Identifying stress-tolerance genes in hyperarid desert soils using functional metagenomics

by

Gideon Jacobus de Jager

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

In the Faculty of Natural & Agricultural Sciences
Department of Genetics
University of Pretoria
Pretoria
12 February 2015

Supervisor: Prof. Don A. Cowan
Co-supervisor: Dr René Sutherland

Declaration

I, **Gideon Jacobus de Jager** declare that the thesis, which I hereby submit for the degree *Magister Scientiae* at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature:

Date: 12 February 2015

Dedication

This thesis is dedicated to my mom, dad and Mari. Without your support and belief I would not have been able to reach this point.

Acknowledgements

I would like to thank the National Research Foundation (NRF) for providing me with a bursary so I could pursue this degree. The financial assistance of the NRF towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Prof Don Cowan, thank you for your supervision and guidance during this research project and for always expecting the best from me and believing in my abilities. You are an inspiration in both science and life.

Dr René Sutherland, thank you for your hands-on supervision and for your continuing support after your departure, despite having so much else on your plate! Dr Eloy Ferreras and Dr Eldie Berger, thank you for taking over from René as my go-to people when I needed advice about which experimental approaches to use and for your guidance in carrying these out. Dr Leandro Guerrero, Dr Pieter de Maayer and Dr Angel Valverde, thank you for helping me acquire valuable skills in bioinformatic and statistical analyses.

I would like to thank the Sequencing Facility at the University of Pretoria and the ION PGM team for their professionalism and particularly Elritha van Zyl for her help in ensuring I had enough good quality DNA to be sequenced.

My thanks go to the following people for taking the time to read various chapters of my thesis and for giving constructive commentary on these: Dr Eloy Ferreras, Dr Evelien Adriaenssens, Dr Pieter de Maayer, Marc Van Goethem and Andi Wilson.

To everyone at CMEG, thank you for being such a great group of people to work and socialise with! Those wine evenings never ceased to entertain (thanks Prof Cowan for the wine). To Marc, the cricket matches, random challenges, honest trailers, tester, #LabStories and countdown beers at Oom Gert's kept us both sane! Thanks for that.

To my friends and family, thank you for your support and for understanding (or trying to, at least) that I sometimes couldn't attend the numerous social gatherings, but also for twisting my arm on other occasions to join the fun and unwind in the midst of thesis writing!

Andi, thank you for your support and understanding, for always believing in me and my abilities and never even letting me forget what I am capable of achieving also!

Summary

Microorganisms represent an intriguing and underexploited resource for the discovery of novel genes. Extremophiles in environments like the Namib Desert are exposed to multiple stress factors including UV radiation, desiccation, osmotic, oxidative, nutrient and temperature stress. These microorganisms are well adapted to tolerate extended exposure to stressors and rapid change in the environmental conditions. Functional metagenomics is a powerful tool for the discovery of novel genes and enzymes. As sequence based approaches are limited by sequence homologues, functional screening allows for the identification of unique genes involved in stress tolerance.

In this study, novel genes involved in stress tolerance were identified by screening a Namib Desert soil metagenomic fosmid library (>1x10^6 clones) in *Escherichia coli* EPI300 under hyperosmotic (NaCl), oxidative (H_2O_2), heat (46°C) and UVB radiation (280 – 315 nm) stress. Increasing the sodium chloride (NaCl) concentration in growth media to 5% resulted in the identification of 12 salt-tolerant clones. All clones, except one, showed significantly increased growth compared to the host (containing an empty vector) in liquid media augmented with 5% NaCl after 48 hours (*P*-value <0.05). The clones contained fosmids of approximately 42 kilobases (kb) with estimated insert sizes ranging from 30 to 34 kb. Fosmid DNA of salt-tolerant clones was sequenced using the Ion Torrent PGM platform at the University of Pretoria, generating 1.4 Gb of sequence data. Sequences were assembled using MIRA and open reading frames were predicted using the RAST server.

Phylogenetic analysis, using conserved proteins, revealed that the metagenomic DNA originated from members of the phyla Deltaproteobacteria, Fibrobacteres, Cyanobacteria, Actinobacteria and Planctomyces. Six putative salt-tolerance genes were identified through functional predictions based on conserved domains present in the encoded proteins. The proteins; an ABC transporter substrate-binding domain, a RelA-SpoT-like hypothetical protein, an HD-hydrolase domain protein, a cation export system protein, a peptidase M29 and a Na^+/H^+ antiporter, all shared less than 60% amino acid identity with the closest homologues. The low amino acid identity of the proteins encoded on the metagenomic DNA, to known sequences, suggests that the inserts were derived from novel taxa, highlighting the untapped microbial communities present in the Namib Desert. Over-expression of the proteins in *E. coli* BL21(DE3) and subsequent NaCl trials revealed that cells expressing the peptidase M29 showed significantly increased salt-tolerance at 3.5% NaCl compared to the host. Future work will entail investigating the mechanism by which the peptidase M29 confers salt-tolerance to *E. coli*, investigating whether any cross-protection to other abiotic

stresses is conferred and determining the optimal parameters for activity and substrate specificity of the enzyme *in vitro*. The identification of novel stress-tolerance genes and proteins holds potential for application in industrial biotechnology, medicine and agriculture and provides insight into the mechanisms employed by extremophiles to tolerate abiotic stress.

Conference Outputs

Sutherland, R., D. de Jager and D. Cowan. Metagenomic Gene Discovery. *2nd International Congress on Functional Metagenomics*, Pretoria, South Africa. 2 – 5 June 2013.

Sutherland, R., D. de Jager and D. Cowan. Metagenomic Gene Discovery. *18th Biennial Congress of the South African Society for Microbiology (SASM)*, Bela-bela, South Africa. 24 – 27 November 2013.

Sutherland, R., D. de Jager and D. Cowan. Metagenomic Gene Discovery. *Genomics Research Institute Symposium*, University of Pretoria, South Africa. November 2013.

***de Jager, D., R. Sutherland and D. Cowan.** Metagenomic Gene Discovery. *10th International Congress on Extremophiles*, St Petersburg, Russia. 7 – 11 September 2014.

de Jager, D., R. Sutherland and D. Cowan. Identification of novel stress-tolerance genes from a Namib Desert soil metagenomic library, through functional screening. *2nd Joint Congress of the South African Society for Bioinformatics (SASBi) and the South African Genetics Society (SAGS)*, Tshwane, South Africa. 23 – 26 September 2014.

de Jager, D., R. Sutherland and D. Cowan. Identification of novel stress-tolerance genes from a Namib Desert soil metagenomic library, through functional screening. *Genomics Research Institute Symposium*, University of Pretoria, Pretoria, South Africa. 31 October 2014.

de Jager, D., R. Sutherland and D. Cowan. Identification of novel stress-tolerance genes from a Namib Desert soil metagenomic library, through functional screening. *Department of Genetics Post-graduate Symposium*, University of Pretoria, Pretoria, South Africa. 28 November 2014.

*This poster was awarded one of five poster prizes for young scientists at this international congress.

Table of Contents

Declaration ii
Dedication iii
Acknowledgements iv
Summary v
Conference Outputs vii
List of Tables xii
List of Figures xiv
Table of Abbreviations xix
Chapter 1: Literature Review 1
1.1 Introduction 1
1.2 Life in Hyperarid Deserts 3
1.3 Stress Response in Bacteria 5
1.3.1 The General Stress Response in *Escherichia coli* 5
1.3.2 Desiccation and Hyperosmotic Stress 6
1.3.3 Microbial Tolerance to High Temperatures 9
1.3.4 UV Radiation Tolerance 10
1.3.5 Oxidative Stress 12
1.4 Metagenomic Gene Discovery 12
1.4.1 Functional Metagenomic Screening 13
1.4.2 Host-vector Systems 13
1.4.3 Functional Metagenomic Screening Techniques for Stress-related Genes 15
1.4.3.1 Biological Activity or Survival Assays 15
1.4.3.2 Heterologous Complementation 16
1.4.3.3 Activity-based Screening 17
1.4.3.4 SIGEX: Substrate-induced Gene Expression 17
1.4.3.5 Compound Configuration Screening 18
1.5 Conclusion 19
1.6 Research Aims and Objectives 20
1.6.1 Aims 20
1.6.2 Objectives 20
Chapter 2: Materials and Methods 22
2.1 Chemicals and Reagents 22
2.2 Antibiotics and Enzymes 22
2.3 Bacterial Strains and Vectors 23

2.4 General Laboratory Techniques 24
2.4.1 Agarose Gel Electrophoresis 24
2.4.2 DNA Quantification 24
2.4.3 DNA Purification 24
2.5 Preparation of Chemically Competent *E. coli* Cells 24
2.6 Transformation of Chemically Competent *E. coli* Cells 25
2.7 The Namib Desert Soil Metagenomic Library 25
2.8 Survival Assays of the Library Host Strain 26
2.8.1 Hyperosmotic Stress-tolerance Assays 27
2.8.2 Oxidative Stress-tolerance Assays 28
2.8.3 Heat-tolerance assays 28
2.8.4 UVB Radiation-tolerance Assays 28
2.9 Screening the Namib Soil Metagenomic Library for Stress-tolerant Clones 29
2.9.1 Hyperosmotic Stress-tolerance Screening 29
2.9.2 Oxidative Stress-tolerance Screening 29
2.9.3 Heat-tolerance Screening 30
2.9.4 UVB Radiation-tolerance Screening 30
2.10 Plasmid DNA Isolation 30
2.10.1 Alkaline Lysis Miniprep 30
2.10.2 QIAprep Spin Miniprep Kit (Qiagen) 31
2.10.3 Non-ionic Detergent (NID) Plasmid Isolation 31
2.11 Fosmid DNA Restriction Digestion Analysis 32
2.12 Salt-tolerance Growth Experiments 32
2.13 Next Generation Sequencing and Sequence Analysis 33
2.14 Cloning of Putative Stress-tolerance Genes 35
2.14.1 PCR Amplification of Novel Genes 35
2.14.2 Directional Cloning of Novel Genes in pET21a 36
2.15 Bioinformatic Analysis of Putative Stress-tolerance Genes and the Encoded Proteins 37
2.16 Heterologous Protein Expression 38
2.17 *In vivo* Stress-tolerance Screening of Sub-cloned Genes 39
2.17.1 Testing for Salt- and Alkali-tolerance 39
2.17.2 Testing for Heat-tolerance 39
Chapter 3: Functional Screening of a Namib Desert Soil Metagenomic Library for Stress-tolerant Clones 41
3.1 Introduction 41

3.2 Survival Assays of *E. coli* EPI300:pCC1FOS Under Different Abiotic Stresses 41
3.2.1 Hyperosmotic Stress-tolerance Assays 41
3.2.2 Oxidative Stress-tolerance Assays 42
3.2.3 Heat-tolerance Assays 43
3.2.4 UVB Radiation-tolerance Assays 43
3.3 Screening the Namib Desert Soil Metagenomic Library for Stress-resistant Clones .. 44
3.3.1 Hyperosmotic Stress-tolerance Screening 44
3.3.2 Oxidative Stress-tolerance screening 45
3.3.3 Heat-tolerance Screening 46
3.3.4 UVB Radiation-tolerance Screening 47
3.4 Fosmid DNA Isolation 47
3.4.1 Comparison of Three Plasmid DNA Isolation Methods 47
3.5 Extraction of Fosmid DNA from STCs and Restriction Digestion Analysis 49
3.6 STC Growth Rate Analysis 50
3.7 Discussion 53
Chapter 4: Bioinformatic and Phylogenetic Analysis of Metagenomic DNA from Salt-tolerant Clones 55
4.1 Assembly of Sequencing Data 55
4.2 Annotation of STCs and Identification of Putative Stress-tolerance Genes 55
4.3 Phylogenetic Analysis of the Metagenomic Inserts of the STCs 63
4.4 GC Content of the Metagenomic DNA of the STCs 69
4.5 Discussion 70
Chapter 5: Heterologous Expression and *in vivo* Testing of Putative Stress-tolerance Proteins 73
5.1 Molecular Cloning of Stress-related Genes and Bioinformatic Analysis of the Encoded Proteins 73
5.1.1 ABC Transporter Substrate-binding Domain 74
5.1.2 RelA-SpoT-like Hypothetical Protein and the HD-hydrolase Domain Protein 79
5.1.3 Cation Export System Protein 84
5.1.4 Peptidase M29 88
5.1.5 Na^+/H^+ Antiporter 90
5.2 *In vivo* Testing of Putative Stress-tolerance Proteins 94
5.2.1 Heterologous Protein Expression 94
5.2.2 *In vivo* Testing for Stress-tolerance 96
5.3 Discussion 100
Chapter 6: General Discussion and Conclusions 104
References 109

Appendix A **126**
Appendix B **128**
Appendix C **152**

List of Tables

Table 1. Enzymes used in this study. 22
Table 2. Vectors used in this study. 23
Table 3. Metagenomic libraries constructed from Namib Desert soil. 26
Table 4. *Nco*I restriction digests of fosmid DNA. 32
Table 5. Primers used for PCR amplification. Bold nucleotides denote restriction enzyme sites. 35
Table 6. Survival rate of *E. coli* EPI300:pCC1FOS under increasing NaCl concentrations. 42
Table 7. Survival rate of *E. coli* EPI300:pCC1FOS under increasing H_2O_2 concentrations in liquid culture. 42
Table 8. Survival rate of *E. coli* EPI300:pCC1FOS under increasing H_2O_2 concentrations in solid media. 43
Table 9. Survival rate of *E. coli* EPI300:pCC1FOS at high temperatures. 43
Table 10. Survival rate of *E. coli* EPI300:pCC1FOS under increasing exposure time to UVB radiation. 44
Table 11. Survival rate of salt-tolerant cultures on 5% NaCl LB agar. 45
Table 12. Putative heat-tolerant clones identified in the Namib Desert soil metagenomic library. 46
Table 13. STC 63 DNA quantity and quality data obtained after extraction with three different methods. 47
Table 14. Statistical significance of STC growth after 48 hrs determined using Student's unpaired t-test. 51
Table 15. Sequence data obtained for each fosmid before and after quality control (trimming). 56
Table 16. Assembly results obtained through MIRA-Gap5 pipeline. 57
Table 17. Number of genes present and hypothetical proteins encoded on each fosmid. 63
Table 18. The GC content of the metagenomic insert of each STC. 70
Table 19. Stress-related genes sub-cloned from salt-tolerant clones. 73
Table 20. BLASTp analysis of putative stress-related proteins. 74
Table 21. Rare codons present in the ABC SBD gene. 78
Table 22. Rare codons present in the RelA-SpoT gene. 83
Table 23. Rare codons present in the HD-hydrolase gene. 83
Table 24. Consensus sequences of HME signature motifs (adapted from NIES (2003)). 86
Table 25. Rare codons present in the CESP gene. 88
Table 26. Rare codons present in the M29 gene. 90
Table 27. Rare codons present in the Na^+/H^+ antiporter gene. 93

Table 28. *E. coli* BL21(DE) and Rosetta(DE3)pLysS transformants used to test stress-tolerance. 94
Table 29. Statistical significance of survival rates determined using Student's unpaired t-test. 99

List of Figures

Figure 1. Extraction of STC 63 using three different plasmid extraction methods. A) Alkaline lysis method, B) QIAprep Spin Miniprep kit (Qiagen), C) Non-ionic detergent method. Lanes 1, 6 and 11: GeneRuler 1 kb DNA Ladder (Fermentas). Lanes 2, 7 and 12: Lambda DNA (50 ng). Lanes: 3 – 5, 8 – 10 and 13 – 15: STC 63 DNA extractions. 48
Figure 2. *Nco*I digestion of STC 1, STC 10 and pCC1FOS. Lane 1: GeneRuler 1 kb DNA Ladder (Fermentas). Lane 2: Lambda DNA (50 ng). Lane 3: STC 1 extracted using the NID method. Lane 4: STC 10 extracted using the NID method. Lane 5: pCC1FOS extracted using the Qiagen kit. 49
Figure 3. Extraction of fosmid DNA from STC 1 – 15 using the Qiagen Miniprep kit extraction method. Lane 1: Lambda DNA (50 ng). Lanes 2 – 16: STC 1 – 15. 49
Figure 4. *Nco*I restriction digestion analysis of fosmid DNA extracted from STCs 1- 15. Lane 1: Lambda DNA (50 ng). Lane 2: Lambda DNA digested with *HindIII* restriction enzyme. Lane 3: GeneRuler 1 kb DNA Ladder (Fermentas). Lanes 4 – 18: STC 1 – 15. Lane 19: pCC1FOS subjected to *Nco*I digestion. 50
Figure 5. Growth of STCs identified from a Namib Desert soil metagenome and *E. coli* EPI300:pCC1FOS (WT:pCC1FOS) in A) LB broth supplemented with Cm and B) LB broth supplemented with Cm and amended with 5% NaCl. Results are shown as the average of triplicate experiments, with error bars showing the standard error of the mean (SEM). 52
Figure 6. Comparison of the genetic content of STC 1, 9, 12 and 18. Yellow arrows show genes that are shared between all four fosmids. Green arrows show genes that are shared between three fosmids only. Red arrows show genes that are shared between two fosmids only. Blue arrows show genes that are unique to a particular fosmid. Hypothetical proteins were removed from the annotations for visual clarity. Graphical representations are not aligned to scale, which may create the illusion that identical genes are different lengths on different fosmids. 59
Figure 7. PCR amplification of the gene encoding the RelA-SpoT-like protein from STC 1, 9 and 12. Lane 1: Kapa Universal DNA Ladder (Kapa Biosystems). Lane 2: RelA-SpoT amplicon from STC 1. Lane 3: STC 9. Lane 4: STC 12. Lane 5: Negative control. 60
Figure 8. Phylogenetic analysis of ATP-dependent DNA Helicase RecG (representing STC 1, 9, 12 and 18) with closely related proteins. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The tree was rooted with the *Microcytis aeruginosa* branch. The bar represents 0.05 nucleotide substitutions per nucleotide position. 64
Figure 9. Neighbour-joining phylogeny of STC 13 based on the RecA amino acid sequences. The numbers at the nodes indicate the bootstrap percentages based on 1000

replicates (values below 50 were excluded). The tree was rooted with the *Bacteriovorus marinus* branch. 65

Figure 10. Neighbour-joining phylogeny of STC 88 based on the aspartyl t-RNA amidotransferase subunit C amino acid sequences. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The tree was rooted with the *Halalkalibacillus halophilus* branch. The bar represents 0.1 nucleotide substitutions per nucleotide position..... 66

Figure 11. Neighbour-joining phylogeny of STC 89 based on the tyrosyl t-RNA synthetase amino acid sequences. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The *Geobacter metallireducens* branch was used to root the tree..... 67

Figure 12. Neighbour-joining phylogeny of STC 90 based on the arginyl t-RNA synthetase amino acid sequences. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The *Dietzia alimentaria* branch was used to root the tree. 67

Figure 13. Neighbour-joining phylogeny of STC 91 on the basis of the RNA polymerase sigma-70 factor amino acid sequences. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The *Halothermothrix orenii* branch was used to root the tree. 68

Figure 14. Neighbour-joining phylogeny of STC 91 on the basis of the CDP-paratose 2-epimerase amino acid sequences. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The *Desulfuromonas acetoxidans* branch was used to root the tree. 69

Figure 15. PCR amplification of the ABC SBD from STC 1 using iProof High-Fideliy DNA Polymerase (Bio-Rad). Lane 1: KAPA Universal DNA Ladder. Lane 2 – 3: ABC SBD amplicons. Lane 4: Negative control. 74

Figure 16. Restriction enzyme digestion of ABC SBD:pET21a clones with Fastdigest *Nde*I and *Hind*III. The vector backbone can be seen at 5.4 kb and the ABC SBD insert at 1.8 kb. Lane 1: KAPA Universal DNA Ladder. Lane 2 – 5: ABC SBD:pET21a clones. Lane 6 – 7: empty pET21a..... 74

Figure 17. SignalP results for A) the ABC SBD and B) the *E. coli* SBD homologue, showing the absence of a signal peptide in A and the presence of a signal peptide in B. The steep gradient of the S-score (in combination with spikes in the C- and Y-scores) is indicative of the presence of a signal peptide. 76

Figure 18. LipoP results for ABC SBD indicating the predicted signal peptidase (Sp) I and II cleavage sites. 77

Figure 19. The Kyte and Doolittle scale was used to determine the hydrophobic character of the entire ABC SBD protein with a window size of 13. Regions with values above 0 are hydrophobic. 77
Figure 20. PCR amplification using iProof polymerase of RelA-SpoT using primers STC1-2F & STC1-2R (A) and primers STC1-2F and STC1-2.1R (B). The HD-hydrolase including its native promoter was also amplified successfully (B). Lanes 1 and 5: KAPA Universal DNA Ladder. Lanes 2, 3 and 6: RelA-SpoT amplicons. Lanes 4, 7 and 9: Negative controls. Lane 8: HD-hydrolase, including the native promoter, amplicon. 79
Figure 21. Colony PCR screening using DreamTaq DNA Polymerase (Thermo Fisher Scientific) for RelA-SpoT:pET21a recombinant plasmids transformed into *E. coli* JM109 cells. A) RelA-Spot cloned into *Nde*I and *Hind*III sites. Lane 1: KAPA Universal DNA Ladder. Lanes 2 – 12: JM109 transformants. Lane 13: Positive control (STC1 DNA). Lane 14: Negative control. B) RelA-SpoT cloned into *Nde*I and *Eco*RI sites. Lane 15: GeneRuler 1 kb DNA Ladder (Fermentas). Lane 16: Negative control. Lanes 17 – 26: JM109 transformants. Lane 27: Positive control (STC 1 DNA). 79
Figure 22. PCR amplification of RelA-SpoT and HD-hydrolase insert in pET21a. Primers used: STC1-2F and STC1-3R. The expected size of the insert is 2447 bp. Lane 1: GeneRuler 1 kb DNA Ladder. Lanes 2 – 4: Amplicons of RelA-SpoT:HD-hydrolase. Lane 5: Negative control. 80
Figure 23. SMART domain prediction of the RelA-SpoT domain A) in the top BLASTp hit, *Bacteriovorax marinus* hypothetical protein, B) in the protein of interest and C) in the *E. coli* RelA protein. 81
Figure 24. The conserved HD domain of HD-hydrolase proteins. The sequence logo was constructed using a multiple sequence alignment of the amino acid sequences of the protein of interest, the closest two BLASTp matches (*Bdellovibrio bacteriovorus* and *Hydrogenedentes* spp.) and the *E. coli* SpoT protein sequence. Residues are coloured according to chemistry: green – polar, blue – basic, red – acidic, black – hydrophobic. The consensus sequence is shown above the image, where: h – hydrophobic residues (A, C, F, I, L, M, V, W and Y), s – small residues (A, C, S, T, D, N, V, G and P), c – charged residues (D, E, R, K and H), H – histidine, D – aspartate. 82
Figure 25. A) PCR amplification of CESP from STC 13 using iProof Polymerase. Lane 1: KAPA Universal DNA Ladder. Lanes 2 and 3: CESP amplicons. Lane 4: negative control. The arrow indicates the band of interest at the expected size of 3.1 kb. B) CESP amplicons after purification from the agarose gel. Lane 5: KAPA Universal DNA Ladder. Lanes 6 – 8: CESP amplicons. 84
Figure 26. Neighbour-joining phylogeny of CESP and selected RND proteins from each of the eight families. The numbers at the nodes indicate the bootstrap percentages based on

1000 replicates (values below 50 were excluded). The bar represents 0.2 nucleotide substitutions per nucleotide position. The branch representing the RND protein from family 8 was used to root the tree. The tree was constructed using MEGA 5.2 software. 85

Figure 27. Multiple sequence alignment using Clustal Omega of the amino acid sequences of various HME-RND proteins and CESP. The region shown is the signature motif in the fourth transmembrane helix of the proteins. C_metallidurans: *Cupriavidus metallidurans*. Shading indicates 70% conservation, or greater, at a particular position. The arrows indicate conserved aspartate and glutamate residues that are essential for metal resistance (GOLDBERG *et al.* 1999). This figure should be interpreted with the aid of Table 24. 86

Figure 28. A) TMHMM results showing the predicted transmembrane helices (red) of CESP and the perisplasmic (pink line) and cytoplasmic regions (blue line) of the protein. B) The Kyte and Doolittle hydrophobicity plot, with a window size of 13, shows the hydrophobic regions of the protein. Regions above 0 are hydrophobic. The vertical black lines illustrate how the hydrophobic regions of CESP (B) correspond to the location of the transmembrane helices (A). 87

Figure 29. PCR amplification of the M29 gene from STC 90 using iProof Polymerase. Lane 1: KAPA Universal DNA Ladder. Lane 2: M29 amplicon. Lane 3: Negative control. 89

Figure 30. Multiple sequence alignment using Clustal Omega of the amino acid sequences of members of the M29 superfamily of aminopeptidases. A highly conserved portion of the C-terminal end of the proteins is shown. The red arrow indicates the putative Zn^{2+} binding residue. The black arrows indicate the putative bestatin ligation residues. Shading indicates 70% conservation, or greater, at a particular position. 90

Figure 31. PCR amplification of the Na^+/H^+ gene from STC 92 using iProof Polymerase. Lane 1: KAPA Universal DNA Ladder. Lane 2: Na^+/H^+ amplicon. Lane 3: Negative control. 91

