD, I. & HAN, M. (1999) A gp330/megalin-related protein is required in the major epidermis of *Caenorhabditis elegans* for completion of molting. *Development,* 126**,** 597-606.

YOCUM, G. D., JOPLIN, K. H. & DENLINGER, D. L. (1998) Up-regulation of a 23 kDa small heat shock protein transcript during pupal diapause in the flesh fly, *Sarcophagan crassipalpis*. *Insect Biochemistry and Molecular Biology,* 28**,** 677-82.

YONG, J., FEI, Y. & WEI, L. (2000) Purification and identification of an antifreeze protein from the leaves of *Ammopiptanthus mongolicus Physiology and Molecular Biology of Plants,* 6**,** 67-73.

YOUNG, M. D., WAKEFIELD, M. J., SMYTH, G. K. & OSHLACK, A. (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology,* 11**,** R14.

YU, S. O., BROWN, A., MIDDLETON, A. J., TOMCZAK, M. M., WALKER, V. K. & DAVIES, P. L. (2010) Ice restructuring inhibition activities in antifreeze proteins with distinct differences in thermal hysteresis. *Cryobiology,* 61**,** 327-34.

ZACHARIASSEN, K. E. & HAMMEL, H. T. (1976) Nucleating agents in the haemolymph of insects tolerant to freezing. *Nature,* 262**,** 285-287.

ZACHARIASSEN, K. E. & KRISTIANSEN, E. (2000) Ice nucleation and antinucleation in nature. *Cryobiology,* 41**,** 257-79.

ZAKHARKIN, S. O., KIM, K., MEHTA, T., CHEN, L., BARNES, S., SCHEIRER, K. E., PARRISH, R. S., ALLISON, D. B. & PAGE, G. P. (2005) Sources of variation in Affymetrix microarray experiments. *BMC Bioinformatics,* 6**,** 214.

ZAMECNIK, J. & JANACEK, J. (1992) Interaction of antifreeze proteins from cold hardened cereal seedlings with ice nucleation active bacteria. *Cryobiology,* 29**,** 718-719.

ZHANG, D. Q., LIU, B., FENG, D. R., HE, Y. M., WANG, S. Q., WANG, H. B. & WANG, J. F. (2004) Significance of conservative asparagine residues in the thermal hysteresis activity of carrot antifreeze protein. *Biochemical Journal,* 377**,** 589-95.

ZHANG, D. Q., WANG, H.-B., BIN, L., FENG, D.-R., HE, Y.-M. & WANG, J.-F. (2006) Carrot antifreeze protein does not exhibit the polygalacturonase-inhibiting activity of PGIP family. *Journal of Genetics and Genomics,* 33**,** 1027-1036.

ZHANG, W. & LAURSEN, R. A. (1998) Structure-function relationships in a type I antifreeze polypeptide. The role of threonine methyl and hydroxyl groups in antifreeze activity. *Journal of Biological Chemistry,* 273**,** 34806-12.

ZHAO, Y. L., ASPINALL, D. & PALEG, L. G. (1992) Protection of membrane integrity in *Medicago sativa* L. by glycine betaine against the effects of freezing. *Journal of Plant Physiology,* 140**,** 541-543.

ZI, Z. J., ZHANG, S. C. & LIU, Q. H. (2008) Vitellogenin functions as a multivalent pattern recognition receptor with an opsonic activity. *PLoS One,* 3**,** 1-7.

ZUCKERKANDL, E. & PAULING, L. (1962) Molecular disease, evolution. and genic heterogeneity IN KASHA, M. & PULLMAN, B. (Eds.) *Horizons in Biochemistry* Academic Press, New York.97-166

ZUCKERKANDL, E. & PAULING, L. (1965) Evolutinary divergence and convergence in proteins. IN BRYSON, V. & VOGEL, H. J. (Eds.) *Evolving Genes and Proteins.* Academic press, New York.97-166