Figure 32. A) TMHMM results showing the predicted transmembrane helices (red) of Na^+/H^+ antiporter and the perisplasmic (pink line) and cytoplasmic regions (blue line) of the protein. B) The Kyte and Doolittle hydrophobicity plot, with a window size of 13, shows the hydrophobic regions of the protein. Regions above 0 are hydrophobic. 92

Figure 33. Multiple sequence alignment using Clustal Omega of the amino acid sequences of members of the NhaP group of Na^+/H^+ antiporters. The region containing the conserved aspartate residue is shown. The residue is indicated with an arrow. Shading indicates 75% conservation, or greater, at a particular position. 93

Figure 34. 10% SDS-PAGE analysis of proteins expressed in *E. coli* BL21(DE3) and Rosetta(DE3)pLysS cells. Total protein extract was analysed in each lane. Lane 1: PageRuler Unstained Protein Ladder (Thermo Fisher Scientific). Lane 2: BL21:pET21a. Lane 3: Rosetta:pET21a. Lane 4: BL21:RelA-SpoT. Lane 5: Rosetta:RelA-SpoT. Lane 6:

BL21:ABC SBD. Lane 7: Rosetta:ABC SBD. Lane 8: BL21:CESP. Lane 9: Rosetta:CESP. Arrows on the right indicate the expected position of each protein. 95

Figure 35. 10% SDS-PAGE analysis of proteins expressed in *E. coli* BL21(DE3) cells. Total protein extract was analysed in each lane. Lane 1: PageRuler Unstained Protein Ladder. Lane 2: BL21:pET21a. Lane 3: BL21:RelA-SpoT:HD-hydrolase. Lane 4: BL21:M29. Lane 5: BL21:CESP. Lane 6: BL21: Na^+/H^+ antiporter. Arrows on the right indicate the expected position of each protein. 95

Figure 36. Growth curves of *E. coli* BL21(DE3) cultures containing stress-related genes in LB broth supplemented with 5% NaCl and ampicillin. Results are shown as the average of triplicate experiments, with error bars showing the standard error of the mean (SEM). 96

Figure 37. Growth curves of *E. coli* BL21(DE3) cultures containing stress-related genes in A) LB broth adjusted to pH 9 and supplemented with ampicillin and B) LB broth adjusted to pH 9 and supplemented with 5% NaCl and ampicillin. Results are shown as the average of triplicate experiments, with error bars showing the SEM. 97

Figure 38. Growth curves of *E. coli* BL21(DE3) cultures containing stress-related genes in LB broth supplemented with ampicillin. Results are shown as the average of triplicate experiments, with error bars showing the SEM. 98

Figure 39. Survival rates of *E. coli* BL21(DE3) cultures containing stress-related genes on LB agar supplemented with 3.5% NaCl, ampicillin and 0.4 mM IPTG. Error bars show the SEM. Asterisks show cultures with a statistically significant survival rate (*P*-value <0.05) compared to the control (BL21:pET21a), as determined using Student's unpaired t-test..... 99

Table of Abbreviations

Term	Abbreviation
Acceleration due to gravity	*g*
Adenosine triphosphate	ATP
Ammonium chloride	NH_4Cl
Ampicillin	Ap
Annealing temperature	Tm
Aridity index	*AI*
Base pairs	bp
Basic local alignment search tool	BLAST
Biological soil crust	BSC
Calcium sulphate	$CaSO_4$
Calcium sulphate dehydrate	$CaSO_4.2H_2O$
Chloramphenicol	Cm
Colony forming units	cfu
Degrees Celsius	°C
Deoxynucleotide triphosphate	dNTP
Deoxyribonucleic acid	DNA
Double-strand break	DSB
et alia	et al
Ethylenediaminetetraacetic acid	EDTA
Fluorescence activated cell-sorting	FACS
Gram	g
Green fluorescent protein	GFP
Heat-tolerant clone	HTC
Hours	hrs
Hydrochloric acid	HCl
Isopropyl β-D-thiogalactosidase	IPTG
Kilobase pairs	kb
Kilometre	km
Kyoto encyclopedia of genes and genomes	KEGG
Litre	L
Luria Bertani medium	LB
Magnesium chloride	$MgCl_2$

Microgram	µg
Microlitre	µL
Micromolar	µM
Millilitre	mL
Millimetre	mm
Millimolar	mM
Minute	min
Molar	M
Nanogram	ng
Nanometre	nm
Open reading frame	ORF
Optical density	OD
Polymerase chain reaction	PCR
Potassium acetate	KOAc
Potassium chloride	KCl
Reactive oxygen species	ROS
Ribonucleic acid	RNA
Salt-tolerant clone	STC
Second	s
Single-strand break	SSB
Sodium chloride	NaCl
Sodium dodecyl sulphate	SDS
Sodium hydroxide	NaOH
Species	sp.
Standard error of the mean	SEM
Survival rate	SR
Transporter classification database	TCDB
Tris (hydroxymethyl)-aminomethane	Tris
Tris-acetic acid EDTA	TAE
Ultraviolet	UV
Units of enzyme	U
Volts	V
Volume per volume	v/v
Weight per volume	w/v

Chapter 1: Literature Review

1.1 Introduction

Microbes have evolved to withstand abiotic stresses, which include limited water availability, extreme temperature, high levels of ultraviolet (UV) and ionizing radiation, reactive oxygen species (ROS), high pressure, steep osmotic gradients, high and low pH, organic solvents and high metal ion concentrations (ROTHSCHILD AND MANCINELLI 2001; DOLHI *et al.* 2013; REED *et al.* 2013). These microorganisms are called extremophiles and the Namib Desert in Namibia is an example of an extreme environment.

The Namib Desert is situated along the South-West African coast, spanning from the southern parts of Angola, through Namibia and into the northern parts of South Africa over a length of 1600 kilometres (km) and is between 50 – 120 km wide (ULLMAN AND BUDEL 2003). It is the oldest desert on Earth, having been arid for at least 80 million years (PRESTEL *et al.* 2008). The Namib Desert consists of both dunes (predominantly the northern and southern parts) and gravel plains (predominantly the central region) (ULLMAN AND BUDEL 2003). The desert receives less than 100 millimetres (mm) mean rainfall annually, with fog and dew events along the coast (up to 60 km inland) providing the moisture equivalent of up to 150 mm rainfall annually (ULLMAN AND BUDEL 2003; ECKARDT *et al.* 2012). When the annual rainfall is taken in consideration with the high rates of evapotranspiration as a result of high temperatures (up to 60°C) and wind (PRESTEL *et al.* 2008; WARREN-RHODES *et al.* 2013) the Namib Desert is classified as an arid to hyper-arid desert (ECKARDT *et al.* 2012; POINTING AND BELNAP 2012). Despite such extreme conditions, microbial life exists and is exposed to numerous abiotic stresses (STOMEO *et al.* 2013; WARREN-RHODES *et al.* 2013).

Adaptations of microorganisms to abiotic stress have been intensively investigated (CSONKA 1989; GARCIA-PICHEL AND CASTENHOLZ 1993; POTTS 1994; ULUSU AND TEZCAN 2001; EMPADINHAS AND COSTA 2006; WARREN-RHODES *et al.* 2006; PLEITNER *et al.* 2012) and thoroughly reviewed (COX AND BATTISTA 2005; RODRIGUES AND TIEDJE 2008; DALY *et al.* 2010; RICHTER *et al.* 2010; GAO AND GARCIA-PICHEL 2011; STORZ AND HENGGE 2011). In general, stress adaptations include modifications to the cell membrane and cell wall, increased production of protein chaperones, active uptake/ release of ions and production of secondary metabolites to help negate the effects of stressors. In addition to general responses, certain stressors activate genes or response pathways that produce enzymes and/ or metabolites specific to coping with a particular stress (LOS *et al.* 2008). These stress

adaptations, whether general or specific, were until recently primarily investigated in single isolates of a genus or species from an extreme environment.

Metagenomics, the study of total community DNA, has greatly broadened the scope of stress adaptation studies by allowing researchers to study the DNA and genes of unculturable (or not-yet-culturable) organisms, which represent between 99 and 99.9% of microbial species in an environmental sample (AMANN *et al.* 1995; HANDELSMAN *et al.* 1998; DANIEL 2004). Bioprospecting, the search for novel genes, enzymes and proteins for use in industrial processes (LORENZ AND ECK 2005), has expanded significantly with the advent of metagenomic techniques (PODAR AND REYSENBACH 2006; LEWIN *et al.* 2013). Metagenomes of extremophile communities are of particular interest, since these may encode enzymes that are stable and active in conditions similar to those at which industrial processes are carried out (LORENZ AND ECK 2005; SARETHY *et al.* 2011). While studies identifying enzymes active in extreme conditions are abundant, a limited number of studies have used functional metagenomic techniques to identify genes and proteins that are directly involved in stress adaptation (CULLIGAN *et al.* 2012; VARIN *et al.* 2012). Knowledge gained through metagenomic analyses of stress adaptation mechanisms, particularly in extremophile communities, has potential applications in medicine, agriculture and industrial biotechnology (LORENZ AND ECK 2005; GAO AND GARCIA-PICHEL 2011; ZHONG *et al.* 2012).

This review discusses the varieties of niches colonised by microorganisms in hyperarid deserts, such as the Namib Desert and the adaptations of these microorganisms to extreme conditions typical of hot and hyperarid deserts. Insights into the construction of metagenomic libraries from extreme environments and the careful planning that is needed when deciding on the vector-host system to be used are given. Following this, function-based screening methods for stress-related genes are discussed, with case studies. Where there are no studies available for a particular method, a potential application of this method to stress adaptation is considered. The review is concluded by comments on how functional metagenomic studies can contribute to understanding how microorganisms are adapted to survive in extreme conditions and the potential commercial application of stress-related genes identified from extremophile communities.

1.2 Life in Hyperarid Deserts

Deserts are one of the most extreme and abundant environments on earth, covering up to 35% of the landmass (POINTING AND BELNAP 2012). These include hot (>18°C), cold (<18°C) and polar deserts. A desert is classified as such if the region has an aridity index (*AI*) of <1. The *AI* (developed by the United Nations Environment Program) of a region is determined by calculating the ratio of precipitation to potential evapotranspiration (BARROW 1992). The index is used to further classify deserts as dry-subhumid (*AI*<0.65 to 0.5), semi-arid (*AI*<0.5 to 0.2), arid (*AI*<0.2 to 0.05) and hyperarid (*AI*<0.05) (POINTING AND BELNAP 2012). Examples of hyperarid regions include the hyperarid core of the Atacama Desert in Chile, the Namib Desert in Namibia and the McMurdo Dry Valleys in Antarctica. Most of the moisture that is present in the soil of these hyperarid deserts is not supplied by rainfall or snow, but rather from fog, dew, marine fog or slightly wet soil just above the permafrost layer in the case of polar deserts (CARY *et al.* 2010; POINTING AND BELNAP 2012). The extreme environmental conditions of these hyperarid regions are not limited to low water availability and high or low temperatures. Numerous other abiotic factors make the continued presence of life increasingly challenging, with soil microbial communities being the most well-adapted and successful life forms in these hyperarid regions (COWAN *et al.* 2002; STOMEO *et al.* 2013).

Hyperarid deserts impose various environmental challenges on microbial communities that reside in them. Abiotic stress factors common to most hyperarid regions are desiccation and osmotic stress. The former is due to low water availability as a result of very limited precipitation and high evaporation rates and the latter is due to low water activity (or water potential) as a result of solutes, such as halite (NaCl) and gypsum ($CaSO_4.2H_2O$), dissolved in the water that would be available. High winds intensify desiccation stress by lowering the relative humidity of the immediate atmosphere around microbial communities and oligotrophic soils also limit the growth of these communities (CHAN *et al.* 2012; POINTING AND BELNAP 2012). Furthermore, high levels of UVA and UVB irradiation in hyperarid regions poses a challenge to cells either via the photochemical production of reactive oxygen species (ROS), which oxidise macromolecules, or by causing direct physical damage to macromolecules, such as DNA and proteins (POTTS 1994; POINTING AND BELNAP 2012). Each desert also has abiotic stress factors that are more specific to the region. For example, soil temperatures of above 60°C or below -20°C are common in hot and cold deserts, respectively (POINTING AND BELNAP 2012). Despite such harsh conditions, microbial life still exists in various forms and niches in hyperarid regions around the world.

The soils of hyperarid deserts are generally very limited in carbon sources, with the result that photoautotrophs, such as Cyanobacteria, chlorophytes, lichens and mosses play a dominant role in carbon fixation and support the survival of heterotrophs in these communities (CARUSO *et al.* 2011; POINTING AND BELNAP 2012). Culture-based isolation methods to determine the microbial diversity of these soils have shown that the soils carry a low microbial biomass (CARY *et al.* 2010). However, molecular phylogenetic techniques have shown that hyperarid soils contain a much higher diversity of microbes than previously thought (COWAN *et al.* 2002; STOMEO *et al.* 2013), thus providing evidence for the need to use culture-independent techniques to fully grasp the diversity found in these regions.

Microbial communities colonise various protective niches in hyperarid regions. Translucent rocks (e.g. quartz and marble) provide shelter from the harsh elements, trap moisture underneath the rock and allow photosynthetically active radiation (PAR) to penetrate through to the soil underneath, while filtering out much of the harmful solar radiation (WARREN-RHODES *et al.* 2006; CARY *et al.* 2010). This niche is probably the most favourable for life and therefore communities that live underneath these rocks, called hypoliths, are highly complex compared to other niches in hyperarid deserts and are sources of productivity and biomass (WARREN-RHODES *et al.* 2006). Common taxa found in hypolithic communities are Cyanobacteria, lichens, mosses and heterotrophic bacteria, such as Acidobacteria and Proteobacteria (COWAN *et al.* 2010; POINTING AND BELNAP 2012). Endoliths are communities of microorganisms that occur in pores (cryptoendoliths) and cracks (chasmoendoliths) of rocks, such as sandstone, limestone and granite. The pores and cracks in which endoliths reside also shelter these communities from abiotic stress factors. As with hypolithic communities, endoliths are diverse and complex in community structure and may consist of Cyanobacteria, lichens, algae, fungi and heterotrophic bacteria, such as Proteobacteria (CARY *et al.* 2010; POINTING AND BELNAP 2012). A third niche occupied in hyperarid deserts, although to a lesser extent as hypoliths and endoliths, is on the surface of rocks. These communities, called epiliths, are directly exposed to the extreme abiotic stress factors. The dominant microorganisms in epilithic communities are usually lichens, mosses and free-living microcolonies of Cyanobacteria and fungi (POINTING AND BELNAP 2012).

Open soil communities, termed biological soil crusts (BSCs), have much greater exposure to abiotic stress factors than hypo- and endoliths. Despite this, BSCs cover the majority of the total area of open soil in many deserts (POINTING AND BELNAP 2012). Studies in hot deserts around the world have shown that the dominant phyla of bacteria in these soils are Actinobacteria, Acidobacteria, Bacteroidetes, Alpha- and Betaproteobacteria and Cyanobacteria. Genera that have been identified in hot desert soils include *Deinococcus*,

Rubrobacter, *Nitrosospira*, *Thermobacterium*, *Hymenobacter*, *Microcoleus*, *Nostoc*, *Kineococcus*, *Bacillus*, *Geodermatopilus* and *Scytonema*, to name a few (GARCIA-PICHEL *et al.* 2001; YEAGER *et al.* 2004; RAINEY *et al.* 2005; GUNDLAPALLY AND GARCIA-PICHEL 2006). However, the relative composition of the BSCs can be quite different, even between different sites in the same desert and BSCs may also contain archaeal phylotypes (CARY *et al.* 2010). As a result of the direct exposure of BSCs to extreme abiotic factors, the microorganisms are well adapted to tolerate extended exposure to, and rapid change in, the environmental conditions (CARY *et al.* 2010; POINTING AND BELNAP 2012; STOMEO *et al.* 2013). This makes BSCs in hyperarid deserts an excellent target for the study of the molecular basis of microbial stress tolerance and response.

1.3 Stress Response in Bacteria

1.3.1 The General Stress Response in *Escherichia coli*

Upon entering the stationary growth phase, under starvation conditions or in response to various stresses, *E. coli*, and other Gram-negative bacteria, accumulate the alternative sigma factor, σ^S, or RpoS (BATTESTI *et al.* 2011). RpoS interacts with the core RNA polymerase and directly, or indirectly, regulates the expression of 10% of genes in the *E. coli* genome (WEBER *et al.* 2005; BATTESTI *et al.* 2011). Genes up-regulated as a result of RpoS include various other transcription factors, genes involved in metabolism, membrane transport, stress adaptation, protein processing and a large set of genes with as-yet-unknown function (WEBER *et al.* 2005). Factors that control the expression of RpoS include the bacterial alarmone, ppGpp, which is produced during the stringent response as a result of nutrient depletion (MAGNUSSON *et al.* 2005; BATTESTI *et al.* 2011; BOUVERET AND BATTESTI 2011) and the response regulator cAMP-CRP, which induces transcription of the *rpoS* gene, predominantly in the late exponential phase (HENGGE 2011). The amount of RpoS present in the cell is not only controlled at transcriptional level, but also at post-transcriptional, translational and post-translational levels (HENGGE 2011). It is, in fact, the prevention of RpoS degradation at post-translational level that allows cells to rapidly respond to exposure of a particular stress (HENGGE-ARONIS 2002). Furthermore, the general stress resistance provided by the RpoS-controlled response to carbon deprivation has been shown to confer cross-resistance to other stresses, such as osmotic and oxidative stress, low pH and heat shock (BATTESTI *et al.* 2011). Detailed discussions of the RpoS-mediated general stress response can be found in reviews by BATTESTI *et al.* (2011) and HENGGE (2011). A comprehensive list of genes directly, or indirectly, up-regulated by RpoS, in *E. coli*, following

hyperosmotic shock, low pH and transition from exponential to stationary growth phase can be found in a genome-wide microarray study by WEBER *et al.* (2005).

1.3.2 Desiccation and Hyperosmotic Stress

Desiccation and hyperosmotic stress both involve the removal of water from the cell, but the way in which this is achieved differs. During desiccation, water is lost to the surrounding air, whereas during hyperosmotic stress, water is lost to the surrounding hypertonic solution (BILLI AND POTTS 2002; WOOD 2011). Microorganism in desert soils may experience either of these water stresses. Desiccation stress is likely to be the major stressor to which BSCs are exposed. This is a result of the low precipitation to potential evapotranspiration ratios of this niche. However, during wetting of the soil due to a precipitation event microorganisms can experience hyperosmotic stress. This would be due to the high levels of salts in desert soils dissolving in the water, creating a hypertonic solution surrounding the cells. Therefore, not only do microorganisms in hyperarid desert soils have to be adapted to extended periods of desiccation, but also to rapid rehydration and subsequent hyperosmotic stress (BILLI AND POTTS 2002; POINTING AND BELNAP 2012; RAJEEV *et al.* 2013).

Slade and Radman (2011) defines desiccation as "water content below 0.1 g H_2O g^{-1} dry mass" and it is one of the most severe water stresses that can be imposed on microorganisms (Potts 1994). Water is essential to metabolic processes and in maintaining the structure of proteins, nucleic acids and cell membranes (POTTS 1994). The most highly desiccation resistant organisms enter a state of metabolic dormancy known as the anhydrobiotic state during severe desiccation stress. Examples of such organisms, called anhydrobiotes, include tardigrades, rotifers, nematodes, resurrection plants, yeast and bacterial spores (CROWE *et al.* 2010; SLADE AND RADMAN 2011). The anhydrobiotic states of these organisms are more resistant to extended periods of desiccation than their vegetative counterparts, which remain metabolically active, albeit at a low level (BILLI AND POTTS 2002). Thus, vegetative cells would be an appropriate model to investigate the desiccation tolerance mechanisms of metabolically active microorganisms.

This has led to in-depth studies being conducted on some of the most desiccation and UV-resistant organisms known, including *Deinococcus radiodurans* (SLADE AND RADMAN 2011) and the Cyanobacteria genus *Chroococcidiopsis* (BILLI 2009; BAQUÉ *et al.* 2013). The ability of these bacteria to survive desiccation stress is considered to be due to a combination of their ability to prevent extensive oxidative damage to proteins, caused by ROS released during desiccation, and their ability to repair double-stranded DNA breaks after rehydration

(FREDRICKSON *et al.* 2008; BAQUÉ *et al.* 2013). Oxidative damage is reduced through a high Mn/ Fe ratio, with manganese complexes scavenging ROS most effectively (FREDRICKSON *et al.* 2008). Furthermore, protein denaturation due to desiccation is reduced by the accumulation of compatible solutes, particularly trehalose and sucrose (WELSH AND HERBERT 1999; BILLI AND POTTS 2002; CROWE *et al.* 2010).

Compatible solutes are neutral, highly soluble organic osmolytes that can be accumulated to very high levels in the cell without disrupting cellular functions (KEMPF AND BREMER 1998; WOOD 2011). Trehalose and sucrose are synthesised, or taken up from the environment if available, during desiccation stress and stabilise proteins by being preferentially excluded from the surface of the protein (POTTS 1994). This allows the protein to remain preferentially hydrated and thus remain in its native state. This phenomenon is known as the preferential exclusion hypothesis. Trehalose and sucrose also form glasses (known as vitrification) during desiccation, which reduces the release of ROS by stabilising the dry cytoplasm, thereby indirectly protecting proteins and nucleic acids from oxidative damage (CROWE *et al.* 1998; FRANÇA *et al.* 2007). However, in extreme cases of air drying it is not even possible to maintain a monolayer of water around proteins (POTTS 1994; BILLI AND POTTS 2002). The maintenance of protein structure during desiccation and the protection from oxidation, by compatible solutes and effective ROS scavengers, allows proteins to perform their specific functions, such as DNA repair, immediately upon rehydration of the cell.

DNA damage is repaired through highly accurate and efficient repair mechanisms, restoring the genome to full functionality (SLADE AND RADMAN 2011). Upon rehydration, single stranded breaks are rapidly repaired, thereby preventing any deleterious or lethal effects as the cell resumes growth. Double-stranded breaks (DSBs), if they are present, are slower to repair, but are repaired in a highly accurate homologous recombination process (SLADE AND RADMAN 2011). The effects that desiccation stress has on cells largely overlaps with those imposed by UV radiation and it is suggested that the resistance of organisms such as *D. radiodurans* and *Chroococcidiopsis* to high levels of UV radiation is a cross-protection effect of their evolutionary adaptation to desiccation (BAUERMEISTER *et al.* 2011; BAQUÉ *et al.* 2013). DNA damage caused by desiccation is largely the same in all cells, but it is the ability of resistant microbes to limit damage to proteins and efficiently repair DNA damage that provides them with an advantage over desiccation sensitive organisms (FREDRICKSON *et al.* 2008).

Hyperosmotic stress is the result of the solution surrounding the cells having a very low water potential or high osmolality, causing a net movement of water out of the cells until

equilibrium is reached (WOOD 2011). The osmolality of a solution is influenced by the total amount of solutes dissolved in the solvent. In natural environments these include various salts, such as NaCl and $CaSO_4$ (ionic solvents), as well as various sugars and polyols (non-ionic solvents) (KEMPF AND BREMER 1998; EMPADINHAS AND COSTA 2006). However, organisms are generally classified in terms of their ability to tolerate solutions with high salinity (NaCl), as this salt is common in nature and used to test osmotic stress responses in the laboratory (WOOD 2011). The degree of water loss from the cell as a result of high osmolality, even in the case of extreme halophiles, is not as severe as that which occurs during desiccation (POTTS 1994).

The rapid exodus of water from the cytoplasm during hyperosmotic stress results in shrinkage of the cytoplasm and can lead to plasmolysis as the cell membrane separates from the cell wall. Cell turgidity therefore decreases substantially, which negatively affects the functions of proteins associated with the cell membrane and increases macromolecular crowding due to the reduced cytoplasmic volume, which can exacerbate misfolding of proteins (DIAMANT *et al.* 2001; WOOD 2011). These events are observed during ionic and non-ionic hyperosmotic stress (SHABALA *et al.* 2009). In response to the reduction in cytoplasmic volume and the osmolality difference across the cellular membrane, cells accumulate compatible solutes in the cytoplasm. This is achieved either through active uptake of these molecules from the environment, if available, or *de novo* synthesis from precursors present in all cells (EMPADINHAS AND COSTA 2006). As in desiccation stress, compatible solutes allow for preferential hydration of macromolecules during times of reduced cellular water through the preferential exclusion hypothesis (POTTS 1994). In addition to the sugars accumulated during desiccation stress (trehalose and sucrose), cells may accumulate amino acids and their derivatives (e.g. proline, glycine betaine and ectoine), polyols (e.g. mannitol), and methylsulfonium compounds (e.g. β-dimethylsulfoniopropionate) as compatible solutes in response to hyperosmotic stress (EMPADINHAS AND COSTA 2006). These molecules serve to increase cell turgidity and balance the intra- and extracellular osmolality, thereby restoring function to membrane-bound proteins, reducing macromolecular crowding and preventing the net efflux of water (POTTS 1994; EMPADINHAS AND COSTA 2006; WOOD 2011).

Hyperosmotic stress induced by ionic osmolytes, particularly NaCl, impose additional stresses on cells such as creating an imbalance of ions across the membrane, affecting pH homeostasis and imposing sodium stress on the cells (PADAN AND KRULWICH 2000; PADAN *et al.* 2005; SHABALA *et al.* 2009). It is widely accepted that potassium ions (K^+) are rapidly transported into the cytoplasm upon hyperosmotic shock together with the accumulation of

an organic counterion such as glutamate (McLaggan *et al.* 1994; Sleator and Hill 2002). The primary purpose of this process is to balance the water potential across the cell membrane while compatible solutes are being imported or synthesised (Sleator and Hill 2002; Krämer 2010). This accumulation of K^+ is a transient response in halosensitive prokaryotes. Halotolerant and halophilic prokaryotes maintain a consistently high cytoplasmic concentration of K^+ (in the form of KCl) in what is termed the "salt-in" approach (Sleator and Hill 2002). Therefore, the intracellular ionic strength closely matches that of the high saline environment in which these organisms thrive. Proteins of these organisms have evolved structurally and functionally to maintain their native state and functionality at such high salt concentrations (Sleator and Hill 2002). Halosensitive prokaryotes might not have specific adaptations for high salinity at the protein structure level and thus have to rely more on compatible solutes to maintain protein structure and cation exporters to balance ionic strength across the cell membrane. When exposed to hyperosmotic shock through the addition of NaCl, cells are simultaneously exposed to high levels of sodium (Na^+), which is toxic (Padan and Krulwich 2000). Na^+ transporters are present in multiple copies in halotolerant and halophilic organisms, and hypersaline environments often have an alkaline pH. Thus halotolerant and halophilic organisms possess membrane transporters, such as Na^+/H^+ antiporters that extrude excess Na^+ from the cell, while importing H^+, thus providing combined tolerance to NaCl-induced stress and alkaline pH (Ventosa *et al.* 1998; Padan *et al.* 2005; Yang *et al.* 2006; Wang *et al.* 2013; Zhang *et al.* 2014).

1.3.3 Microbial Tolerance to High Temperatures

Thermophiles have evolved various molecular adaptations to support their survival and growth at temperatures ranging from 50°C to 121°C. An example of this is the high GC content of nucleic acids, which increases the stability of DNA and RNA and thus prevents the unwinding of double-stranded nucleic acids at high temperatures (Trivedi *et al.* 2005; Gerday 2011). Another adaptation of thermophiles is the production of reverse-gyrase enzymes. These introduce positive supercoils into circular DNA, decreasing the possibility of unwinding at high temperatures (Rodríguez and Stock 2002). Proteins have evolved in thermophiles to be more stable at high temperatures by increasing the abundance of charged amino acids, such as glutamic acid, arginine and lysine. These form ionic bonds and salt bridges, which are stronger bonds than the hydrophobic interactions and van der Waals forces provided by non-charged amino acids. The combined strength of the ionic bonds, especially when clustered, is enough to maintain protein structure and function at the high temperatures at which thermophiles live (Goldstein 2007). These adaptations allow proteins to perform their functions at high temperatures and thus further contribute to thermal

tolerance of thermophiles. An extensive review of the adaptations of thermophiles to high temperatures is beyond the scope of this review. Therefore, the reader is directed to the following reviews which discuss the topic in greater detail (GOTTSCHAL AND PRINS 1991; FERRERA AND REYSENBACH 2001; TRIVEDI *et al.* 2005; GOLDSTEIN 2007; CAVA *et al.* 2009; JAENICKE AND STERNER 2013).

The heat shock response occurs in all known living cells when temperatures are raised above the normal growth range (LIM AND GROSS 2011). In *E. coli*, an increase in temperature activates sigma 32 (σ^{32}), a transcription factor that controls the expression of a host of genes known as heat shock proteins (HSPs). The most universal of these are Hsp70 and Hsp60 known as DnaK and GroEL, respectively in prokaryotes (RICHTER *et al.* 2010; VABULAS *et al.* 2010; LIM AND GROSS 2011). These are chaperone proteins that assist in the correct folding of newly produced proteins and refolding of misfolded proteins (MAYER AND BUKAU 2005). Chaperones are not the only proteins activated by σ^{32}. This transcription factor controls a regulon that includes proteins involved in maintaining biological membrane integrity, stabilising nucleic acids, stabilising transcription and translation machinery, transport proteins, other transcription factors and proteases (LIM AND GROSS 2011). The up-regulation of such diverse genes highlights the far-reaching effects of heat shock.

The protein quality control system, which is implicated in the maintenance of the structure and function of proteins and the degradation of misfolded proteins, is of crucial importance in the ability of cells to survive heat shock events (TURGAY 2011). At the centre of the ability of thermophiles to thrive at such high temperatures is the combination of protein, nucleic acid and membrane adaptations. This allows proteins such as transcription factors, chaperones and proteases to carry out their functions more efficiently at high temperatures than their mesophilic relatives (RICHTER *et al.* 2010; VABULAS *et al.* 2010). Furthermore, without these adaptations it would not be possible for these bacteria to tolerate other stress factors frequently encountered in extreme environments, such as high levels of UV radiation, desiccation, osmotic and oxidative stress.

1.3.4 UV Radiation Tolerance

Although ultraviolet (UV) radiation is only a small part of the total solar radiation that reaches the surface of the earth, it can cause severe damage to living organisms (GAO AND GARCIA-PICHEL 2011). UV radiation can be separated based on which region of the electromagnetic spectrum the wavelengths fall into: UVA (315 – 400 nm), UVB (280 – 315 nm) and UVC (100 – 280 nm), but only UVA and UVB reach the surface of the earth, as UVC is absorbed by the

ozone layer (GAO AND GARCIA-PICHEL 2011). As a result of reduced water vapour in the atmosphere, hyperarid deserts around the world receive much higher doses of UVA and UVB radiation than other biomes (POINTING AND BELNAP 2012). UV radiation is absorbed by various molecules in cells, such as proteins, nucleic acids and metabolically active small molecules. The absorption of UVA by these small molecules can produce ROS, which oxidise various targets within the cell (DALY 2012). However, as a result of the shorter wavelength and higher energy state of UVB, it causes direct damage to proteins and DNA. For example the formation of thymine dimers and single-strand breaks (SSBs) in DNA can interrupt transcription and DSBs can result in lethal chromosomal mutations (DALY 2009).

In order to mitigate the deleterious effects of both UVA and UVB radiation, microbial communities and microorganisms implement various strategies. Antioxidants may be produced to reduce the number of ROS, growth in biofilms or colonies provides protection by sacrificing surface layers of cells and secondary metabolites that absorb UV radiation may be produced (GARCIA-PICHEL AND CASTENHOLZ 1991; GARCIA-PICHEL AND CASTENHOLZ 1993; GAO AND GARCIA-PICHEL 2011). These metabolites, called 'microbial sunscreens,' include scytonemin, mycosporines, melanins and carotenoids (GAO AND GARCIA-PICHEL 2011). Although these strategies limit the deleterious effects of UV radiation, protein and DNA damage is still incurred by BSCs.

The ability of an organism to efficiently and accurately repair DNA breaks is fundamental to its ability to survive UV radiation stress (MINTON 1996). Under normal conditions SSBs and DSBs in DNA are efficiently repaired, thus preventing any deleterious or lethal mutations from occurring. However, UVA and UVB damage occurs simultaneously. This means that before DNA breaks can be repaired, the proteins that are meant to repair them become oxidised by ROS resulting from UVA radiation, which results in cell death (KRISKO AND RADMAN 2010). Even though the same number of DNA breaks may occur in radiation-sensitive and radiotolerant bacteria under UV stress, it is the ability of radiotolerant bacteria to repair these breaks that provides them with the survival advantage (DALY 2009; SLADE AND RADMAN 2011; DALY 2012). It is now thought that protein, and not DNA, is the primary critical target of UV radiation and desiccation stress (DALY 2009). Both of these factors cause oxidative damage to macromolecules and the ability of radiotolerant and desiccation-tolerant organisms to reduce oxidative damage to proteins allows efficient DNA repair and thus survival of UV radiation stress (DALY *et al.* 2007; FREDRICKSON *et al.* 2008; SLADE AND RADMAN 2011; DALY 2012).

1.3.5 Oxidative Stress

Oxidative stress is considered to be a unifying stress under conditions typical to hyperarid regions around the world. During desiccation and UV stress, ROS are produced that react with protein side chains, which result in reduction, or loss, of function. The three most common ROS are superoxide radicals (O_2^-), hydroxyl radicals ($OH^\bullet$) and hydrogen peroxide (H_2O_2) (SLADE AND RADMAN 2011). These molecules oxidise carbonyl groups on proteins, in a process called carbonylation, which is the most common form of oxidative damage in cells and results in decreased enzyme activity (SLADE AND RADMAN 2011). The radiotolerant *D. radiodurans* minimises ROS damage by producing enzymes that convert these radicals into less harmful molecules. Examples of such enzymes are catalase, superoxide dismutase (SOD) and hydrogen peroxidase. These enzymes, produced by *D. radiodurans*, are up to 17 times more efficient at scavenging ROS than their counterparts produced by radiation-sensitive *E. coli* (SLADE AND RADMAN 2011). Furthermore, *D. radiodurans* uses manganese ions (Mn^{2+}) in complexes and carotenoids to help protect its proteins and DNA from oxidative damage. Even though UV, desiccation and oxidative stress response pathways have been thoroughly investigated in *D. radiodurans*, much research still needs to be done in hyperarid microbial communities, such as BSCs, to elucidate how these organisms survive the simultaneous onslaught of abiotic stresses (CARY *et al.* 2010; POINTING AND BELNAP 2012; VERDE *et al.* 2012).

1.4 Metagenomic Gene Discovery

Metagenomics is the study of the combined DNA of all microorganisms in a particular environment (HANDELSMAN *et al.* 1998). It involves the extraction of the total DNA from an environmental sample and analysing the total organismal diversity and function. This method was developed because culture methods were unable to capture the true microbial diversity present in environmental samples (COWAN *et al.* 2002). Metagenomics have the advantage of being culture independent and therefore provide potential for the discovery of novel genes or operons from bacteria that have not been classified or cultured.

The search for such novel genes, operons or pathways is called metagenomic gene discovery (COWAN *et al.* 2005). This is achieved by cloning the total DNA isolated from an environmental sample into suitable cloning vectors, such as bacterial artificial chromosomes (BACs), cosmids, fosmids and plasmids, creating a metagenomic library (RONDON *et al.* 2000; ENTCHEVA *et al.* 2001; GABOR *et al.* 2004). The metagenomic library can be screened

for novel genes in several ways. Firstly, an *in silico* approach can be followed by mining sequence data for genes that are homologous to those in available databases and can thereby predict gene products, protein structure and function (TUFFIN *et al.* 2009). The disadvantage of this approach is that identification of entirely novel genes is improbable, as the discovery of functional genes or operons is dependent on what is already available in the databases. The second method for the discovery of novel genes from metagenomes, discussed below, is functional screening.

1.4.1 Functional Metagenomic Screening

Functional metagenomic screening is a useful tool for discovering completely novel genes or classes of genes, as the process is not dependent on sequences of known genes deposited in databases (TUFFIN *et al.* 2009). However, it is highly dependent on the ability of the heterologous host to produce a functional product from the metagenomic DNA harboured in the vector, highlighting the importance of the host-vector system used in functional metagenomic screening (EKKERS *et al.* 2012). Nonetheless, many novel genes have been discovered from extreme and other environments in this way (UCHIYAMA AND MIYAZAKI 2009).

1.4.2 Host-vector Systems

The heterologous host used during functional screening of metagenomic libraries is arguably the most important factor affecting screening efficiency. The choice of which host to use should ideally be based on the dominant bacterial phylum present in the environment from which the metagenome was constructed (EKKERS *et al.* 2012). This increases the probability that the host will recognise foreign promoters and express foreign genes. Codon bias discrepancy is minimised and correct folding of proteins is more likely, thus increasing the likelihood of expressing functional proteins. In screening for high temperature stress adaptation, the use of an extremophile, such as *Thermus thermophilus*, as an expression host has various advantages. This was demonstrated in a study by ANGELOV *et al.* (2009), where 10 unique clones showing xylanase activity were identified by functional screening in *T. thermophilus*. Only two unique xylanase active clones were identified in *E. coli*, with 10 clones harbouring the same fosmid identified in both hosts. The advantage of using an extremophile as expression host could be due to several reasons. Extremophile transcription machinery, such as polymerases and sigma factors, are well adapted to extreme conditions (STORZ AND HENGGE 2011). These are more likely (than mesophilic proteins) to be functional when screening metagenomic libraries at high temperatures. Similarly, translational and

post-translational machinery, such as ribosomes and chaperones, are well adapted to high temperatures. Thus, the expression of functionally active stress-related proteins is more likely. The study by ANGELOV *et al.* (2009) not only illustrates the advantage of using an extremophilic host, but shows that using different hosts can increase the overall hit-rate of metagenomic screening.

Several recent studies have reported the use of broad-host-range vectors functional across species and phylum barriers. AAKVIK *et al.* (2009) constructed an RK2-based plasmid that functioned as both a fosmid and a bacterial artificial chromosome (BAC). The plasmid, pRS44/pTA44, was designed using the pCC1FOS fosmid (Epicentre Biotechnologies) as the parental plasmid. The broad-host-range vector was shown to be stable and functional when transferred as a fosmid from *E. coli* to *Pseudomonas fluorescens* and *Xanthomonas campestris* with an average insert size of 35 kb. As a BAC the vector required a helper plasmid, pRS48, for replication support and was stable and functional when transferred to *P. fluorescens* with insert sizes of up to 200 kb. This study showed that large-insert plasmids can be stably maintained across species barriers, an important aspect for broad-range functional metagenomic screening. Another study by CRAIG *et al.* (2010) made use of an IncP1-α group (also possessing the RK2 replicon) broad-host-range cosmid cloning vector, pJWC1, to illustrate the ability of six different Proteobacteria to express metagenomic DNA. It was found that the six different hosts showed greatly varied phenotypic outputs for the three traits investigated: antibacterial activity, colony morphology and pigmentation, with very little overlap. This study showed that species from the same phylum will not necessarily serve as equally efficient functional screening hosts and that successful screening does not only depend on the host, but also on the target of the functional screen.

To extend this expression technology to Gram-positive organisms, broad-host-range plasmids (pEBP) that were stable and active in *E. coli*, *Pseudomonas putida* and *Bacillus subtilis*, a Gram-positive bacterium belonging to the phylum Firmicutes, were constructed (TROESCHEL *et al.* 2012). These vectors made it possible to screen metagenomic libraries simultaneously in both Gram-negative and Gram-positive bacteria. This is advantageous as genetic and physiological backgrounds differ substantially between species and thus where one species fails to express a heterologous protein, another species might be successful (TROESCHEL *et al.* 2012).

Another major factor that influences the ability of the heterologous host to express genes in the metagenomic insert is the presence of transcription termination signals (GABOR *et al.* 2004). When a transcription terminator is located between the vector promoter and the first

gene in the insert, transcription of this gene will not occur (GABOR *et al.* 2004) and if the host transcription machinery does not recognise the exogenous promoters of the downstream genes then these will also not be expressed. The same problem arises when using substrate-induced gene expression (SIGEX – see below), as a termination signal could be located between a gene being expressed and the green fluorescent protein (GFP) on the vector (YUN AND RYU 2005). Thus, a positive clone would not be detected. These issues were addressed by TERRON-GONZALEZ *et al.* (2013) who altered the expression vector and the expression host, *E. coli*, to be less sensitive to termination signals in the metagenomic insert. This was achieved by inserting the viral elements, T7-RNA polymerase and lambda phage transcription anti-termination protein N, into the vectors and host strain. When screening for carbenicillin resistance genes in a metagenomic library, the frequency of resistant clones increased six-fold, compared to when the host and vector remained unedited. The implementation of the SIGEX system to this altered and inducible host-vector platform resulted in a 10^4-fold enrichment of clones expressing GFP. This system increases the probability of obtaining positive hits, particularly in large-insert libraries, by addressing a major obstacle in functional metagenomic screening.

1.4.3 Functional Metagenomic Screening Techniques for Stress-related Genes

Current literature is biased towards the discovery of genes encoding enzymes, or pathways producing small molecules, with potential use in industry, medicine or bioethanol production (TAUPP *et al.* 2011). Consequently, high-throughput screening methods have been developed to identify clones expressing these enzymes, or producing small molecules, in a metagenomic library. However, such methods have not been firmly established in screening for stress-related genes. In this section, the application of various well-established functional screening methods to stress-related screening is discussed.

1.4.3.1 Biological Activity or Survival Assays

In the case of stress adaptation screening, biological activity refers to the ability of the heterologous host to survive under a particular stress while the wild-type (hosting an empty vector) fails to do so. Such survival assays are simple and powerful, as the result is either ‘growth’ or ‘no growth.’ Drawbacks of survival assays and heterologous complementation (see below) are the high detection threshold of expression and the laborious nature of screening thousands of clones. In terms of the detection threshold, the host has to survive for expression of a gene product to be detected. Many clones may initially express a gene product in response to the stress, but at very low levels which are eventually insufficient to

allow survival of the clone. Nonetheless, the entire metagenomic insert in the surviving clones is sequenced and mined for the gene, or genes, conferring stress-tolerance to the host. The identification of the gene, or genes, responsible would be dependent on known genes or protein domains involved in that particular stress, which somewhat negates the advantage of this form of functional screening; i.e. the identification of entirely novel genes (CULLIGAN *et al.* 2012).

This matter is resolved through the use of transposon mutagenesis. This approach makes use of the random insertion of a transposable element into the fosmid, potentially resulting in gene inactivation (GAILLARD *et al.* 1986; MUÑOZ-LÓPEZ AND GARCÍA-PÉREZ 2010). This method was applied in two studies to identify salt-tolerance genes from a human gut metagenome (CULLIGAN *et al.* 2012; CULLIGAN *et al.* 2013). The authors identified two genes that had not previously been implicated in salt-tolerance and one entirely novel gene with no known homologues that conferred salt-tolerance. The metagenomic library was screened on LB agar plates supplemented with 6.5% sodium chloride (NaCl). Transposon mutagenesis was performed on the surviving clones and these were replica-plated on LB agar plates with and without 6.5% NaCl, which allowed the researchers to determine in which clones the gene conferring salt-tolerance was knocked-out. Sequencing from the ends of the transposon provides the sequence of the inactivated gene, thereby eliminating extensive sequence mining. The studies are classic examples of how functional metagenomic survival assays combined with transposon mutagenesis can uncover genes that have not been implicated in the adaptation to a particular stress or discover completely novel stress-related genes.

1.4.3.2 Heterologous Complementation

Heterologous complementation is a rapid screening method that has frequently been applied to metagenomics. It involves the use of a mutant host, either lacking, or containing an inactive copy of, a particular gene of interest. The mutant host containing metagenomic DNA is then plated on media augmented with a specific substrate which the mutant is unable to utilise or it is subjected to selective growth conditions. The complementation of the function of the knocked-out gene in the mutant host by the metagenomic DNA results in an active or surviving clone.

In a study by SIMON *et al.* (2009), the authors screened small- and large-insert libraries constructed from glacial ice for novel DNA polymerase I genes by complementation screening in a mutant *E. coli* host. The host contained a cold-sensitive (<20°C) lethal

mutation in the *polA* gene encoding DNA polymerase I. The metagenomic libraries were screened at 18°C and any surviving clones therefore expressed DNA polymerase I that was active at lower temperatures. To confirm that these *polA* genes were responsible for the survival of the mutant *E. coli* host, the genes were sub-cloned and expressed in the *E. coli* host containing the cold-sensitive mutation and screened at 18°C.

1.4.3.3 Activity-based Screening

Activity based functional metagenomic screens have mostly been used to discover novel enzymes with industrial potential, such as lipases (COUTO *et al.* 2010), esterases (HEATH *et al.* 2009; HU *et al.* 2012), xylanases (ANGELOV *et al.* 2009) and other hydrolases (SIMON AND DANIEL 2009). These assays are typically based on the use of chromogenic or fluorogenic substrates in plate assays, resulting in a zone of clearance or a colour change around clones expressing functional copies of the desired enzyme. This approach has been widely applied in screening metagenomes from extreme environments to identify alkali-, heat-, cold- or halo-tolerant enzymes with potential commercial applications (SARETHY *et al.* 2011; HU *et al.* 2012; MOHAMED *et al.* 2013).

1.4.3.4 SIGEX: Substrate-induced Gene Expression

Substrate-induced gene expression (SIGEX) is a high-throughput metagenomic screening technology developed by UCHIYAMA *et al.* (2005) for the discovery of catabolic operons. The method is based on a reporter system where a promoterless vector containing a green fluorescent protein (GFP) gene is used in shotgun cloning of the metagenomic DNA. A particular substrate is then added to the metagenomic library culture. If an operon is induced by this substrate in the metagenomic DNA, GFP will be co-expressed as it is under the control of the inducible operon promoter. The cells within the culture are then subjected to fluorescence-activated cell sorting (FACS), where the cells expressing GFP are selectively recovered. Cells from the fluorescent population are isolated and further analysed using standard molecular biology techniques. A similar reporter based assay, termed product-induced gene expression (PIGEX), was developed by UCHIYAMA AND MIYAZAKI (2010). This approach requires two types of cells within one population. The first type, called sensor cells, contains a GFP expression vector under the control of a product-inducible promoter. The second type of cell is the host containing the metagenomic DNA. In a proof-of-principle paper, the authors placed GFP under the control of a benzoate-responsive transcriptional activator in the sensor cells. The activator, BenR, was not responsive to benzamide, an amidase substrate, but was responsive to benzoate, the product of amidase-mediated

benzamide conversion. In cultures containing sensor and library cells, a fluorescence signal indicates the conversion of benzamide to benzoate. Several novel amidases encoded by the metagenomic DNA were identified in a high-throughput method with a low detection threshold (UCHIYAMA AND MIYAZAKI 2010).

Flow cytometry has been successfully applied to study stress responses and adaptation in various organisms (BUNTHOF *et al.* 1999; AMANULLAH *et al.* 2003; DA SILVA *et al.* 2005; PAPADIMITRIOU *et al.* 2007; DA SILVEIRA AND ABEE 2009; TRACY *et al.* 2010). However, FACS has not been applied in screening metagenomic libraries for stress-adaptation genes. The principle of SIGEX could be applied to screen for genes activated in the metagenomic DNA during exposure of the host cell to stresses such as extreme temperature, non-physiological pH levels, UV radiation as well as oxidative, osmotic and organic solvent stress. The advantages of this approach would be its high-throughput mechanism and low detection threshold of genes expressed in response to a particular stress. Survival assays and heterologous complementation allow the discovery of the final effector gene in terms of survival, whereas a SIGEX-based method could reveal intermediate genes involved in the stress response pathway. This approach could provide novel insights into the overall stress adaptation pathways of microbial communities in extreme environments.

A third reporter-based assay is termed metabolite-regulated expression (METREX). It was developed by WILLIAMSON *et al.* (2005) and involves a quorum-sensing biosensor promoter upstream of a GFP-encoding gene. These biosensor promoters are activated when a certain metabolite, acting as a signal molecule, reaches a threshold level. The host contains this biosensor and the vector containing the metagenomic DNA is transformed into the cells, which then allows for the detection of any biologically active small molecules that induce quorum sensing in the host. This method can be applied in stress response screening, as such signal molecules may be expressed in response to abiotic stresses (MAGNUSSON *et al.* 2005).

1.4.3.5 Compound Configuration Screening

Compound configuration screening relies on the identification of chromatographic peaks which differ from those of the host and is a method that can be used to screen metagenomic libraries for novel natural products, such as bioactive small molecules (WANG *et al.* 2000). Through the use of high performance liquid chromatography (HPLC) techniques, chromatographic profiles are created for the host and the metagenomic clones. Unique peaks in the metagenomic samples are then analysed and spectrophotometric techniques,

such as two-dimensional nuclear magnetic resonance spectroscopy, can be used to determine the structure and identity of the novel product. This method was successfully applied by WANG *et al.* (2000) who identified five novel compounds, known as terragines, from a soil metagenomic library. However, the throughput capacity of HPLC could be a potential bottleneck of this approach.

A stress specific application of compound configuration screening could be the identification of novel microbial sunscreens. These are small molecules capable of absorbing UV radiation and are synthesised by microorganisms to protect intracellular structures from damage caused by UV radiation (GAO AND GARCIA-PICHEL 2011). The production of these small molecules from metagenomic DNA in response to UV exposure could be detected via compound configuration screening.

1.5 Conclusion

Extremophiles have evolved numerous mechanisms to survive at the limits of life. Adaptation strategies of microbes have been studied at the molecular, physiological and community levels, but are not yet fully understood (STORZ AND HENGGE 2011; POINTING AND BELNAP 2012). Functional metagenomics holds potential to further provide insight into these survival mechanisms and uncover novel stress-related genes and survival strategies at both the individual cell and community levels. Novel stress-related genes have various potential applications. Desiccation or osmotic stress-related genes could improve resistance of crops to drought or soils with high salinity (SHI *et al.* 2003; ZHONG *et al.* 2012). Microbial sunscreens identified through UV-tolerance screening would be applicable to the skin-care industry, where such molecules are routinely incorporated into sunscreen and skin lotions (GAO AND GARCIA-PICHEL 2011). Antioxidant proteins or molecules identified through oxidative stress screening could be incorporated as part of cancer treatment plans (SOTGIA *et al.* 2011). Proteins conferring tolerance to, or enzymes active at, high temperatures have potential application in industrial microbiology and biotechnology where processes are often carried out at high temperatures and alkaline pH (SARETHY *et al.* 2011). Engineering microorganisms tolerant to, and active at, high temperatures could make these processes more efficient, thereby reducing overall cost (SARETHY *et al.* 2011). Functional metagenomics has only recently begun to be applied in identifying stress-related genes from metagenomes constructed from extremophile communities. Stress-related metagenomic screening has the potential to expand and develop into a valuable and applicable field.

1.6 Research Aims and Objectives

Hyperarid desert soils present a viable target for novel gene discovery through functional metagenomics. The mechanisms by which bacteria survive in soils of these extreme environments have not been fully elucidated (CARY *et al.* 2010; POINTING AND BELNAP 2012). The microorganisms that reside in these soils are exposed to high levels of UV radiation, extreme desiccation, osmotic and oxidative stress, and to low or high temperatures. The question that this project therefore addresses is:

Which genes are involved in the tolerance of microbial communities to multiple abiotic stresses in hyperarid Namib Desert soil?

Through functional screening of a Namib Desert soil metagenome, this project will identify novel genes involved in stress tolerance and survival of microoorganisms in hyperarid deserts. Novel stress-related genes can have possible applications in agriculture, medicine, industrial microbiology and biotechnology (STORZ AND HENGGE 2011).

1.6.1 Aims

To identify novel genes from a Namib Desert soil metagenome which confer tolerance to one or more of the following stress factors:

- oxidative stress;
- hyperosmotic stress;
- UVB radiation;
- high temperature.

1.6.2 Objectives

A. Determine the survival threshold of *E. coli* EPI300:pCC1FOS under each of the stress conditions (Chapter 3).

B. Screen the Namib Desert soil metagenomic library cultures at the survival threshold, as determined in objective A, to identify clones resistant to the particular stress (Chapter 3).

C. Select surviving colonies, extract and sequence the metagenomic DNA of the fosmid clones (Chapter 4).

D. Perform bioinformatic analyses of the metagenomic sequences. Assemble and annotate the metagenomic inserts. Identify functional domains/ proteins that could possibly contribute to tolerance of the specific stress (Chapter 4).

E. Sub-clone genes of interest and over-express the protein in *E. coli* BL21(DE3) and Rosetta(DE3)pLysS strains under stress as proof-of-function (Chapter 5).

Chapter 2: Materials and Methods

2.1 Chemicals and Reagents

All chemicals and reagents used in this project were of analytic grade and obtained from various sources.

2.2 Antibiotics and Enzymes

Unless otherwise stated, all media used with cultures containing the pCC1FOS fosmid contained chloramphenicol (Cm) (Sigma-Aldrich, St. Louis, MO, USA) at a final concentration of 12.5 µg/mL. Cultures containing the pET21a plasmid contained ampicillin (Ap) (Sigma-Aldrich) at a final concentration of 100 µg/mL.

Enzymes used in this study are shown in Table 1.

Table 1. Enzymes used in this study.

Enzyme	Function	Target Site	Supplier
*Nco*I	Restriction digestion of fosmid DNA	5' C↓CATGG 3'	Thermo Fisher Scientific (Waltham, MA, USA)
FastDigest *Hind*III	Restriction digestions of gene fragments and pET21a plasmid DNA for cloning of novel genes	5' A↓AGCTT 3'	Thermo Fisher Scientific
FastDigest *Nde*I		5' CA↓TATG 3'	Thermo Fisher Scientific
FastDigest *Eco*RI		5' G↓AATTC 3'	Thermo Fisher Scientific
FastDigest *Sal*I		5' G↓TCGAC 3'	Thermo Fisher Scientific
T4 DNA Ligase	Ligation of insert and vector DNA	Phosphate backbone of DNA	Thermo Fisher Scientific
ATP-dependent Fast-Link DNA Ligase	Ligation of insert and vector DNA	Phosphate backbone of DNA	Epicentre Biotechnologies (Madison, WI, USA)

2.3 Bacterial Strains and Vectors

Escherichia coli EPI300 (Epicentre Biotechnologies) containing either the pCC1FOS fosmid (hereafter referred to as *E. coli* EPI300:pCC1FOS) or a fosmid with a metagenomic DNA insert were grown in Luria-Bertani (LB) broth (1% [w/v] NaCl; 1% [w/v] tryptone; 0.5% [w/v] yeast extract), or LB agar (1% [w/v] NaCl; 1% [w/v] tryptone; 0.5% [w/v] yeast extract; 1.2% [w/v] agar bacteriological) containing 12.5 µg/mL Cm. The cultures were routinely supplemented with Fosmid Autoinduction Solution (Epicentre Biotechnologies) according to the manufacturer's instructions or with 5 mM filter-sterilised L-arabinose (Sigma Aldrich), to induce a high copy number of the fosmid, and incubated at 37°C, unless otherwise stated, with constant shaking at 180 rpm for 16 to 20 hours (hrs). After overnight incubation each culture was sub-cultured into fresh LB broth containing 12.5 µg/mL Cm and grown at 37°C with constant shaking at 180 rpm. Exponential phase cultures (optical density at 600 nm (OD_{600}) of 0.1 to 0.15) were used in subsequent experiments. The OD_{600} was measured using the Multiskan Go spectrophotometer (Thermo Fisher Scientific).

E. coli JM109 (Promega, Madison, WI, USA), BL21(DE3) and Rosetta(DE3)pLysS (Novagen, USA) strains containing the pET21a vector, with or without an insert, were cultured in LB broth, salt-reduced LB broth (0.5% NaCl) or on LB agar. Growth media was routinely supplemented with Ap (100 µg/mL) and cultures were incubated at 37°C, unless otherwise stated, with constant shaking at 180 rpm when applicable.

The vectors used in this study are shown in Table 2.

Table 2. Vectors used in this study.

Plasmids/ Vectors	Features	Final Antibiotic Concentration (µg/mL)	Source
pCC1FOS	Cm^R	12.5	Epicentre Biotechnologies
pET21a	Ap^R	100	Thermo Fisher Scientific

2.4 General Laboratory Techniques

2.4.1 Agarose Gel Electrophoresis

Agarose gels were prepared as 1% [w/v] agarose (SeaKem® LE Agarose, Lonza, Basel, Switzerland), dissolved in 1 x TAE buffer (40 mM tris base, 20 mM acetic acid, and 1 mM ethylenediaminetetraacetic acid (EDTA) [pH 8.0]). DNA fragments were resolved by mixing samples in a 1:1 ratio with GelRed loading dye (Anatech, Johannesburg, South Africa) or 6 x loading dye (Thermo Fisher Scientific). In the latter case, SYBR Safe DNA Gel Stain (Invitrogen, Carlsbad, CA, USA) or GRGreen Nucleic Acid Stain (Excellgen, Rockville, MD, USA) was used to pre-stain the gel in a ratio of 1:50 000. Electrophoresis was performed at 80 – 100 volts (V) for 45 minutes (min), unless stated otherwise. Lanes loaded with GeneRuler 1 kb DNA Ladder (Fermentas, Vilnius, Lithuania) or KAPA Universal DNA Ladder (Kapa Biosystems, Cape Town, South Africa) were routinely included to estimate the size of DNA fragments. DNA was visualised under UV illumination using the Bio-Rad Gel Doc XR+ Molecular Imager (Bio-Rad, Hercules, CA, USA) and Image Lab Software Version 4.1 (Bio-Rad).

2.4.2 DNA Quantification

Unless otherwise stated, DNA was quantified using the NanoDrop 2000 (Thermo Fisher Scientific) or from agarose gel images using Image Lab Software Version 4.1 (Bio-Rad) and GeneRuler 1 kb DNA Ladder (Fermentas) as reference to construct a calibration curve.

2.4.3 DNA Purification

DNA was routinely purified from PCR reactions, enzyme digestion reactions and agarose gels using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Hoerdt, France) or the GeneJET Gel Extraction and DNA Cleanup Micro Kit (Thermo Fisher Scientific).

2.5 Preparation of Chemically Competent *E. coli* Cells

All procedures were carried out under sterile conditions. *E. coli* (EPI300, JM109, BL21(DE3) or Rosetta(DE3)pLysS) was streaked out on LB agar (no antibiotics) from a glycerol stock and incubated overnight at 37°C. A single colony was picked and inoculated into 5 mL LB broth (no antibiotics) and incubated overnight at 37°C with constant shaking at 180 rpm. From this culture, 1 mL was inoculated into 200 mL fresh LB broth. This culture was

incubated at 37°C with constant shaking at 180 rpm. The OD_{600} was measured every 15 to 20 min until a value between 0.35 and 0.4 was reached. All solutions used from here onwards were ice cold and all glassware and tubes were pre-chilled and kept on ice throughout the procedure. All centrifugation steps were carried out at 4°C. The culture was chilled on ice for 30 min after which the 200 mL volume was divided into four 50 mL falcon tubes. Cells were harvested by centrifugation at 3000 x *g* for 15 min in an Eppendorf Centrifuge 5810 R (Eppendorf, Hamburg, Germany). The supernatant was decanted and cells resuspended in 20 mL 100 mM $MgCl_2$. Cells were harvested by centrifugation at 2000 x *g* for 15 min, the supernatant was decanted and the cells were resuspended in 40 mL 100 mM $CaCl_2$. The suspension was incubated on ice for 20 min, after which cells were harvested by centrifugation at 2000 x *g* for 15 min. The supernatant was decanted and cells were resuspended in 10 mL 85 mM $CaCl_2$, 15% [v/v] glycerol. Cells were harvested by centrifugation at 1000 x *g* for 15 min. Finally, the supernatant was decanted and cells were resuspended in 1 mL 85 mM $CaCl_2$, 15% glycerol. Cells were stored at -80°C in 50 µL aliquots in sterile 1.5 mL eppendorf tubes (SAMBROOK *et al.* 1989; GREEN AND SAMBROOK 2012).

2.6 Transformation of Chemically Competent *E. coli* Cells

Plasmid DNA was added directly to 25 µL or 50 µL *E. coli* cells, which had been thawed on ice, and mixed by gentle agitation. If the plasmid DNA had been extracted from a cell culture then 2 µL was added. However, if the plasmid DNA was from a ligation reaction then 5 µL was added to the cells. The cells were incubated on ice for 20 min and transformed via heat-shock by incubating the cultures at 42°C for 45 seconds (s). The cells were immediately returned to ice for two min, after which 950 µL LB broth (no antibiotics) was added to the cells. The culture was incubated at 37°C, with shaking, for two hrs. The culture was spread-plated onto five LB agar plates (200 µL per plate), supplemented with 12.5 µg/mL Cm or 100 µg/mL Ap. The LB agar plates were pre-warmed to 37°C. The cultures were incubated overnight at 37°C. Transformation was confirmed via plasmid extractions or colony polymerase chain reactions (PCRs). Transformed cells were stored at -80°C as 30% [v/v] glycerol stocks until required (SAMBROOK *et al.* 1989; GREEN AND SAMBROOK 2012).

2.7 The Namib Desert Soil Metagenomic Library

The Namib Desert soil metagenomic library screened in this project was constructed from desert surface soil samples obtained from the following coordinates: S 23º 14.575’ E 015º 21.219’, using aseptic techniques. The construction of the library did not form part of this

project. In short, the metagenomic library was constructed using the CopyControl™ Fosmid Library Production Kit with pCC1FOS™ Vector (Epicentre Biotechnologies). Four different vector:insert ratios were used to construct the library, which resulted in four sub-libraries being produced. After transformation of *E. coli* EPI300 with the fosmid vectors containing metagenomic DNA, the colonies on each LB agar plate were pooled and stored as triplicate 30% [v/v] glycerol stocks at -80°C until required. This resulted in between 60 and 70 pooled samples in each sub-library, labelled according to the ratio of the library and then alphabetical letters, for example 3:1 A, 3:1 B etc. When the end of the alphabet was reached, letters were doubled (e.g. 3:1 AA, 3:1 BB) and when the end of the alphabet was reached again, samples were labelled as follows, 3:1 AB, 3:1 AC etc. This was designated as the Namib Desert soil metagenomic library (NA). In order to make screening of the metagenomic library less intensive, every 18 samples were further pooled into a single sample, which would then be labelled according to the first and last sample pooled into that sample, e.g. 3:1 A-R, 3:1 S-JJ, 3:1 KK-AC and 3:1 AD-AN. This resulted in four samples for each sub-library. This final pooled library was designated as the Namib Pooled library (NAP). The NAP was then used in all subsequent screening experiments. The Namib Desert soil metagenomic libraries constructed with different vector:insert ratios and the number of clones contained within each sub-library are shown in Table 3.

Table 3. Metagenomic libraries constructed from Namib Desert soil.

Library (vector:insert)	Number of Clones
3:1	133 500
2:1	37 750
1:1	557 500
1:3	417 500
Total Number of Clones	**1 146 250**

2.8 Survival Assays of the Library Host Strain

E. coli EPI300:pCC1FOS cultures were subjected to various levels of each stress condition investigated to determine the survival threshold of the host strain. As a positive control, exponential phase *E. coli* EPI300:pCC1FOS cultures were serially diluted up to 10^{-6}, using quarter strength Ringer's solution (Sigma-Aldrich), and 100 µL was plated, in triplicate, on LB agar plates without the stress condition. These cultures were used to determine the survival rate of the host under each of stress condition. The same procedure was followed

when using the other *E. coli* strains mentioned below. All survival assays were performed at least twice, in triplicate.

2.8.1 Hyperosmotic Stress-tolerance Assays

E. coli EPI300:pCC1FOS was inoculated into LB broth and induced to produce high fosmid copy numbers for 16 to 20 hrs at 37°C with constant shaking at 180 rpm. The culture was then inoculated into fresh LB broth and incubated at 37°C with constant shaking at 180 rpm until the exponential growth phase was reached. The culture was then diluted to 10^{-4} in quarter strength Ringer's solution and spread-plated (100 µL), in triplicate, onto LB agar plates containing 2% [w/v], 3%, 3.5%, 4% or 5% NaCl, considering the salt content of the LB medium, to investigate tolerance to hyperosmotic stress-tolerance. The cultures were incubated at 37°C overnight.

E. coli BL21:pET21a and Rosetta:pET21a were inoculated into 0.5% NaCl LB broth supplemented with 0.5% glucose and incubated overnight. These overnight cultures were then re-inoculated 1/100 into fresh 0.5% NaCl LB broth and grown to an OD_{600} of between 0.4 and 0.8, after which the cultures were induced with 0.4 or 0.8 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated overnight at 37°C, with constant shaking at 180 rpm. Cultures were then normalised to an OD_{600} of 0.6 in 0.5% NaCl LB broth containing Ap. Cultures were then spread-plated (100 µL) in triplicate on LB agar plates containing 2%, 3%, 4%, 5% or 6% NaCl. The cultures were incubated at 37°C for 72 hrs and the number of colonies on each LB agar plate was recorded every 24 hrs.

The colony forming units (cfu) per millilitre under each stress level was determined using the average number of colonies from triplicate plates using the following equation:

$$\text{cfu/mL} = \frac{\text{number of colonies} \times \text{dilution factor}}{\text{volume plated (mL)}}$$

The survival rate under each stress level was calculated using the formula below:

$$\text{Survival rate (SR)} = \frac{\text{cfu/mL on stress plate}}{\text{cfu/mL on LB agar plate}} \times 100$$

2.8.2 Oxidative Stress-tolerance Assays

The survival threshold of the library host under H_2O_2 (Merck, Darmstadt, Germany) treatment, to investigate oxidative stress-tolerance, was determined using two different approaches. In the first approach, exponential phase *E. coli* EPI300:pCC1FOS cultures were diluted 1:1 with LB broth. H_2O_2 was added to a final concentration of 1 mM, 2 mM, 3 mM, 4 mM, 5 mM or 6 mM. The cultures were incubated for an additional 30 min at 37°C with constant shaking, after which 100 µL was plated in triplicate on LB agar plates, which were then incubated at 37°C overnight. The survival rate under each stress level was calculated.

In the second approach, H_2O_2 was added to LB agar to a final concentration of 1 mM, 1.5 mM, 2 mM or 2.5 mM. Exponential phase *E. coli* EPI300:pCC1FOS cultures were spread-plated (100 µL) on the stress plates in triplicate and incubated at 37°C for 48 hrs. The results were recorded every 24 hrs, for a maximum of 72 hrs, and the survival rate calculated.

2.8.3 Heat-tolerance assays

Exponential phase *E. coli* EPI300:pCC1FOS cultures were spread-plated (100 µL) in triplicate, on LB agar plates and incubated at 45.5°C and 46°C for 48 hrs, after which the survival rate was calculated.

2.8.4 UVB Radiation-tolerance Assays

Exponential phase *E. coli* EPI300:pCC1FOS cultures were spread-plated (100 µL) in triplicate on LB agar plates and incubated at room temperature for approximately 10 min, to allow the liquid media to be absorbed into the agar plates. These plates were then subjected to UVB irradiation (280 – 315 nm) by removing the lid and placing the plates upside-down on a Spectroline Ultraviolet Transilluminator (Spectronics Corporation, Westbury, USA), at half-strength, for 10, 15, 20, 25, 30 or 45 s. This equated to exposure intensities of 1179; 1768; 2357; 2947; 3537 and 5305 mJ/cm^2, at each time point, respectively. The intensity (or dose) of the UV irradiation was calculated using the equation below. Cultures were then incubated a 37°C overnight and the survival rate was calculated.

$$\text{UVB intensity} = \frac{\text{Energy (Joules)}}{\text{surface area of petri dish (cm}^2\text{)}}$$

Where, energy (joules) was calculated using the equation: $\mathrm{E(Joules)} = \mathrm{Power\ (Watts)} \times \mathrm{time(s)}$. The Spectroline Ultraviolet Transilluminator contained a lamp of 15 watt (http://www.spectroline.com/content/transilluminators-lab). This value was halved in the calculations above, as the transilluminator was used at half of its maximum power. Petri dishes of diameter 90 mm were used in this study.

2.9 Screening the Namib Soil Metagenomic Library for Stress-tolerant Clones

Induced library and control (*E. coli* EPI300:pCC1FOS) cultures in the exponential growth phase were exposed to the level of stress at which *E. coli* EPI300:pCC1FOS had a survival rate of 0%. In general, the library was screened by spread-plating 100 µL of each culture, in triplicate, on LB agar plates containing the appropriate stress and incubating the cultures at 37°C until surviving colonies appeared or for a maximum of 72 hrs. Surviving colonies were picked manually and cultured in LB broth containing 12.5 µg/mL Cm and incubated overnight at 37°C with constant shaking at 180 rpm to confirm that the clones were resistant to Cm and thus contained a fosmid. The single clones were then grown to exponential phase and subjected to the particular stress again, to confirm resistance. Screening of the metagenomic library under each stress was performed at least twice, in triplicate.

2.9.1 Hyperosmotic Stress-tolerance Screening

The Namib Desert soil metagenomic library was screened for salt-tolerance clones by adding NaCl to LB agar at a final concentration of 5% [w/v].

2.9.2 Oxidative Stress-tolerance Screening

The Namib Desert soil metagenomic library was screened by adding H_2O_2 to exponential phase cultures at a final concentration of 6 mM. The cultures were then incubated at 37°C, with shaking, for an additional 30 min, after which 100 µL was spread-plated onto LB agar, in triplicate, and incubated overnight at 37°C.

The library was also screened for oxidative stress-tolerant clones by adding H_2O_2 to LB agar at a final concentration of 0.7 mM, 0.725 mM, 0.75 mM or 1 mM.

The library was screened using a third approach. Enrichment of oxidative stress-tolerant clones within the library was attempted before screening on LB agar stress plates. Overnight induced cultures were inoculated into fresh LB broth to an OD_{600} between 0.05 and 0.08.

H_2O_2 was then added to the cultures at a final concentration of 0.725 mM. The cultures were incubated at 37°C, with constant shaking at 180 rpm, until an OD_{600} of between 0.1 and 0.15 was reached. The optical density of the cultures was measured after one hour and every 30 min thereafter. Exponential phase cultures were spread-plated (100 µL) onto LB agar plates containing 0.725 mM H_2O_2. The cultures were incubated at 37°C overnight.

2.9.3 Heat-tolerance Screening

Screening for heat-tolerant clones in the Namib Desert soil metagenomic library was performed by spread-plating 100 µL of exponential phase cultures in triplicate on LB agar plates and incubating the cultures at 46°C for a maximum of 72 hrs.

2.9.4 UVB Radiation-tolerance Screening

Exponential phase library cultures were plated out on LB agar plates, allowed to stand for 10 min and exposed to UVB radiation (as in section 2.8.4) for 45 s, after which the cultures were incubated at 37°C overnight.

2.10 Plasmid DNA Isolation

Fosmid library clones were inoculated in 5 mL LB broth containing 12.5 µg/mL Cm and 5 mM arabinose or Autoinduction Solution (Epicentre Biotechnologies) and incubated for 16 hrs at 37°C and constant shaking at 180 rpm. *E. coli* JM109:pET21a cells were inoculated into LB broth containing 100 µg/mL Ap and incubated overnight at 37°C and constant shaking at 180 rpm. The cultures were then transferred to sterile 1.5 mL eppendorf tubes and the cells were pelleted by centrifugation at 12 000 rpm for one minute in an Eppendorf 5424 R bench-top centrifuge (Eppendorf). Fosmid DNA was then extracted from the cell pellet in one of three ways, described in full below, and stored at -20°C until required.

2.10.1 Alkaline Lysis Miniprep

Cell pellets were resuspended in 100 µL of Solution I (50 mM glucose; 25 mM Tris-HCl [pH 8.0]; 10 mM EDTA [pH 7.5]) and incubated at room temperature for five min. Thereafter 200 µL Solution II (0.2 M NaOH; 1% [w/v] SDS) was added and the suspension mixed by inverting the tubes repeatedly. Immediately following this, 150 µL Solution III (5M KOAc; glacial acetic acid [pH 4.8]) was added, the suspension mixed by inverting the tubes repeatedly and incubating on ice for five min. The samples were centrifuged at 15 000 rpm

for 10 min in an Eppendorf 5424 R centrifuge. The supernatant was recovered and 1 volume phenol:chloroform (1:1) was added. The samples were mixed by inversion and centrifuged as before at 4°C. The aqueous phase was transferred to a new eppendorf tube and 1 volume chloroform:isoamyl alcohol (24:1) was added and the samples mixed and centrifuged as before. The aqueous phase was recovered and 2 volumes ice-cold 99% ethanol was added and the DNA precipitated by incubating the samples at -20°C for at least 10 min. The samples were then centrifuged at 15 000 rpm for 15 min at 4°C. The supernatant was discarded and the pellet was washed in 200 µL cold 70% ethanol by centrifugation at 15 000 rpm for 5 min at 4°C. The pellet was dried in a heating block at 60°C for 10 min or until dry and resuspended in 30 µL nuclease-free water. RNA was removed by the addition of 1 µL RNase A (10 mg/mL, Thermo Fisher Scientific) to the final suspension and incubation at 37°C for 30 min or on bench-top overnight (SAMBROOK *et al.* 1989; GREEN AND SAMBROOK 2012).

2.10.2 QIAprep Spin Miniprep Kit (Qiagen)

Plasmid DNA was extracted as per the manufacturer's instructions. Specific alterations included the immediate addition of Buffer N3 after the addition of Buffer P2 and mixing by inversion, as increased incubation time in Buffer P2 may result in genomic DNA contamination. During the final step of the protocol, DNA was eluted using 30 µL nuclease-free water.

2.10.3 Non-ionic Detergent (NID) Plasmid Isolation

This plasmid extraction method is based on a protocol developed by LEZIN *et al.* (2011). Cell pellets were resuspended in 150 µL extraction buffer (5% [w/v] sucrose; 20 mM EDTA [pH 8.0]; 50 mM Tris-HCl [pH 8.0]; 0.75 M NH_4Cl; 0.5% [v/v] Triton X-100; 0.075 µg/mL lysozyme) and incubated at 65°C for 5 min. The suspension was centrifuged at 15 000 rpm for 10 min. The supernatant was transferred to a new eppendorf tube, to which 120 µL isopropanol was added and the DNA was precipitated at -20°C for at least 10 min. The samples were then centrifuged at 7000 rpm for 10 min at room temperature. The DNA pellet was then washed with 200 µL cold 70% ethanol by centrifugation at 15 000 rpm for 5 min. The ethanol was discarded and the pellet was dried, resuspended and treated with RNase A as in the alkaline lysis method.

2.11 Fosmid DNA Restriction Digestion Analysis

Fosmid DNA restriction digestion analysis was performed using the *Nco*I restriction enzyme (Thermo Fisher Scientific). The restriction digests were set up as in Table 4.

Table 4. *Nco*I restriction digests of fosmid DNA.

Reagent	Volume Added (µL)
Fosmid DNA (~32 ng/µL)	17
10X Buffer Tango	2
*Nco*I	1
Total	**20**

Digestion reactions were incubated at 37°C for 16 hrs. The enzyme was inactivated after digestion by incubation at 65°C for 20 min. The DNA was then separated on a 1% agarose gel via electrophoresis at 80 V for two-and-half to three hrs.

2.12 Salt-tolerance Growth Experiments

Salt-tolerant clones (STCs) were induced for 16 to 20 hrs and re-inoculated in fresh LB broth supplemented with Cm, and in LB broth supplemented with Cm and 5% [w/v] NaCl. All cultures had an OD_{600} of 0.001 at 0 hrs. The cultures were then incubated at 37°C, 170 – 180 rpm and OD_{600} readings were recorded at 1, 3, 6, 9, 12, 24 and 48 hrs post-inoculation. Readings were recorded in triplicate and the experiment was performed three times. The readings at each time-point (technical and biological repeats) for each STC was averaged and used to construct growth curves. The Shapiro-Wilk test was used to determine whether the OD_{600} readings at 48 hrs followed a normal distribution. Where data were not normally distributed a log-transformation was performed to fit a normal distribution. To determine whether growth of STCs was significantly different from the wild-type, Student's *t*-test (unpaired) for independent samples was performed. This tests the null hypothesis which states that the means of two independent samples (in this case, the OD_{600} of the wild-type and STC cultures at 48 hrs) are the same. A *P*-value of <0.05 allows the null hypothesis to be rejected. Thus, the difference observed between the means of the samples would not be due to chance alone. R software (version 3.1.0) was used to perform all statistical analyses (http://www.r-project.org/).

2.13 Next Generation Sequencing and Sequence Analysis

Fosmid DNA was sequenced on the Ion Torrent PGM (Personal Genome Machine), using the 316 chip and a unique barcode for each fosmid, at the Sequencing Facility of the University of Pretoria, South Africa.

A modified plasmid extraction method to that described in section 2.10.1 (QIAprep Spin Miniprep Kit (Qiagen)) was used to isolate fosmid DNA for Ion Torrent sequencing. From the overnight culture, 3 mL was used to harvest cells; i.e. 1.5 mL was added to the same eppendorf tube, twice. The protocol was then followed as per the manufacturer's instructions. The flow-through in the collection tube was discarded using a pipette, to ensure the sides of the spin column stayed dry and free of any chemicals that could contaminate the final elution. The final DNA elution step was performed twice, using 20 µL nuclease-free water and 10 µL nuclease-free water, respectively. These steps were taken to ensure high-quantity (>20 ng/µL) and high-quality DNA was obtained to achieve quality Ion Torrent sequencing results. Fosmid DNA concentration and quality for Ion Torrent sequencing was determined using the Qubit 2.0 Fluorometer (Invitrogen).

Quality data on the reads for each fosmid were obtained using PRINSEQ (http://edwards.sdsu.edu/cgi-bin/prinseq/prinseq.cgi?home=1) (SCHMIEDER AND EDWARDS 2011) and quality control was performed using various python scripts on Linux. Firstly, low quality bases were trimmed from the 5' and 3' ends of all reads. Secondly, all reads shorter than 80 nucleotides or longer than 360 – 430 nucleotides (dependent on individual fosmids) were discarded. Lastly, nucleotides were trimmed from the 3' end until every read satisfied the quality parameter Q24. Reads were then assembled into contiguous sequences (contigs) using MIRA (CHEVREUX *et al.* 1999) through the Centre for High Performance Computing (http://www.chpc.ac.za) and contigs were joined into scaffolds or complete fosmids using Gap5 (BONFIELD AND WHITWHAM 2010). Any contigs that could not be joined into complete fosmids were manually joined by 12 N's.

Vector sequences were manually trimmed from the ends of the assembled fosmids. Open reading frames (ORFs) were predicted by submitting the assembled inserts on the RAST (Rapid Annotation using Subsystem Technology) server (http://rast.nmpdr.org/) (OVERBEEK *et al.* 2014) and using CLC Main Workbench version 7.0.2 (CLC bio, Aarhus, Denmark) (allowing ATG, GTG and TTG as alternate start codons), MetaGeneMark (http://exon.gatech.edu/GeneMark/meta_gmhmmp.cgi) (ZHU *et al.* 2010), ORFinder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), Glimmer (https://ccb.jhu.edu/software/glimmer/)

(Salzberg *et al.* 1998) and Softberry fgenesB, which was also used to predict operons (http://www.softberry.com/berry.phtml) (Solovyev and Salamov 2011). The RAST annotation was downloaded as a GFF file and a graphic representation of the genes present on each fosmid was created using this file in CLC Main Workbench 6.8.3. Translated ORFs were also compared to the protein and conserved domain databases (CDD) using the BLASTp function on the National Centre for Biotechnology Information (NCBI) website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Promoter regions were predicted using Softberry BPROM (http://www.softberry.com/berry.phtml). Gaps left by the annotation programs on the fosmid was submitted to BLASTx analysis and manually annotated if applicable. RAST annotated genes were manually checked at random, using BLASTx as before, to ensure there were no incorrectly annotated genes. The metagenomic insert of STC 89 was submitted on the antiSMASH server; an online program that identifies secondary metabolite biosynthesis gene clusters (MEDEMA *et al.* 2011) (http://antismash.secondarymetabolites.org/). The amino acid sequence of a putative bacteriocin gene identified on STC 89 was submitted on the BAGEL3 (http://bagel.molgenrug.nl/) and BACTIBASE (http://bactibase.pfba-lab-tun.org/) servers, which compared the protein to databases of known bacteriocins via BLASTp analysis.

Phylogenetic analyses were performed on the metagenomic DNA, treating each fosmid as a separate sample. A housekeeping gene, or otherwise the best-candidate gene, was selected from each fosmid for phylogenetic analysis. The BLASTp function (NCBI) was used to compare the amino acid sequence of the selected gene to known proteins. The amino acid sequence of the protein, instead of the nucleic acid sequence of the gene encoding the protein, was used. This is because the nucleic acid sequence could give a much lower identity score when comparing the same gene between species due to codon usage bias. The amino acid sequence negates this effect, due to the redundancy of the genetic code, and thus gives a higher and more accurate identity score when comparing the same protein between species. The ten genera with the closest match to the query protein sequence in the BLASTp results were selected to infer the phylogenetic origin of the segment of DNA on which the selected gene was located. The type species and type strain of each genus (as listed on http://www.bacterio.net/) was used in further analysis if present in the list of BLASTp results. If the type species or type strain was not present, then the species or strain with the closest match to the query was used. Before the amino acid sequence of the relevant protein from each species was obtained, the identity of the species on the NCBI database was checked by submitting the 16S rRNA gene sequence on http://rdp.cme.msu.edu/seqmatch, searching only type strains. The amino acid sequence of the selected protein was downloaded from the NCBI for each species. The amino acid

sequences were aligned using ClustalW (http://www.phylogeny.fr) (Larkin *et al.* 2007). The alignment was curated using Gblocks (http://www.phylogeny.fr). The curated alignments were used to construct phylogenetic trees through the neighbour-joining method, with 1000 bootstrap replicates, in MEGA (Molecular Evolutionary Genetic Analysis) Software version 5.2 (TAMURA *et al.* 2011). The trees were rooted with the branch of the most distantly related species.

2.14 Cloning of Putative Stress-tolerance Genes

2.14.1 PCR Amplification of Novel Genes

Primers used to amplify novel genes in this study (Table 5) were designed manually for directional cloning into pET21a (Novagen). CLC Main Workbench version 7.0.2 was used to analyse assembled metagenomic fosmid insert DNA and thus assisted in primer design. Primer quality and compatibility of primer pairs was determined using the online primer analysis tool NetPrimer (http://www.premierbiosoft.com), where a score of above 75 for each primer generally indicates good quality oligonucleotides.

Table 5. Primers used for PCR amplification. Bold nucleotides denote restriction enzyme sites.

Primer ID	Sequence (5' – 3')	Target
STC131F	GCAT**CATATG**CTCAACGCCGTCATCG	Putative cation efflux system protein (CESP)
STC131-IF	TCCTACATCATCTCCATCCT	3' half of CESP
STC131R	TTC**AAGCTT**TCATGGCAGGCTCCTGG	CESP
STC131-IR	AGAAGATGTTGAGGGTGG	5' half of CESP
STC1-1F	GCG**CATATG**AGGATGAAAGGCTTC	ABC transporter, substrate-binding protein, family 5
STC1-1R	ACGCG**AAGCTT**ACTTTACTTCCTGGTA	ABC transporter, substrate-binding protein, family 5
STC1-2F	ACGG**CATATG**TCCTCTGAGGCGCTG	RelA-SpoT-like hypothetical protein
STC1-2R	CTGC**AAGCTT**CTATCGAACGGCATCC	RelA-SpoT-like hypothetical protein
STC1-2.1R	CTGC**GAATTC**CTATCGAACGGCATCC	RelA-SpoT-like hypothetical protein

Primer ID	Sequence (5' – 3')	Target
STC1-3F	CGAG**GAATTC**ACTCCTGGTCCATGTTTC	HD-hydrolase, including promoter
STC1-3R	ATTA**GTCGAC**CGTCACGCAGCCTTCTTC	HD Hydrolase
STC90-1F	AGT**CATATG**AACGACCCGCGGATCG	Peptidase M29
STC90-1R	TAT**AAGCTT**CGCTGCGTCGCTAGCCCA	Peptidase M29
STC92-1F	CGT**GAATTC**GTGGAGCACCTGGAGATT	Na^+/H^+ antiporter
STC92-1.1F	CCGT**CATATG**GAGCACCTGGAGATT	Na^+/H^+ antiporter
STC92-1R	AGCC**AAGCTT**CTAGATCTCCAACCGCTG	Na^+/H^+ antiporter
T7F	TAATACGACTCACTATAGGG	Genes cloned into pET21a
T7R	GCTAGTTATTGCTCAGCGG	Genes cloned into pET21a

Genes with putative roles in stress-tolerance were amplified via conventional PCR, using iProof High-Fidelity DNA Polymerase (Bio-Rad). Between 50 and 100 ng of fosmid DNA was used as template in 25 µL reaction volumes containing 1X HF buffer, 200 µM dNTPs, 0.48 µM of each primer, 0.6, 0.7 or 1 mM $MgCl_2$ and 0.5 U iProof High-Fidelity DNA Polymerase. Negative controls (no DNA template) were routinely included. Thermal cycling conditions were as follows: initial denaturation at 98°C for 3 min; 30 cycles of denaturation at 98°C for 30 s, annealing at 70°C for 30 s and extension at 72°C for 1 min 30 s or 1 min 45 s; and a final extension step at 72°C for 7 or 10 min. The primer pairs with different annealing temperatures (Tm) were as follows: T7F and T7R (Tm = 47°C), T7F and STC131-IF (Tm = 51.5°C) and STC131-IR and T7R (Tm = 57.7°C). Reactions were performed using Bio-Rad C1000 or T100 Thermal Cyclers (Bio-Rad).

2.14.2 Directional Cloning of Novel Genes in pET21a

PCR products and pET21a plasmid DNA were digested with the applicable FastDigest restriction enzymes according to the manufacturer's instructions. The digested PCR products were then ligated into pET21a using T4 DNA Ligase (Thermo Fisher Scientific) or the Fast-Link DNA Ligation Kit (Epicentre Biotechnologies), according to the manufacturer's instructions, with the following alterations: An insert:vector molar ratio of 5:1 was used in all ligation reactions. Reactions were set up in a total volume of 10 – 15 µL, incubated at 22°C for 10 min and thereafter at 16°C overnight (T4 DNA Ligase) or at room temperature overnight (Fast-Link DNA Ligase). Enzymes were inactivated at 70°C for 10 or 15 min. The ratio of vector to insert DNA to be added to the reaction was calculated using the following equation:

$$ngI = \frac{ngV \times TI}{Tv} \times R$$

Where, ngI = nanograms of insert DNA, ngV = nanograms of vector DNA, T_I = size of insert (bp), T_V = size of vector (bp) and R = insert:vector molar ratio.

E. coli JM109 cells were transformed with the ligation reaction as described in section 2.6. Transformants were screened for recombinant constructs by colony PCR or plasmid extractions using the Qiagen miniprep kit and subsequent restriction enzyme digestion. The extracted plasmid DNA was digested with the appropriate enzymes and the products separated on a 1% agarose gel, to confirm that the insert is the expected size. Colony PCRs were performed using DreamTaq DNA Polymerase (Thermo Fisher Scientific). One half of a colony was added to a 25 µL reaction containing 1X DreamTaq Buffer, 160 µM dNTPs, 0.48 µM of each primer and 0.025 U DreamTaq DNA Polymerase. Thermal cycling conditions were as follows: initial denaturation at 95°C for 6 min; 30 cycles of denaturation at 95°C for 30 s, annealing at 70°C for 30 s and extension at 72°C for 1 min 30 s or 1 min 45 s; and a final extension step at 72°C for 7 or 10 min. Recombinant plasmids were extracted from positive colonies using the Qiagen miniprep kit and the DNA was quantified. The cloned gene was amplified via conventional PCR using iProof polymerase protocol described before and the T7 primers, which are designed to bind to regions flanking the multiple cloning site of the pET21 vector (section 2.14.1). The cleaned PCR product was sequenced via dye terminator sequencing at Stellenbosch University Central Analytic Facilities. The sequencing results were subsequently analysed with BioEdit Sequence Alignment Editor Version 7.2.5 (HALL 1999). Sequence-verified recombinant plasmids were subsequently used to transform *E. coli* BL21(DE3) and Rosetta(DE3)pLysS cells.

2.15 Bioinformatic Analysis of Putative Stress-tolerance Genes and the Encoded Proteins

The expected protein size of novel proteins was determined by submitting the amino acid sequence on the ProtParam tool on the ExPASy server (GASTEIGER *et al.* 2005). This tool also provides other useful characteristics of the protein, such as the molecular weight, isoelectric point, estimated half-life *in vivo*, Grand Average Hydropathy (GRAVY) and amino acid content. SignalP (http://www.cbs.dtu.dk/services/SignalP/) (PETERSEN *et al.* 2011) and LipoP (http://www.cbs.dtu.dk/services/LipoP/) (JUNCKER *et al.* 2003) were used to determine whether any signal peptides were present and TMHMM was used to predict transmembrane helices in membrane proteins (http://www.cbs.dtu.dk/services/TMHMM/). Rare codon

composition, for expression in *Escherichia coli* was determined by submitting the nucleic acid sequence of the stress-related gene on Rare Codon Caltor (http://people.mbi.ucla.edu/sumchan/caltor.html). Putative transporter proteins were submitted for BLASTp analysis on the Transporter Classification Database (http://www.tcdb.org/) (SAIER *et al.* 2014). Functional domains were predicted by submitting the amino acid sequences on the Simple Modular Architecture Research Tool (SMART) server (http://smart.embl-heidelberg.de/) (SCHULTZ *et al.* 1998). BioEdit Sequence Alignment Editor Version 7.2.5 was used to generate Kyte and Doolittle hydrophobicity plots. Protein alignments were performed using the Clustal Omega server online (http://www.ebi.ac.uk/Tools/msa/clustalo/) (SIEVERS *et al.* 2011). Sequence logos of conserved domains or motifs were constructed by submitting a multiple sequence alignment of the amino acid sequences of the region of interest on WebLogo (version 3.4) (http://weblogo.threeplusone.com/) (CROOKS *et al.* 2004).

2.16 Heterologous Protein Expression

E. coli BL21(DE3) and Rosetta(DE3)pLysS cells harbouring constructs, BL21:pET21 and Rosetta:pET21 (controls) were inoculated into 0.5% NaCl LB broth and incubated overnight. The overnight cultures were then re-inoculated 1/100 into fresh 0.5% NaCl LB broth and grown to an OD_{600} of between 0.4 and 0.8, after which the cultures were induced with 0.4 or 0.8 mM IPTG and incubated overnight at 37°C and constant shaking at 180 rpm.

Heterologous protein expression was confirmed via denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Total cell protein extracts were prepared as follows: cells (1 mL) were harvested from induced cultures by centrifugation at 12 000 rpm for 1 – 2 min in an Eppendorf 5424 R bench-top centrifuge (Eppendorf). Cells were resuspended in 200 µL deionised distilled water (ddH_2O) and stored at -20°C. Protein samples were prepared by adding 20 µL of cells to 5 µL 5X loading buffer (300 mM Tris-HCl [pH 6.8]; 5% [w/v] SDS; 50% [v/v] glycerol; 10% [v/v] β-mercaptoethanol; 25 mM EDTA; 0.2% [w/v] bromophenol blue), denaturing at 100°C for 10 min and centrifugation at 15 000 rpm for 10 min. Samples (3 µL) were loaded onto a gel containing a 4% stacking gel and a 10% separating gel. Electrophoresis was carried out in the Mini PROTEAN Tetra Cell system (Bio-Rad) containing 1X running buffer (25 mM Tris; 192 mM glycine; 0.1% SDS) at 100 – 130 V until the dye front reached the bottom of the gel. Gels were then stained with Coomassie Brilliant Blue staining solution (50% [v/v] ddH_2O; 40% [v/v] methanol; 10% [v/v] acetic acid; 2.5% [w/v] Coomassie Brilliant Blue) for 20 min, rinsed with water and destained in destaining solution (50% [v/v] ddH_2O; 40% [v/v] methanol; 10% [v/v] acetic acid) for 2 hrs

to overnight. PageRuler Unstained Protein Ladder (Thermo Fisher Scientific) (3.5 µL) was routinely included as a size marker to confirm the size of the heterologously expressed protein.

2.17 *In vivo* Stress-tolerance Screening of Sub-cloned Genes

2.17.1 Testing for Salt- and Alkali-tolerance

The OD_{600} of overnight induced cultures were measured spectrophometrically and normalised to an OD_{600} of 0.6 using 0.5% NaCl LB broth containing Ap. Serial dilutions up to 10^{-6} were made in quarter strength Ringer's solution. Each dilution (100 µL) was spread-plated in triplicate on LB agar plates containing 3.5% and 5% NaCl and corresponding control LB agar plates (1% NaCl), supplemented with 0.4 mM IPTG. The survival rate of each culture was calculated. The Shapiro-Wilk test was used to determine whether the survival rate data were normally distributed. Where data were not normally distributed a log-transformation was performed to fit a normal distribution. To determine whether the survival rate was significantly different from the wild-type Student's *t*-test (unpaired) for independent samples was performed. This tests the null hypothesis which states that the means of two independent samples (in this case, the survival rate of the wild-type and experimental cultures on 3.5% NaCl LB agar plates) are the same. A *P*-value of <0.05 allows the null hypothesis to be rejected. Thus, the difference observed between the means of the samples would not be due to chance alone. The same test was used to compare the survival rate of each *E. coli* BL21(DE3) transformant on 3.5% NaCl LB agar to every other culture and not only to the wild-type. R software (version 3.1.0) was used to perform all statistical tests.

Salt-tolerance in liquid culture was tested by inoculating overnight induced cultures 1:100 into LB broth containing 1% and 5% NaCl, as well as alkaline LB broth (pH adjusted to 9 using 3 M NaOH) containing 1% and 5% NaCl. 220 µL of each culture was transferred to a sterile 96-well microplate (Greiner Bio-one, Kremsmünster, Autria). Plates were incubated at 37°C in the Multiskan Go spectrophotometer (Thermo Fisher Scientific) for 72 hrs. Optical density was measured at 600 nm (OD_{600}) automatically every hour for the first 12 hrs and every 3 hrs thereafter.

2.17.2 Testing for Heat-tolerance

The OD_{600} of overnight induced cultures were measured spectrophometrically and normalised to an OD_{600} of 0.6 using 0.5% NaCl LB broth containing Ap. Serial dilutions up to

10^{-6} were made in quarter strength Ringer's solution. Each dilution (100 µL) was spread-plated in triplicate on LB agar plates, incubated at 46°C for 48 hrs, and on corresponding control LB agar plates incubated at 37°C for 24 hrs.

Chapter 3: Functional Screening of a Namib Desert Soil Metagenomic Library for Stress-tolerant Clones

3.1 Introduction

Functional metagenomic screening can result in the identification of entirely novel stress-related genes, as illustrated by CULLIGAN *et al.* (2013). The discovery of novel genes, or genes not previously implicated in stress-response, is not dependent on databases of known genes and this technique is thus capable of vastly expanding our understanding of how microorganisms respond to abiotic stress (HANDELSMAN 2011). Various functional screening methods can be used to detect stress-response genes from a metagenomic library. A simple and effective method is screening for biological activity, such as stress complementation under conditions that normally inhibit growth of the host strain.

In this chapter, a functional metagenomic approach was used to screen a Namib Desert soil library (section 2.7) for genes conferring stress-resistance to an *E. coli* host under salt (NaCl), oxidative (H_2O_2), heat and UVB radiation stress. This is the first study to elucidate which stress-adaptation genes are present in microbes surviving in Namib Desert soil.

3.2 Survival Assays of *E. coli* EPI300:pCC1FOS Under Different Abiotic Stresses

In order to determine the survival threshold of the metagenomic library host, *E. coli* EPI300:pCC1FOS was subjected to increasing intensities of each stress. This was achieved by increasing the concentration of NaCl in solid media for hyperosmotic stress, H_2O_2 in solid and liquid media for oxidative stress, increasing incubation temperature for heat stress and extending exposure time to UVB radiation, until a survival rate of 0% was observed. All results shown are the average of triplicate experiments.

3.2.1 Hyperosmotic Stress-tolerance Assays

The survival rates of *E. coli* EPI300:pCC1FOS cultures in exponential phase challenged with hyperosmotic stress through the addition of NaCl to LB agar were calculated (Table 6). *E. coli* EPI300:pCC1FOS cultures were unable to grow on media containing 4% or 5% NaCl.

Table 6. Survival rate of *E. coli* EPI300:pCC1FOS under increasing NaCl concentrations.

NaCl Concentration in Solid Media (%)	Survival Rate (%) ± SEM*
1	100 ± 0.0
2	40.4 ± 1.2
3	11.6 ± 4.8
3.5	11.6 ± 4.1
4	0 ± 0.0
5	0 ± 0.0

*SEM: Standard error of the mean.

3.2.2 Oxidative Stress-tolerance Assays

The survival rates of *E. coli* EPI300:pCC1FOS cultures in exponential phase challenged with oxidative stress through the addition of H_2O_2 to LB broth were calculated (Table 7). *E. coli* EPI300:pCC1FOS cultures were unable to survive at final concentrations of 5 or 6 mM H_2O_2.

Table 7. Survival rate of *E. coli* EPI300:pCC1FOS under increasing H_2O_2 concentrations in liquid culture.

H_2O_2 Concentration in Liquid Media (mM)	Survival Rate (%) ± SEM
0	100 ± 0.0
1	0.02 ± 0.001
2	$4.9 \times 10^{-3} \pm 6.6 \times 10^{-4}$
3	$1.2 \times 10^{-3} \pm 1.8 \times 10^{-4}$
4	$2.1 \times 10^{-4} \pm 6.4 \times 10^{-5}$
5	0 ± 0.0
6	0 ± 0.0

The survival rates of *E. coli* EPI300:pCC1FOS challenged with oxidative stress through the addition of H_2O_2 to LB agar supplemented with H_2O_2 were calculated (Table 8). Cultures were unable to grow on media containing H_2O_2 at a final concentration of 1 mM or more.

Table 8. Survival rate of *E. coli* EPI300:pCC1FOS under increasing H_2O_2 concentrations in solid media.

H_2O_2 Concentration in Solid Media (mM)	Survival Rate (%) ± SEM
0	100 ± 0.0
0.5	$2.1 \times 10^{-3} \pm 4.3 \times 10^{-4}$
1	0 ± 0.0
1.5	0 ± 0.0
2	0 ± 0.0
2.5	0 ± 0.0

3.2.3 Heat-tolerance Assays

The survival rates of *E. coli* EPI300:pCC1FOS cultures incubated at 45.5°C and 46°C for 24 hrs on LB agar were calculated (Table 9). Cultures were unable to tolerate temperatures of 46°C, or more.

Table 9. Survival rate of *E. coli* EPI300:pCC1FOS at high temperatures.

Temperature (°C)	Survival Rate (%) ± SEM
37	100 ± 0.0
45.5	$7.8 \times 10^{-4} \pm 2.3 \times 10^{-4}$
46	0 ± 0.0

The colonies observed on LB agar plates incubated at 45.5°C were colourless, as opposed to the cream/ beige colour of healthy *E. coli* colonies. These colourless colonies were inoculated into LB broth and incubated overnight at 37°C, with shaking. No growth was observed in liquid culture, which showed that the colonies initially grew, but could not sustain growth at 45.5°C and were unable to grow at 37°C thereafter.

3.2.4 UVB Radiation-tolerance Assays

The survival rates of *E. coli* EPI300:pCC1FOS cultures exposed to UVB radiation (280 – 315 nm) on LB agar for various periods and subsequent incubation for 24 hrs at 37°C were calculated (Table 10). Cultures were unable to tolerate exposure to UVB radiation for periods of 45 s, or more.

Table 10. Survival rate of *E. coli* EPI300:pCC1FOS under increasing exposure time to UVB radiation.

UVB Exposure Time (s)	Survival Rate (%) ± SEM
0	100 ± 0.0
10	$5.8 \times 10^{-4} \pm 4.0 \times 10^{-4}$
15	$4.9 \times 10^{-3} \pm 4.2 \times 10^{-4}$
20	$5.4 \times 10^{-4} \pm 1.9 \times 10^{-4}$
30	$3.3 \times 10^{-4} \pm 1.8 \times 10^{-4}$
45	0 ± 0.0

All surviving colonies were located around the margins of the LB agar plates, where the flux of UVB radiation was possibly reduced by the sides of the petri dish. At 45 s, no growth was observed on any part of the plates.

3.3 Screening the Namib Desert Soil Metagenomic Library for Stress-resistant Clones

A Namib Desert soil metagenomic library containing approximately 1.15×10^6 clones was screened for stress-tolerance genes using the biological activity functional screening method. Stress-resistant library clones were identified as those surviving at a stress level at which *E. coli* EPI300:pCC1FOS growth was completely inhibited, i.e. where 0% survival of the host was observed.

3.3.1 Hyperosmotic Stress-tolerance Screening

The metagenomic library was screened for salt-tolerance at a final concentration of 5% NaCl in LB agar. After 72 hrs incubation at 37°C, sub-library cultures 3:1 S-JJ, 1:3 S-JJ and 1:3 AE-AQ showed growth on all three replicate LB agar plates, with a total of 114 colonies (see section 2.7 for information on how the metagenomic library was labelled and stored). The library host had 3.18×10^7 cfu/mL. on LB agar. This value was used to calculate the survival rate of the salt-tolerant clones (STCs). Table 11 shows the distribution of these clones across the library cultures and survival rate of the cultures.

Table 11. Survival rate of salt-tolerant cultures on 5% NaCl LB agar.

Sub-library	Number of Colonies	Survival Rate (%) ± SEM
3:1 S-JJ	97	$1.0 \times 10^{-3} \pm 6.8 \times 10^{-5}$
1:3 S-JJ	10	$1.0 \times 10^{-4} \pm 2.5 \times 10^{-5}$
1:3 AE-AQ	7	$7.3 \times 10^{-5} \pm 2.5 \times 10^{-5}$

These STCs were analysed further as described in section 3.5, 3.6 and Chapter 4.

3.3.2 Oxidative Stress-tolerance screening

The first screening of the Namib Desert soil pooled library in liquid culture at a final H_2O_2 concentration of 6 mM resulted in substantial growth of both library and control cultures. The library screening was subsequently performed at a final H_2O_2 concentration of 7 mM. Six colonies were obtained across all library cultures: two colonies from sub-library 1:3 KK – AD, one colony from 1:1 KK – AC and three colonies from 1:3 A – R. The colonies were inoculated into LB broth supplemented with Cm and the experiment was repeated at final H_2O_2 concentrations of 7, 8 and 9 mM. No growth was observed on any of the plates and it was concluded that these clones were false-positives.

The library was also screened for oxidative stress-tolerant clones by adding H_2O_2 to LB agar at a final concentration of 1 mM. After 72 hrs incubation at 37°C, no growth was observed on any of the library plates. The stringency of the screen was therefore decreased to a concentration of 0.75 mM H_2O_2. Five irregularly shaped clusters of growth were observed in sub-library 1:1 A – R after 72 hrs. A sample of each cluster was streak-plated on a LB agar plate and a single colony of each was picked and subjected to screening at 0.75 mM H_2O_2. After 72 hrs incubation at 37°C, no growth was observed in the centre of the plates, although some colonies had grown around the margins of the plate where condensation on the curve of the agar was observed. It was concluded that these colonies were probably not resistant to oxidative stress. No growth was observed in the control culture, thus the stringency of the screen was further reduced to 0.7 and 0.725 mM H_2O_2. No growth was observed in library cultures at either H_2O_2 concentration.

An alternative strategy of enriching oxidative stress-tolerant clones within the library was attempted. The entire Namib Desert soil metagenomic library was pooled into a single culture. This liquid culture, along with the control culture (*E. coli* EPI300:pCC1FOS), was treated with 0.725 mM H_2O_2 before being screened on LB agar plates supplemented with

0.725 mM H_2O_2. A control culture, not treated with H_2O_2, was also included. The principle behind this strategy was that clones resistant to this level of H_2O_2 should have a growth advantage over sensitive clones. Resistant clones should be dominating the pooled culture, making their identification within the library more likely after enrichment. However, no resistant library clones were identified.

It was concluded that the screening methods used here were not sufficiently sensitive to identify oxidative stress-tolerant clones.

3.3.3 Heat-tolerance Screening

The screening of the Namib Desert soil metagenomic pooled library at 46°C resulted in the growth of a total of 23 colonies (Table 12). In order to qualify as a positive hit, growth had to be observed across all three replicate cultures of the sub-library.

Table 12. Putative heat-tolerant clones identified in the Namib Desert soil metagenomic library.

Sub-library	Number of Colonies
3:1 A – R	3
3:1 AD – AN	6
1:1 S – JJ	3
1:3 S – JJ	3
1:3 KK – AD	8

In order to confirm a heat-tolerant phenotype, the putative heat-tolerant clones were screened at 46°C again. Each culture was spot-plated (3 µL), in triplicate, onto LB agar plates and incubated at 46°C for 48 hrs. After this treatment, only four cultures showed growth and were further evaluated. These cultures were: 3:1 A – R colony 1, 3:1 AD – AN colony 3, 1:3 KK – AD colony 7 and 1:3 KK – AD colony 8. Five replicates of each culture were re-plated (3 µL) onto LB agar and incubated at 46.3°C and 46.5°C for 96 hrs. Neither culture 3:1 A – R colony 1 nor the control culture showed growth at these temperatures. Therefore, 3:1 AD – AN colony 3 and 1:3 KK – AD colonies 7 and 8 were identified as heat-tolerant clones (HTCs). These cultures, designated as HTC 1, 2 and 3, were stored as 30% glycerol stocks. The investigation of these clones will be the subject of another Master's project.

3.3.4 UVB Radiation-tolerance Screening

The Namib Desert soil metagenomic pooled library was screened for UVB radiation tolerant clones at the 45 s exposure threshold. No surviving colonies were identified. The exposure time was subsequently decreased to 30, 25, 20, 15 and 10 s, but no surviving colonies were identified. It was concluded that the UVB screening method used here was not sufficiently sensitive to identify UVB radiation resistant clones.

3.4 Fosmid DNA Isolation

3.4.1 Comparison of Three Plasmid DNA Isolation Methods

Three different plasmid extraction methods were evaluated for the extraction of high-yield, high-quality fosmid DNA. These methods were: alkaline lysis miniprep, QIAprep Spin Miniprep kit (Qiagen) and non-ionic detergent (NID) plasmid extraction (LEZIN *et al.* 2011). These were evaluated for DNA yield and quality (as indicated by the $A_{260/280}$ and $A_{260/230}$ ratios). A random clone, STC 63, was selected for this experiment. STC 63 was extracted with all three methods, in triplicate. The DNA was analysed by agarose gel electrophoresis and quantified using ImageLab software. DNA extracts were also analysed with a Nanodrop 2000 (Thermo Fisher Scientific) (Table 13 and Figure 1).

Table 13. STC 63 DNA quantity and quality data obtained after extraction with three different methods.

Extraction Method	Concentration (NanoDrop) (ng/µL)	Concentration (ImageLab) (ng/µL)	$A_{260/280}$	$A_{260/230}$
Alkaline lysis 1	3108	13.5	2.03	2.23
Alkaline lysis 2	3078	23.6	2.04	2.24
Alkaline lysis 3	3675	11.6	2.07	2.31
Qiagen 1	26.1	55.2	1.98	1.82
Qiagen 2	25.3	37.6	1.98	1.69
Qiagen 3	21.1	53	2.02	1.64
NID 1	300	12	1.71	0.71
NID 2	382	12.6	1.72	0.77
NID 3	416	13.2	1.91	1.15

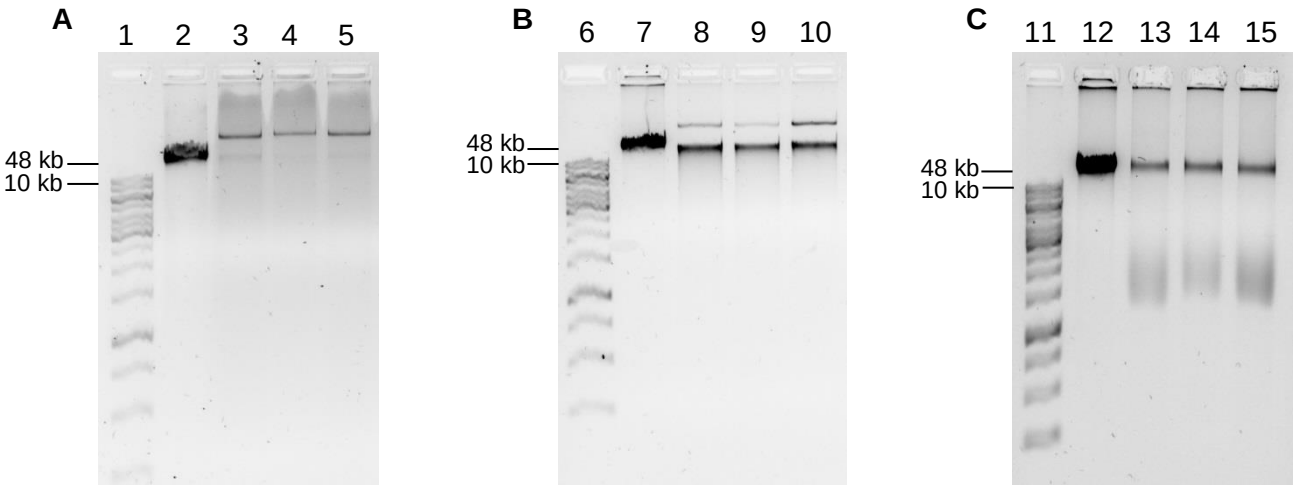


Figure 1. Extraction of STC 63 using three different plasmid extraction methods. A) Alkaline lysis method, B) QIAprep Spin Miniprep kit (Qiagen), C) Non-ionic detergent method. Lanes 1, 6 and 11: GeneRuler 1 kb DNA Ladder (Fermentas). Lanes 2, 7 and 12: Lambda DNA (50 ng). Lanes: 3 – 5, 8 – 10 and 13 – 15: STC 63 DNA extractions.

The DNA quantification values from the NanoDrop (Table 13) varied considerably between the three methods, whereas the ImageLab software used to estimate DNA concentrations from the intensity of the GeneRuler bands on the agarose gel images (Figure 1) was more consistent. It was calculated that the Qiagen kit provided the highest yield of fosmid DNA, supported by the visual intensity of the bands in the gel image (Figure 1). Two bands were observed in lanes 8, 9 and 10 in Figure 1 B. These bands were assumed to be different isoforms of the fosmid and both bands were used for the quantification of the DNA using the ImageLab software. The $A_{260/280}$ ratios (Table 13) of all samples were acceptable (ideal readings should be between 1.8 and 2.0) showing that the DNA samples were relatively free of protein contamination. The $A_{260/230}$ ratios of the phenol-chloroform and Qiagen methods were at acceptable levels (ideal readings should be between 2.0 and 2.2), showing the DNA samples were relatively free of organic contaminants, such as phenol. The average $A_{260/230}$ ratio of the Qiagen samples was 1.72, which is acceptable. However, the NID samples had extremely low $A_{260/230}$ ratios. This was probably due to residual EDTA, which absorbs near 230 nm, from the extraction buffer contaminating the DNA sample. The ability of the *Nco*I restriction enzyme to digest NID-extracted fosmids was evaluated (Figure 2). STC 1 and 10 extracted using the NID method and the fosmid (pCC1FOS) extracted using the QIAprep Spin Miniprep kit (Qiagen), were digested with the *Nco*I enzyme. The enzyme was unable to digest the NID-extracted fosmids, showing that this method did not provide a DNA sample that is sufficiently pure for use in downstream enzymatic reactions. On the basis of these experiments, the Qiagen kit was elected as the extraction method that provided the ideal balance between yield and quality of fosmid DNA.

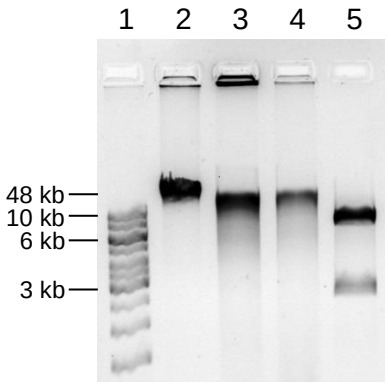


Figure 2. *Nco*I digestion of STC 1, STC 10 and pCC1FOS. Lane 1: GeneRuler 1 kb DNA Ladder (Fermentas). Lane 2: Lambda DNA (50 ng). Lane 3: STC 1 extracted using the NID method. Lane 4: STC 10 extracted using the NID method. Lane 5: pCC1FOS extracted using the Qiagen kit.

3.5 Extraction of Fosmid DNA from STCs and Restriction Digestion Analysis

Fosmids were extracted using the QIAprep Spin Miniprep kit (Qiagen) and 15 of the 114 STCs (1 – 15) were subjected to *Nco*I digestions (Figure 3 and Figure 4). After successful extractions and digestions of these 15 clones, the remainder of the STCs were extracted with the Qiagen kit and subjected to *Nco*I restriction digestion analysis (Appendix A: Figure A1 and A2). This method was used to distinguish clones from each other, as each unique clone provided a unique restriction pattern. A total of 12 STCs unique clones were identified (STC 1, 9, 11, 12, 13, 18, 82, 88, 89, 90, 91 and 92). These were further analysed as described in section 3.6 and Chapter 4.

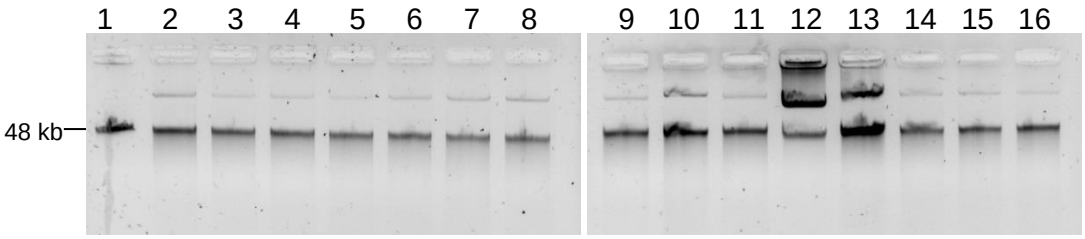


Figure 3. Extraction of fosmid DNA from STC 1 – 15 using the Qiagen Miniprep kit extraction method. Lane 1: Lambda DNA (50 ng). Lanes 2 – 16: STC 1 – 15.

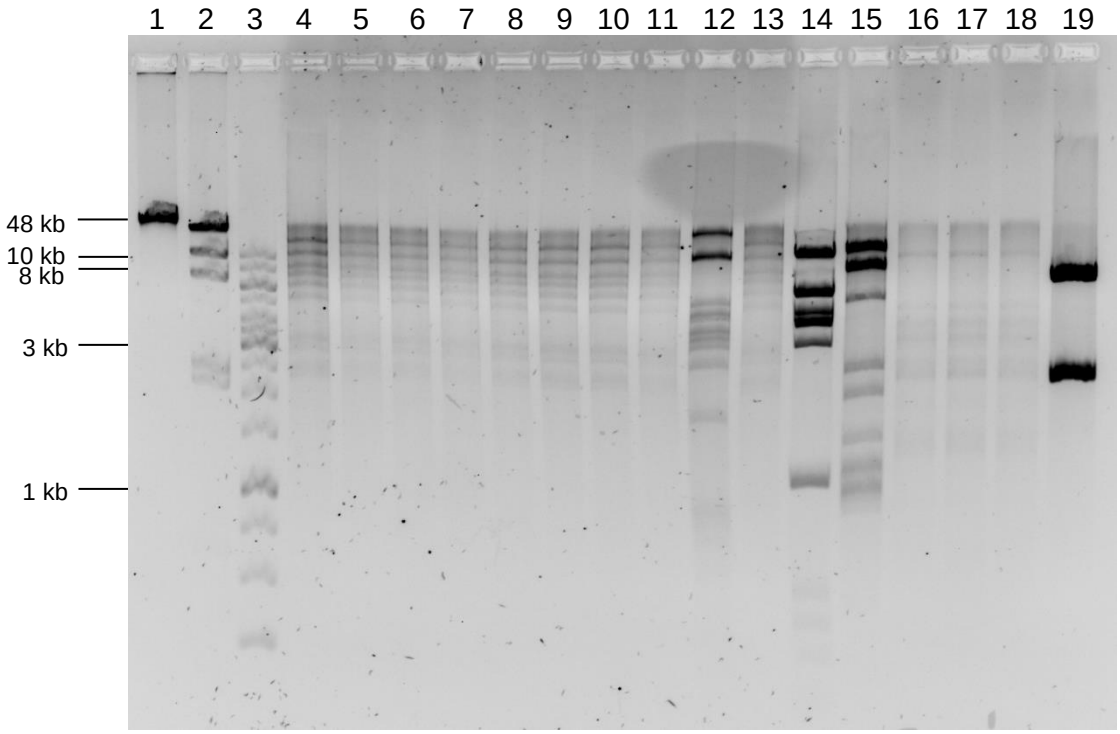


Figure 4. *Nco*I restriction digestion analysis of fosmid DNA extracted from STCs 1- 15. Lane 1: Lambda DNA (50 ng). Lane 2: Lambda DNA digested with *HindIII* restriction enzyme. Lane 3: GeneRuler 1 kb DNA Ladder (Fermentas). Lanes 4 – 18: STC 1 – 15. Lane 19: pCC1FOS subjected to *Nco*I digestion.

3.6 STC Growth Rate Analysis

The ability of the 12 STCs to confer a survival or growth advantage over *E. coli* EPI300:pCC1FOS (control) in liquid culture supplemented with 5% NaCl was investigated. The OD_{600} reading at each time-point was obtained by averaging the readings from triplicate biological experiments and triplicate technical repeats of the measurements at each time-point. The OD_{600} at 48 hrs was used to determine whether cultures showed significantly greater growth in 5% NaCl LB broth compared to the host. All STCs, except STC 9, conferred a statistically significant growth advantage to *E. coli* EPI300 when compared to the host containing an empty vector (Table 14 and Figure 5).

Table 14. Statistical significance of STC growth after 48 hrs determined using Student's unpaired t-test.

Culture	*P*-value*
STC 1	$10.0e^{-05}$
STC 9	0.91
STC 11	$2.1e^{-04}$
STC 12	$7.8e^{-05}$
STC 13	$1.2e^{-05}$
STC 18	$1.4e^{-04}$
STC 82	$1.8e^{-04}$
STC 88	$8.3e^{-05}$
STC 89	0.047
STC 90	$2.6e^{-04}$
STC 91	0.027
STC 92	$3.1e^{-05}$

**P*-values <0.05 show significant differences in growth compared to the host.

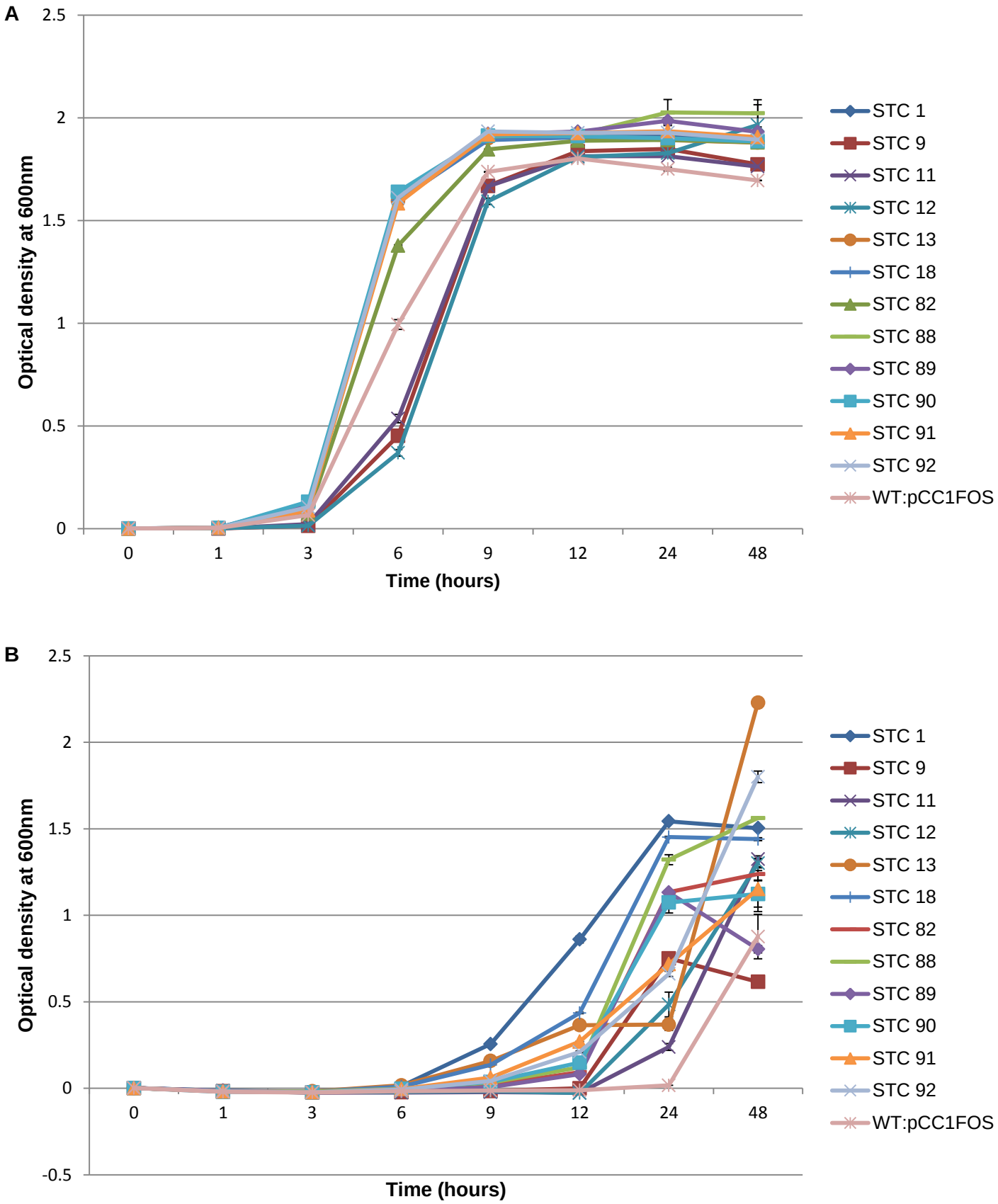


Figure 5. Growth of STCs identified from a Namib Desert soil metagenome and *E. coli* EPI300:pCC1FOS (WT:pCC1FOS) in A) LB broth supplemented with Cm and B) LB broth supplemented with Cm and amended with 5% NaCl. Results are shown as the average of triplicate experiments, with error bars showing the standard error of the mean (SEM).

3.7 Discussion

The hyperosmotic osmotic stress screening approach used in this study (biological activity screening or survival assays) was simple and effective. Here, 114 salt-tolerant clones (STC) were identified from a Namib Desert soil metagenomic library containing approximately 1.15 x 10^6 clones. This equates to a positive hit-rate of 0.01% for STCs, which is reduced to 0.001% after identification of unique insert sequences, as shown by *Nco*I digestion analysis. CULLIGAN *et al.* (2013) identified 53 STCs from a human gut metagenomic library containing 23 000 clones via biological activity screening, with a hit-rate of 0.23%. A possible explanation for this higher hit-rate could be because microbes in the human gut are constantly exposed to osmotic stress; whereas microbes in hyperarid desert soils are probably better adapted to desiccation than osmotic stress, when compared to microbes from the human gut. Furthermore, only hyperosmotic stress induced by an ionic chemical (NaCl) was investigated in this study. The ability of the STCs to tolerate non-ionic hyperosmotic stress, such as that induced by mannitol, could be investigated in future to identify genes that provide protection to ionic- and non-ionic-induced hyperosmotic stress.

Three heat-tolerant clones (HTC) were identified from the Namib Desert soil metagenomic library screened in this project. This was expected, as soil temperatures in the Namib Desert soil can rise to almost 60°C (unpublished measurements taken by Dr J. B. Ramond). Investigating the mechanisms of heat-tolerance provided by these clones could be the subject of a future Masters project.

Screening pooled libraries, as in this study, increased the chance that any two stress-tolerant clones contained the same metagenomic insert and a method to distinguish unique clones, such as restriction enzyme digestion had to be employed. This approach reduced the number of STCs identified in this study from 114 to 12 (Figure 4 and A2). This redundancy within the metagenomic library could be avoided by individually picking and end-sequencing each clone during library construction. This would ensure only unique clones were stored. The library would be smaller, with little or no redundancy. However, this approach can be cost-prohibited.

No UVB radiation- or oxidative stress-tolerant clones were identified in this study. UVB radiation can cause pyrimidine dimers in cellular DNA and can also result in DNA single- and double-strand breaks (DSBs) (CADET *et al.* 2005; BAUERMEISTER *et al.* 2009). Reactive oxygen species can also induce DSBs in DNA (SLADE AND RADMAN 2011). Such damage could potentially inactivate the chloramphenicol (Cm) resistance gene in the fosmid DNA,

rendering the library clones sensitive to Cm, which was routinely included in growth media. Alternatively, the UVB screening method used could not be sufficiently sensitive to detect UVB radiation resistant clones. More sophisticated methods of UV-tolerance screening could be employed, such as those described by BAUERMEISTER *et al.* (2009) to study radiation-resistance of *Deinococcus radiodurans*. Cells were exposed to UV wavelengths of 315 – 400 nm (UVA) and 280 – 400 nm (UVA and UVB) using a combination of high pressure metal halogenide lamps and filters, which allow only light of the desired wavelength to pass through. The employment of such methods was beyond the scope of this project, where four different abiotic stresses were investigated. Enrichment of oxidative stress-tolerant clones in the Namib Desert soil metagenomic library was attempted over 24 hrs. However, this enrichment may have not been performed for a sufficient period of time. Enrichment could be attempted over a period of 10 days with daily sub-culturing into fresh media containing a low concentration of H_2O_2, after which the detection of oxidative stress-tolerant clones could be attempted by screening on solid media containing H_2O_2.

Growth experiments of STCs showed all but two clones had significantly greater tolerance to 5% NaCl after 48 hrs, compared to the host strain. Growth of the library host was inhibited at 5% NaCl up to 24 hrs, after which growth was observed. Upon hyperosmotic shock, *E. coli* synthesises the compatible solute trehalose to balance the water potential across the plasma membrane (WOOD 2011). High amounts of trehalose would be required to achieve this balanced state. It was hypothesised that sufficient amounts of trehalose had been synthesised after 24 hrs to obtain a balanced osmotic state across the cell membrane and growth could continue. The same principle can be used to explain the sudden increase in growth of STC 13 after 24 hrs. As a result of additional genes present on the metagenomic insert potentially involved in mediating hyperosmotic stress, STC 13 could achieve rapid growth after the initial osmotic shock was overcome, resulting in significantly greater salt-tolerance compared to the host.

In this study, functional screening of a Namib Desert soil metagenomic library resulted in the identification of both salt- and heat-tolerant clones. These are prevalent abiotic stresses imposed on microbes living in Namib Desert soil (STOMEO *et al.* 2013; WARREN-RHODES *et al.* 2013). Investigating the genes responsible for conferring salt- and heat-tolerance to these clones could result in the identification of novel stress-tolerance genes.

Chapter 4: Bioinformatic and Phylogenetic Analysis of Metagenomic DNA from Salt-tolerant Clones

4.1 Assembly of Sequencing Data

Fosmid DNA of unique STCs identified in section 3.5 was isolated and sequencing was performed using the Ion Torrent platform (see section 2.13). The quality of the Ion Torrent sequencing data obtained for each fosmid was analysed on PRINSEQ. Low quality data was trimmed using Python scripts as described in section 2.13. Table 15 shows the sequence data before and after quality control. In total, 12 fosmids were sequenced, with an output of 1.1 Gb of sequence data. The final result of the assembly of each fosmid is shown in Table 16. Coverage was calculated using the following formula: coverage (c) = (LN)/G, where L = average sequence length after trimming (from Table 15), N = number of sequences (obtained from Gap5, used to assemble final contigs representing each fosmid, Table 16) and G = fosmid size (insert + vector).

Assembly of the metagenomic inserts of STC 11 and 82 was unsuccessful, as the sequencing reads were dominated by *E. coli* genomic DNA reads. Contamination with genomic DNA was also observed in the other samples, but to a lesser extent. This can be due to the co-purification of genomic DNA with fosmid DNA during the plasmid extraction procedure, or as a result of contaminating DNA from previous sequencing projects on the Ion Torrent platform. STC 11 and 82 were thus excluded from further analyses.

4.2 Annotation of STCs and Identification of Putative Stress-tolerance Genes

Metagenomic inserts of STCs were annotated using RAST and these annotations were graphically depicted using CLC Main Workbench version 6.8.4. Other open reading frame (ORF) predicting programs (Glimmer, MetaGeneMark, ORFinder, FgenesB and CLC Main Workbench) were used to validate the RAST ORF predictions. Where discrepancies in start or stop codon positions were observed for a particular gene, the consensus positions were accepted and the annotation was edited. Functional assignments by RAST of predicted genes were manually curated by comparing ORFs to the NCBI protein database using BLASTx. Putative stress-related proteins were identified by comparing ORFs to the conserved domain database (CDD) on the NCBI website using BLASTx to elucidate putative functions by identifying conserved functional domains. Tables B1 – B10 in Appendix B list the genes found on each fosmid.

Table 15. Sequence data obtained for each fosmid before and after quality control (trimming).

	Before Trimming			After Trimming		
Fosmid	**#Sequences**	**Mean Sequence Length (bp)**	**#Bases**	**#Sequences**	**Mean Sequence Length (bp)**	**#Bases**
STC 1	448 735	314.63 ± 70.77	141 187 637	403 721	293.39 ± 56.44	118 447 986
STC 9	293 304	330.03 ± 91.35	96 797 904	265 240	311.06 ± 75.93	82 506 356
STC 11	317 334	330.63 ± 84.22	104 920 468	300 094	314.05 ± 66.46	94 243 675
STC 12	374 703	342.46 ± 91.30	128 320 253	353 116	326.37 ± 74.92	115 247 601
STC 13	402 696	336.37 ± 83.75	135 453 666	384 808	317.74 ± 70.98	122 268 886
STC 18	453 652	307.65 ± 74.11	139 564 340	431 569	289.74 ± 59.97	125 044 034
STC 82	242 173	328.05 ± 89.04	79 444 634	225 054	311.92 ± 71.13	70 198 548
STC 88	263 308	328.91 ± 82.25	86 604 419	249 868	311.05 ± 66.55	77 720 350
STC 89	296 402	329.08 ± 78.80	97 539 029	282 387	310.08 ± 64.15	87 563 664
STC 90	199 216	321.31 ± 84.77	64 010 317	186 383	305.12 ± 66.76	56 868 369
STC 91	262 032	318.17 ± 83.49	83 372 009	241 084	300.16 ± 66.79	72 364 510
STC 92	309 966	346.59 ± 94.26	107 430 089	293 742	329.04 ± 80.85	96 651 673
Average/ Total	**3 863 521**	**327.60 ± 83.61**	**1 264 644 765**	**3 617 066**	**309.76 ± 68.04**	**1 119 125 652**

Table 16. Assembly results obtained through MIRA-Gap5 pipeline.

Fosmid	#Contigs	#Sequences	Insert Size (bp)	Fosmid Size (bp)	xCoverage
STC 1	1	76 640	31 296	39 435	570
STC 9	11	19 999	34 057	42 196	147
STC 12	5	2 656	21 038	29 177	31
STC 13	3	33 847	26 333	34 472	312
STC 18	2	66 630	31 971	40 110	481
STC 88	10	24 115	24 944	33 083	227
STC 89	4	34 010	20 971	29 110	362
STC 90	6	23 537	31 907	40 046	179
STC 91	7	21 818	20 098	28 237	232
STC 92	3	55 817	31 829	39 968	460
Average	**5**	**35 907**	**27 444**	**35 583**	**321**

4.2.1 STC 1, 9, 12 and 18

The *Nco*I restriction digestion analysis showed different profiles for STC 1, 9, 12 and 18. However, sequencing and annotation (Appendix B: Figure B2, Tables B1 – B4) showed that these fosmids contained many of the same genes. Local alignment of the *recA* present on all four fosmids showed 100% nucleotide identity between STC 1 and STC 9 and 99% identity between STC1 and STC 12 and 18 (Figure B1). The 99% identity is due to the insertion of a cytosine at position 1044 of the RecA genes of STC 12 and 18 (position 1057 in the alignment), which resulted in a frame-shift. The RecA genes of STC 12 and 18 were thus 102 nucleotides longer than the RecA genes of STC 1 and 9. This insertion is possibly due to a sequencing error. Therefore, these inserts probably originated from individuals of the same species. However, differences in gene content were still observed, hence the different restriction digestion patterns.

Given the growth advantage that STC 1 and 18 showed compared to STC 9 and 12 in LB broth augmented with 5% NaCl (Figure 5, Table 14), the genetic content of these four clones were compared (Figure 6). The gene or genes responsible for this increased salt-tolerance should be present on the former two and either absent or truncated on the latter two fosmids. It was observed that a RelA-SpoT-like protein (among others) was present on both STC 1 and 18, but absent on STC 9 and 12. This protein, in conjunction with an HD-hydrolase domain-containing protein, is responsible for the synthesis of the bacterial alarmone

(p)ppGpp (GENTRY AND CASHEL 1996; HOGG *et al.* 2004). This molecule induces the stringent response in bacteria during amino acid starvation and other stresses and is also accumulated in plants and yeast in response to various abiotic and biotic stresses (VAN DER BIEZEN *et al.* 2000; BRAEKEN *et al.* 2006; BATTESTI *et al.* 2011; OCHI *et al.* 2012). In Gram-positive bacteria, (p)ppGpp is synthesised by a single enzyme (Rel) containing both the synthetase and hydrolase domain necessary for the reaction. However, in Gram-negative bacteria these domains are often found in two different proteins (HE *et al.* 2013). The closest relative of STC 1 is *Bdellovibrio bacteriovorus*, a Gram-negative bacterium (Figure 8). It was determined through a conserved domain database (CDD) search that the closest BLASTp hit (a hypothetical protein) of the RelA-SpoT-like protein identified in this study contained the synthetase domain and that the HD-hydrolase domain protein (present on all four fosmids) contained the related hydrolase domain. It is possible that STC 9 and 12 showed reduced growth under NaCl stress, compared to STC 1 and 18, because the RelA-SpoT synthetase domain protein was absent. The HD-hydrolase domain cannot produce (p)ppGpp and thus STC 9 and 12 probably have other mechanisms of salt-tolerance that are not as effective as the stringent response. The RelA-SpoT-like and HD-hydrolase domain proteins were thus identified as the most likely mechanism through which STC 1 (and STC 18) conferred salt-tolerance to *E. coli* EPI300. This hypothesis was tested by attempting to amplify the gene encoding the RelA-SpoT-like protein from STC 1, 9 and 12 using the primers in Table 5 and the iProof Polymerase protocol described in section 2.14.1. A band of the expected size (~1.2 kb) was observed in the STC 1 sample, but there was no amplification of DNA of this size in the STC 9 and 12 samples (Figure 7). A band of above 10 kb is present in both of the latter samples and could be attributed to non-specific amplification of DNA. The negative control sample showed no amplification of DNA. The absence of a band corresponding to the RelA-SpoT-like gene in the STC 9 and 12 supports the above hypothesis.

An oligopeptide ABC transporter substrate-binding protein was identified on all four fosmids. These proteins are commonly found as part of uptake ABC-transporter systems (BERNTSSON *et al.* 2010). However, the other components of the ABC transport system were not present on these fosmids. The substrate-binding protein identified in this study contained an OppA-like domain, which facilitates transport of di- and oligopeptides into the cell (MONNET 2003). This could be beneficial during hyperosmotic stress, as proline and glycine betaine are known compatible solutes and could be obtained from such oligopeptides (EMPADINHAS AND COSTA 2006). The above-mentioned proteins are discussed in greater detail in sections 5.1.1 and 5.1.2. The genetic organisation of STC 1, 9, 12 and 18 are graphically presented in Figure B2.

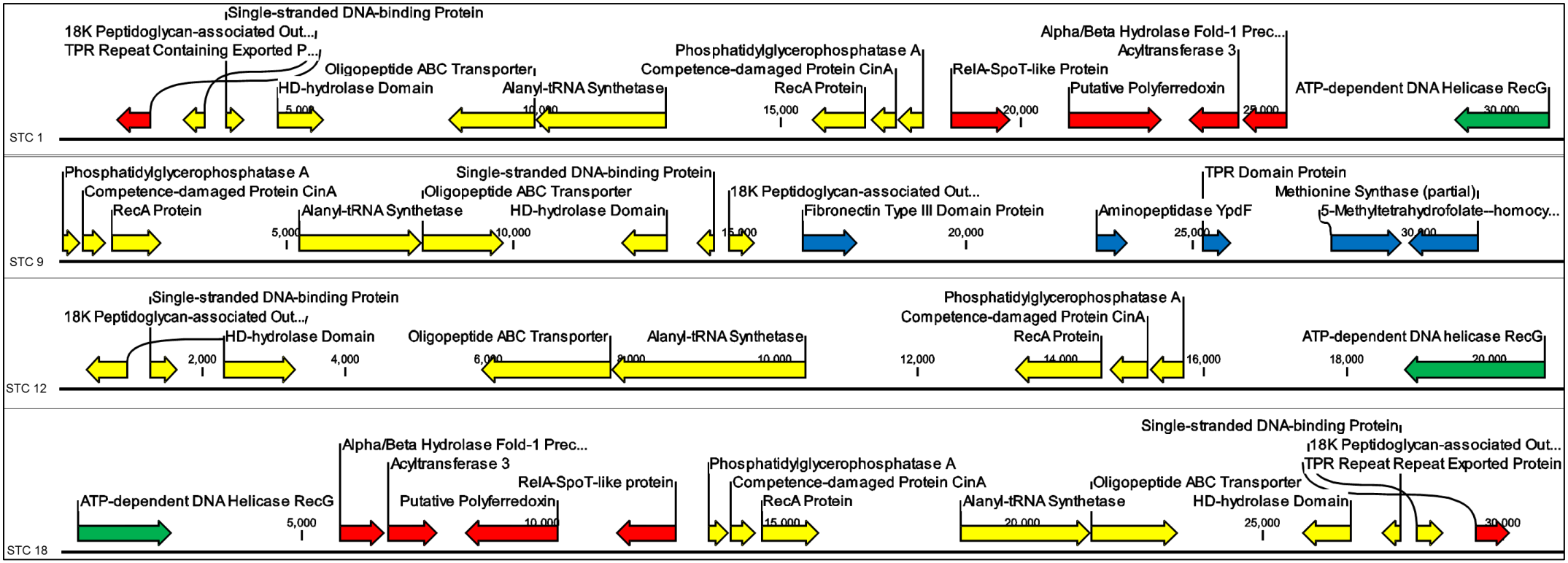


Figure 6. Comparison of the genetic content of STC 1, 9, 12 and 18. Yellow arrows show genes that are shared between all four fosmids. Green arrows show genes that are shared between three fosmids only. Red arrows show genes that are shared between two fosmids only. Blue arrows show genes that are unique to a particular fosmid. Hypothetical proteins were removed from the annotations for visual clarity. Graphical representations are not aligned to scale, which may create the illusion that identical genes are different lengths on different fosmids.

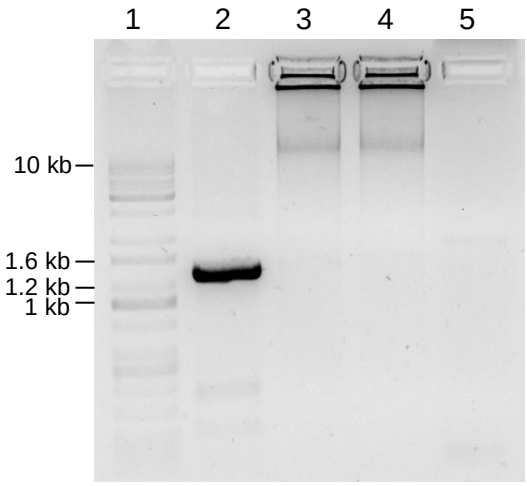


Figure 7. PCR amplification of the gene encoding the RelA-SpoT-like protein from STC 1, 9 and 12. Lane 1: Kapa Universal DNA Ladder (Kapa Biosystems). Lane 2: RelA-SpoT amplicon from STC 1. Lane 3: STC 9. Lane 4: STC 12. Lane 5: Negative control.

4.2.2 STC 13

This fosmid contained a cobalt-zinc-cadmium resistance protein, also annotated as a cation export system protein (Figure B3, Table B5). This is an integral membrane protein containing domains similar to those implicated in cation and heavy metal export. Cation export proteins are important in maintaining membrane potential and balancing osmotic potential across the cellular membrane, particularly during osmotic upshock (SHABALA *et al.* 2009). This cation exporter is further discussed in section 5.1.3.

4.2.3 STC 88

The metagenomic insert of STC 88 (Figure B4, Table B6) encoded a large-conductance mechanosensitive channel protein (MscL), which was identified as potentially being involved in osmotic stress-tolerance. This protein forms a multimeric pore in the cellular membrane and opens in response to membrane tension. Membrane tension increases during hypo-osmotic shock and at a certain tension threshold the channel opens and releases pressure to prevent cell lysis (MOE AND BLOUNT 2005). During hyperosmotic shock, as induced in this project by the addition of NaCl to growth media, the cellular membrane tension decreases as water moves out of the cell. It was thus concluded that this mechanosensitive channel protein would not be investigated further as a potential salt-tolerance protein.

4.2.4 STC 89

The salt-tolerant phenotype of STC 89 could not be assigned to any of the functionally annotated genes present on the fosmid. A penicillin-binding protein, a methyltransferase and three aminotransferase enzymes were identified on STC 89 (Figure B5, Table B7). However, no antibiotic-producing gene clusters were identified on the antiSMASH server, but a putative bacteriocin gene was predicted. Bacteriocins are antibacterial peptides produced by some bacteria and have been considered as possible alternatives to traditional antibiotics (COTTER *et al.* 2013). One of the ways in which bacteriocins kill bacteria is by forming pores in the cellular membrane (COTTER *et al.* 2013). No matches to the putative bacteriocin were found via BAGEL3. Two bacteriocins were present in the BACTIBASE results, Colicin-Ia and Colicin-10 with E-values of 0.55 and 6.5, respectively. These two bacteriocins are channel-forming peptides that act against *E. coli* and closely related species. However, the high E-values indicate that these matches could have occurred by chance alone. When comparing the amino acid sequence of the putative bacteriocin to the NCBI protein database using BLASTp, significant matches (E-value of $3e^{-6}$) to hypothetical proteins only, were found. This leads to two possible conclusions: this gene was incorrectly annotated by antiSMASH or a novel bacteriocin gene has been identified.

4.2.5 STC 90

A notable gene present on STC 90 was a gene encoding a phage-shock protein, PspC, which is a transcriptional regulator that forms part of a stress-response regulon (MODEL *et al.* 1997; DARWIN 2005). The other phage-shock genes that usually form part of this operon, *pspA, -B, -D, -E* and *-F* could not be identified on the metagenomic insert. No genes could be characterised as being involved in salt-tolerance, but a putative thermophilic metalloprotease, peptidase M29, was identified (Figure B6, Table B8). This protease could confer salt- or heat-tolerance to *E. coli* by degrading denatured and aggregated proteins and thus prevent lethal protein aggregates from forming (LIM AND GROSS 2011). This protein was therefore selected for *in vivo* assays of salt- and heat-tolerance and is further discussed in section 5.1.4.

4.2.6 STC 91

STC 91 contained an alkyl hydroperoxide reductase, which been shown to be involved in oxidative stress-tolerance (ROCHA AND SMITH 1999). The presence of genes involved in tolerance to stresses, other than hyperosmotic stress, showed that organisms present in

Namib Desert soil are adapted to the multitude of abiotic stresses of this environment. However, the salt-tolerance phenotype could not be attributed to any of the annotated genes (Figure B7, Table B9).

4.2.7 STC 92

The metagenomic insert of STC 92 contained a gene encoding a Na^+/H^+ antiporter (Figure B8, Table B10). These transport proteins are essential in maintaining intracellular Na^+ and pH levels in bacteria, including halophiles, and have been shown to confer salt- and combined salt- and alkali-tolerance to *E. coli* (MAJERNÍK *et al.* 2001; ZHANG *et al.* 2014). This gene was thus targeted for further investigation under salt and combined salt and alkali stress (section 5.1.5). A potassium (K^+) channel protein was also identified. The influx of K^+ into cells is a well-known response of bacteria to osmotic upshock, as discussed in section 1.3.2. This gene could therefore also be responsible for conferring salt-tolerance. A likely scenario could be that the combined action of these two genes would confer a more salt-tolerant phenotype than either gene in isolation. However, the K^+ channel protein was initially not identified and thus was not sub-cloned for further characterisation.

4.2.8 Hypothetical Proteins

A high proportion of the ORFs predicted on the metagenomic inserts of each fosmid were hypothetical proteins (Table 17). No function could be predicted for the majority of these hypothetical proteins, as many did not contain any putative conserved domains. Furthermore, the salt-tolerant phenotype of STC 88, 89, 90 and 91 could not be attributed to any of the functionally annotated genes present on these fosmids. It could therefore be assumed that one or more of these hypothetical proteins could be responsible for conferring salt-tolerance. The high number of hypothetical proteins predicted on the fosmids in this project is indicative of the sequence novelty of the metagenomic DNA and the limited amount of functional data available in databases.

Table 17. Number of genes present and hypothetical proteins encoded on each fosmid.

Fosmid	No. of Genes	No. of Hypothetical Proteins	% Hypothetical Proteins
STC 1	30	17	57
STC 9	30	17	57
STC 12	22	13	59
STC 18	30	17	57
STC 13	25	12	48
STC 88	28	13	46
STC 89	23	11	48
STC 90	38	12	32
STC 91	24	13	54
STC 92	35	11	31

4.3 Phylogenetic Analysis of the Metagenomic Inserts of the STCs

Section 2.13 provides a detailed description of how phylogenetic analyses were performed. In brief, a housekeeping, or conserved gene, was identified on each STC. The amino acid sequences of this protein and the ten closest homologues were aligned using ClustalW and the alignment curated using Gblocks. The curated alignment was used to construct a neighbour-joining tree, with 1000 bootstrap replicates, using MEGA version 5.2, to discern the origin of the metagenomic DNA insert. The trees were rooted with the branch of the most distantly related species. STC 1 also represents STC 9, 12 and 18, as the metagenomic DNA of these fosmids are thought to originate from the same species.

4.3.1 STC 1, 9, 12 and 18

The ATP-dependent DNA helicase RecG amino acid sequence of STC 1 was used to construct the phylogenetic tree. This protein showed 46% amino acid identity to the *B. bacteriovorus* sequence, which represented the closest related species (Figure 8). This is a Gram-negative bacterium that parasitizes other Gram-negative bacteria (STOLP AND STARR 1963). The metagenomic DNA appears to originate from a member of the phylum Proteobacteria and class Deltaproteobacteria. The habitats of the species in Figure 8 include soil, sewage, aquatic environments (marine and freshwater), hot springs and deep-sea hydrothermal vents, sediments and brine lakes (STOLP AND STARR 1963; SONNE-HANSEN

AND AHRING 1999; LAPIDUS *et al.* 2011; AKLUJKAR *et al.* 2012; ANDERSON *et al.* 2012). *Flexistipes sinusarabici* is described as a halophilic and moderately thermophilic bacterium and *Thermodesulfatator indicus* and *Thermodelsulfobacerium hveragerdense* are thermophilic and together with *Desulfaromonas acetoxidans* are sulphur and sulphate-reducing bacteria (SONNE-HANSEN AND AHRING 1999; BOND *et al.* 2002). This indicated that the STC 1, 9, 12 and 18 metagenomic DNA inserts probably originated from a halotolerant or halophilic thermophile. This was expected, as the Namib Desert soil represents a hot environment, with high concentrations of salts (STOMEO *et al.* 2013).

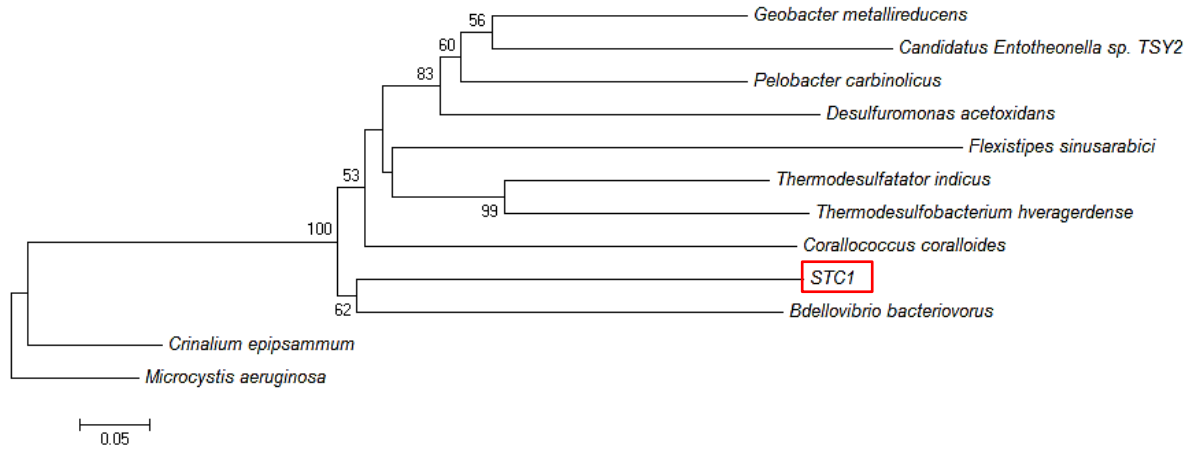


Figure 8. Phylogenetic analysis of ATP-dependent DNA Helicase RecG (representing STC 1, 9, 12 and 18) with closely related proteins. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The tree was rooted with the *Microcytis aeruginosa* branch. The bar represents 0.05 nucleotide substitutions per nucleotide position.

4.3.2 STC 13

The RecA amino acid sequence, which was used to construct the STC 13 phylogenetic tree, showed 76% identity to the *Fibrobacter succinogenes* sequence. This Gram-negative, cellulose-degrading bacterium, which represented the closest related species (Figure 9), is a member of the phylum Fibrobacteres and is one of the three most abundant members of the intestinal microflora in the rumen, across ruminant species (FORSBERG *et al.* 1997). However, no cellulose or polysaccharide-degrading genes were identified on the metagenomic insert of STC 13. Further clustering occurred with two Clostridial species, *Clostridium novyi* and *Desulfotomaculum thermocisternum*. The former is prevalent in soil, aquatic environments and the intestinal tract of humans and animals (NISHIDA AND NAKAGAWARA 1964; SELZER *et al.* 1996; POXTON *et al.* 1997). The latter was isolated from a hot oil reservoir in the North Sea (NILSEN *et al.* 1996). Rumen- or intestine-dwelling bacteria could frequently be introduced to Namib Desert soil microbial communities via the dung of

indigenous herbivorous macrofauna, such as Zebra, Oryx, Springbok and donkeys. The effect of these introductions on the microbial communities has yet to be investigated. Several Proteobacteria are present in the STC 13 tree, with two species overlapping with STC 1.

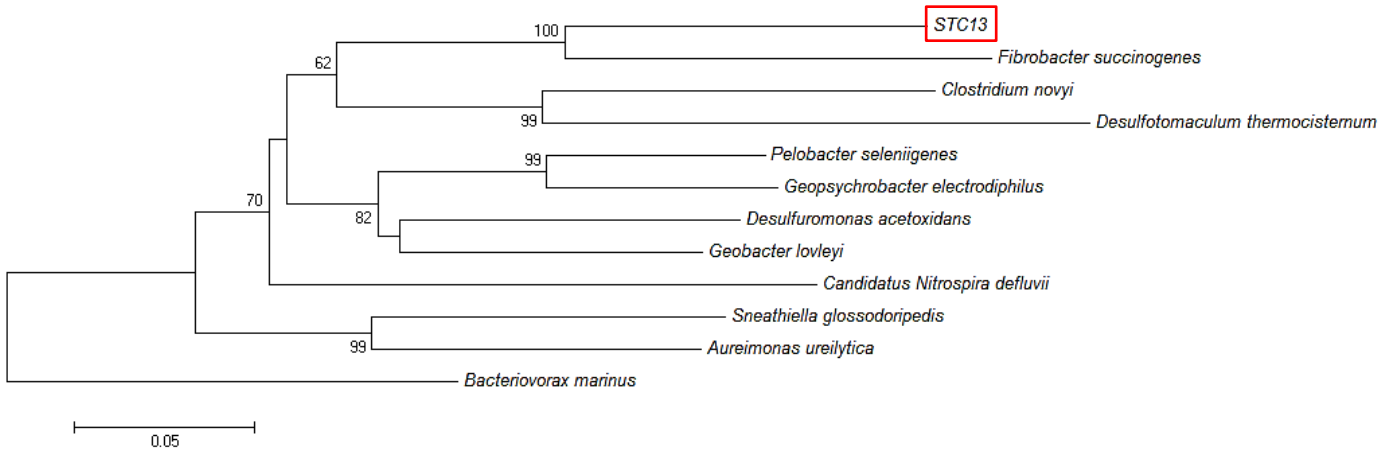


Figure 9. Neighbour-joining phylogeny of STC 13 based on the RecA amino acid sequences. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The tree was rooted with the *Bacteriovorus marinus* branch.

4.3.3 STC 88

Figure 10 shows the phylogenetic relationship of STC 88, based on the aspartyl t-RNA amidotransferase subunit C amino acid sequences. This protein from STC 88 showed 46% amino acid identity to the *Halothece sp. PCC 7418* sequence, which represented the closest related species. *Halothece* are extremely halotolerant Cyanobacteria of the order Chroococcales known for their resistance to harsh conditions (MARGHERI *et al.* 2008). However, the majority of the species in this phylogenetic analysis form part of the phylum Firmicutes, a common phylum in desert soils (POINTING AND BELNAP 2012). The re-occurrence of sulphur- and metal-reducing bacteria *Geobacter metallireducens* and *D. acetoxidans*, as well as Proteobacteria (*Pelobacter*) shows that bacteria capable of these metabolic processes are prevalent in the Namib Desert soil samples used to construct the metagenomic library.

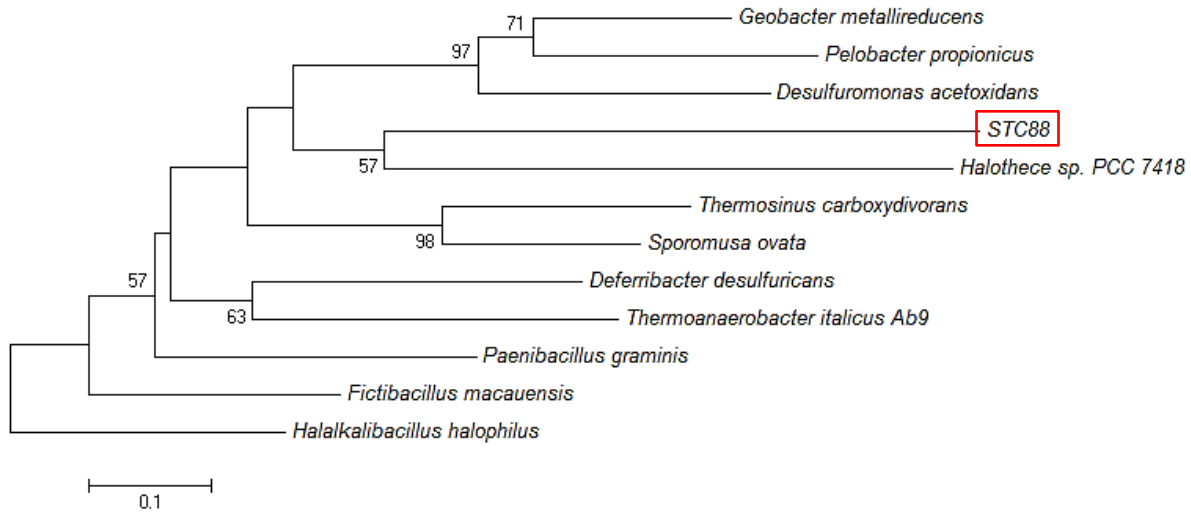


Figure 10. Neighbour-joining phylogeny of STC 88 based on the aspartyl t-RNA amidotransferase subunit C amino acid sequences. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The tree was rooted with the *Halalkalibacillus halophilus* branch. The bar represents 0.1 nucleotide substitutions per nucleotide position.

4.3.4 STC 89

The tyrosyl t-RNA synthetase amino acid sequence was used to construct the STC 89 phylogenetic tree. This protein showed 64% amino acid identity to the *Conexibacter woesei* sequence, which represented the closest related species (Figure 11). The phylogenetic analysis showed that the metagenomic DNA of STC 89 clusters strongly with the Actinobacteria, a common phylum in desert soils (POINTING AND BELNAP 2012). *C. woesei*, *Patulibacter americanus*, *Solirubrobacter soli* and *Rubrobacter radiotolerans* are all of the sub-class Rubrobacteridae and the order Solirubrobacterales. These are Gram-positive bacteria known for the high GC content of their genomes (MONCIARDINI *et al.* 2003). This is advantageous in stress-adaptation, particularly UV-radiation, which induces the formation of thymine dimers, which in turn results in replication errors and the subsequent introduction of mutations. A high GC content is advantageous, as the number of thymine dimers is reduced upon UV exposure due to the reduced frequency of thymine in the genome. The advantage of this adaptation is exemplified by the extremely radiation-resistant bacterium, *R. radiotolerans* (FERREIRA *et al.* 1999). UV-resistant microorganisms are common in desert soils, as there is higher UV irradiance on the surface, compared to temperate environments, due to the reduced moisture content of the atmosphere (POINTING AND BELNAP 2012).

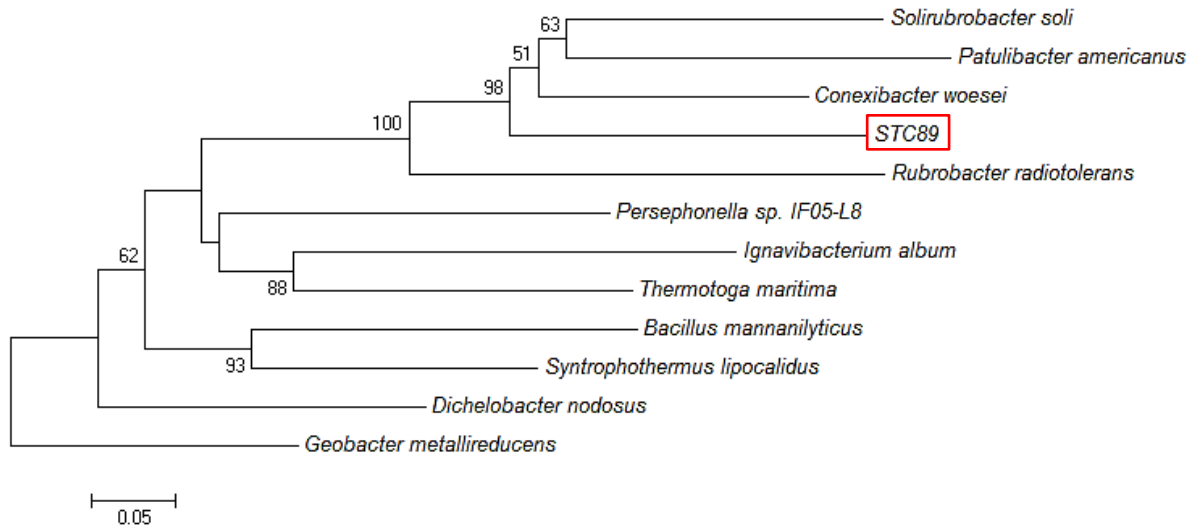


Figure 11. Neighbour-joining phylogeny of STC 89 based on the tyrosyl t-RNA synthetase amino acid sequences. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The *Geobacter metallireducens* branch was used to root the tree.

4.3.5 STC 90

The arginyl t-RNA synthetase amino acid sequence, which was used to construct the STC 90 phylogenetic tree, showed 45% amino acid identity to the *Patulibacter medicamentivorans* sequence, which represented the closest related species and is a Gram-positive bacterium (Figure 12). STC 90 clustered with the Solirubrobacterales, which are bacteria known for the high GC content of their genomes (Almeida *et al.* 2013). A notable feature of STC 90 is that it contained three complete protease-encoding genes and one partial protease gene (Figure B6, Table B8).

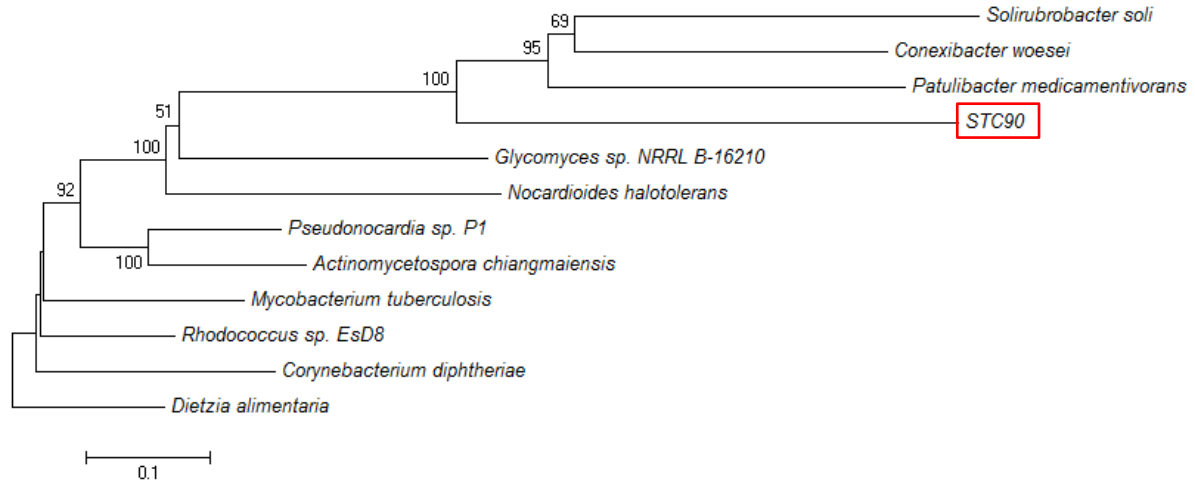


Figure 12. Neighbour-joining phylogeny of STC 90 based on the arginyl t-RNA synthetase amino acid sequences. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The *Dietzia alimentaria* branch was used to root the tree.

4.3.6 STC 91

Figure 13 shows the phylogenetic relationship of STC 91, based on the RNA polymerase sigma-70 factor amino acid sequences. This protein from STC 91 showed 66% amino acid identity to the *Blastopirellula marina* sequence, which represented the closest related species. *B. marina* is a Gram-negative halotolerant marine bacterium (SCHLESNER *et al.* 2004). STC 91 clustered with species from the phylum Planctomycetes and the Family Planctomycetaceae, which include all four species above STC 91 in the tree in Figure 13. These bacteria are important in sequestering carbon from sulfatated heteropolysaccharides within the environments in which they are found, which include marine and terrestrial habitats and have been identified in Namibian coast upwelling systems (WOEBKEN *et al.* 2007).

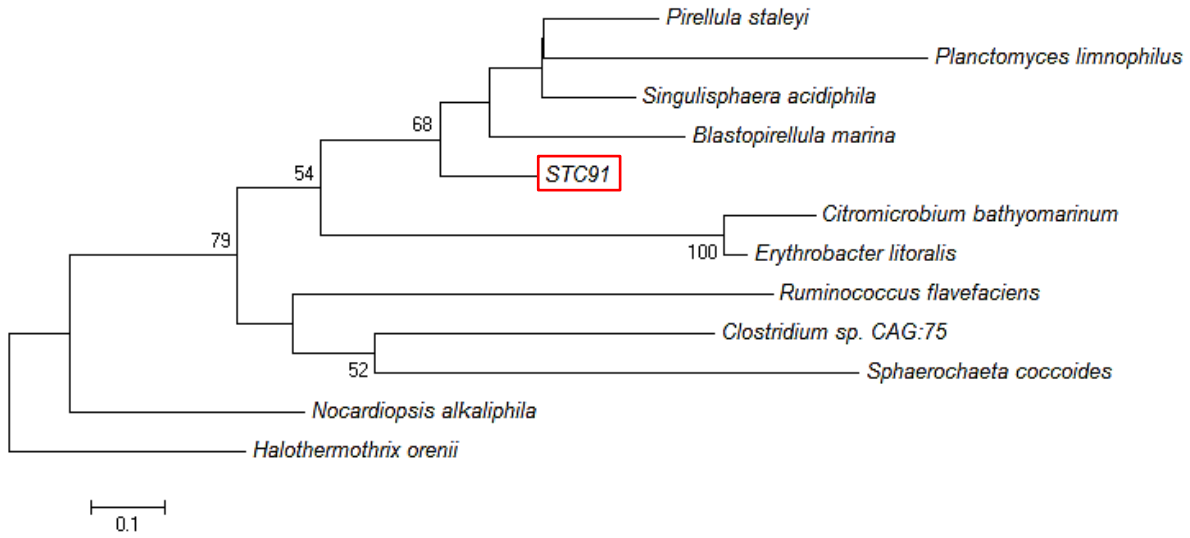


Figure 13. Neighbour-joining phylogeny of STC 91 on the basis of the RNA polymerase sigma-70 factor amino acid sequences. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The *Halothermothrix orenii* branch was used to root the tree.

4.3.7 STC 92

The cell division inhibitor, CDP-paratose 2-epimerase, amino acid sequence was used to construct the STC 92 phylogenetic tree. This protein showed 62% amino acid identity to the *Syntrophobacter fumaroxidans* sequence, which represented the closest related species (Figure 14). This species, as well as *Anaeromyxobacter sp. Fw109-5* and *Geobacter daltonii*, are Deltaproteobacteria. *S. fumaroxidans* is a Gram-negative, aquatic bacterium capable of oxidising propionate and fermenting fumarate (HARMSEN *et al.* 1998). However, it was difficult to draw any conclusions of the fumarate fermenting abilities of the species from

which STC 92 originated, due to the weak phylogenetic clustering observed. The use of threonine dehydratase, also present on STC 92 (Figure B8, Table B10), amino acid sequences to construct the neighbour-joining phylogenetic tree of STC 92 did not improve the clustering or bootstrap values associated with diverging branches (results not shown).

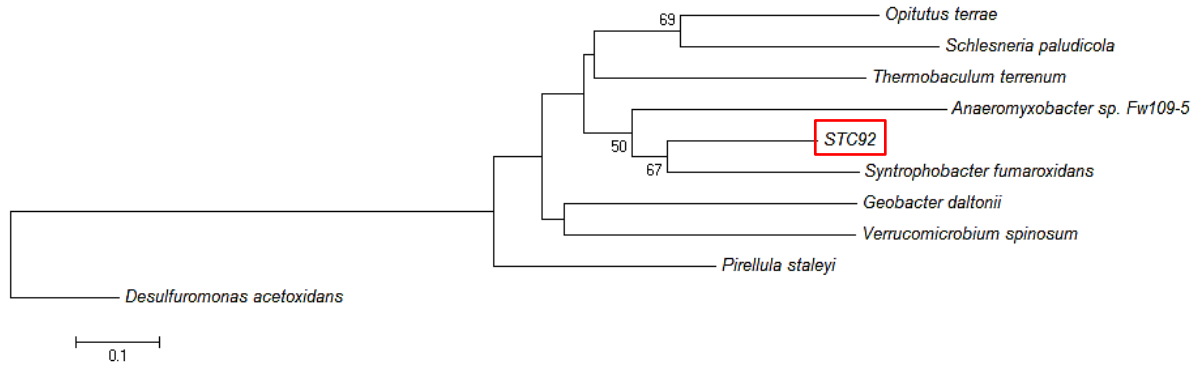


Figure 14. Neighbour-joining phylogeny of STC 91 on the basis of the CDP-paratose 2-epimerase amino acid sequences. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The *Desulfuromonas acetoxidans* branch was used to root the tree.

4.4 GC Content of the Metagenomic DNA of the STCs

The GC content of the metagenomic insert of all STCs was remarkably high, ranging from 63.5 to 70.9%. STC 1, 13, 88 and 91 each had a considerably higher GC content compared to the genome of the closest related species. This could be an adaptation of the microorganisms that survive in the Namib Desert soil in response to high UV irradiance. Furthermore, other abiotic stresses such as desiccation, heat and oxidative stress also damage DNA. It is possible that the observed GC content stabilises DNA structure through base-stacking (YAKOVCHUK *et al.* 2006). STC 89, 90 and 92 each had GC content close to that of its closest relative. A study by KARLIN *et al.* (1997) showed that the dinucleotide CpG is overrepresented in halophiles, but underrepresented in thermophiles. This could explain the increased GC content of the metagenomic inserts conferring salt-tolerance to *E. coli* EPI300. Table 18 shows the GC content of each STC compared to the closest relative as determined in the phylogenetic analysis.

Table 18. The GC content of the metagenomic insert of each STC.

Fosmid	%GC	Closest Relative	%GC
STC 1	63.5	*Bdellovibrio bacteriovorus*	49.9
STC 13	67.9	*Fibrobacter succinogenes*	48.1
STC 88	66.9	*Halothece sp. PCC 7418*	42.9
STC 89	67.0	*Conexibacter woesei*	72.7
STC 90	70.9	*Patulibacter medicamentivorans*	74.1
STC 91	69.4	*Blastopirellula marina*	57.0
STC 92	64.7	*Syntrophobacter fumaroxidans*	59.9

4.5 Discussion

Sequence assembly of 10 fosmids that conferred salt-tolerance to *E. coli* EPI300 was successfully achieved. The annotation of ORFs and assignment of putative function to each gene resulted in the identification of six putative stress-tolerance genes, based on the putative domains identified in the encoded proteins and the general function of the protein families to which these belonged. The putative stress-tolerance genes encoded a RelA-SpoT-like protein and HD-hydrolase domain protein (STC 1), an oligopeptide ABC transporter substrate binding domain (STC 1), a cation export system protein (STC 13), a thermophilic protease, peptidase M29 (STC 90) and a Na^+/H^+ antiporter (STC 92). Even though these were not the only stress-related genes present on the STCs, surprisingly few stress-related genes could be identified across the STCs. The large percentage of hypothetical proteins highlights both the novelty of the metagenomic sequences and the limited functional data available in databases to provide putative functions to novel proteins. Future research could be aimed towards investigating the ability of these hypothetical proteins to confer stress-tolerance, or perform other functions.

A possible explanation why most of the fosmids could not be assembled in their entirety is the presence of genomic DNA contamination from the *E. coli* host. This results in fewer reads mapping to the metagenomic DNA, thereby reducing coverage and decreasing the probability that the assembler, in this case MIRA, would assemble reads correctly. In order to avoid this problem the contaminating genomic DNA could be digested with a specific endonuclease, such as Plasmid-Safe™ DNase (Epicentre Biotechnologies). This DNase uses single-stranded nicks in double-stranded DNA as a starting point for digestion. During fosmid extraction, single-stranded nicks should occur through mechanical methods in

genomic DNA as a result of its large size and thus only genomic DNA will be degraded. However, in experiments not included here it was observed that both genomic and fosmid DNA was degraded by this enzyme. It is thought that due to the relatively large size of the fosmids (~40 kb) single-stranded nicks were also introduced into the fosmid DNA and thus also acted as a substrate for the DNase. An alternative method would be to separate the fosmid DNA from genomic DNA using agarose gel electrophoresis and then excising and purifying the fosmid DNA from the agarose gel. However, if the presence of *E. coli* DNA is due to pre-sequencing contamination, i.e. DNA from a previous project is still present in the capillaries of the Ion Torrent platform, then this issue cannot be prevented with the approached described above. In this case, the contigs representing the metagenomic insert were manually joined by 12 N's. This was sufficient for the purposes of this study. However, to generate complete insert sequences, primer walking could be applied.

A variety of programs were used to predict ORFs in the metagenomic DNA. There were numerous instances where at least one of these programs predicted a different start or stop codon for a particular ORF. This could be explained by the fact that the programs do not all use the same models for ORF prediction. In cases where programs do use the same prediction model, such as the Markov model, differences could be due to the different model training sets used (ANDERSON 2012). The correct prediction of the start codon of an ORF is particularly important if the gene is targeted for subsequent over-expression, purification and functional characterisation, as the gene has to be cloned in-frame with the expression vector promoter.

To get an indication of the species from which the metagenomic DNA originated, the inserts were searched for housekeeping or conserved genes that could provide some phylogenetic information. None of the inserts contained genes encoding 16S rRNA, a commonly used phylogenetic marker (WU *et al.* 2013). However, there is some evidence that genes encoding 16S rRNA undergo horizontal gene transfer (HGT), which complicates phylogenetic analyses and reduces confidence in the evolutionary information provided by this phylogenetic marker (HARRINGTON *et al.* 2012). Other informative genes were present on the STC fosmids. These included *recA* and *recG*, which are housekeeping genes that are less likely to undergo HGT and are thus more robust phylogenetic markers (WU *et al.* 2011; WU *et al.* 2013), various genes encoding t-RNA synthetases and transferases and two genes encoding LSU ribosomal proteins. When there is a limited set of genes available from which to infer phylogenetic relationships, alternative genes to 16S rRNA and *recA* have to be considered. The closest match of all encoded proteins across the STCs to a known sequence was the RecA protein from STC 13 that showed 76% amino acid identity to the

RecA sequence of *Fibrobacter succinogenes*. Despite the relatively high sequence identity (compared to encoded proteins form other STCs and their homologues) the sequences are not similar enough to assign STC 13 to a particular genus. In the majority of cases the lowest classification level to which the insert could be resolved was order. The phyla represented in the STC sequences were Deltaproteobacteria, Fibrobacteres, Cyanobacteria, Actinobacteria and Planctomyces. In a study by MAKHALANYANE *et al.* (2013), the authors investigated the bacterial diversity of Namib Desert soil. The dominant phyla identified were Actinobacteria and Proteobacteria, with Acidobacteria, Cyanobacteria, Bacteroidetes and Chloroflexi representing the lesser abundant phyla. The proteins encoded on metagenomic inserts sequenced in this study which were used as phylogenetic markers showed very low amino acid identity to known sequences, often with identity below 60%. This suggests that the fosmid sequences were derived from novel taxa, highlighting the untapped microbial communities present in the Namib Desert soil.

Chapter 5: Heterologous Expression and *in vivo* Testing of Putative Stress-tolerance Proteins

5.1 Molecular Cloning of Stress-related Genes and Bioinformatic Analysis of the Encoded Proteins

Six genes potentially responsible for conferring salt-tolerance to STC 1, 13, 90 and 92 were identified in Chapter 4. In this chapter, the proteins encoded by these genes were classified and characterised using various bioinformatic tools. The genes were sub-cloned into the expression vector, pET21a, over-expressed and tested for salt-tolerance activity in *E. coli* BL21(DE3) and Rosetta(DE3)pLysS strains.

The stress-related genes targeted for analysis are shown in Table 19. The nucleic acid sequence of each gene is provided in Appendix C. Table 20 shows percentage identity of each protein to the closest match in the non-redundant protein sequences on the NCBI website, as determined by BLASTp analysis.

Table 19. Stress-related genes sub-cloned from salt-tolerant clones.

Fosmid	Gene	Abbreviation	Size (bp)	Protein Size (aa)	Protein Size (kDa)
STC 1	ABC transporter substrate-binding domain	ABC SBD	1812	603	67.8
	RelA-SpoT-like hypothetical protein	RelA-SpoT	1248	415	48.2
	HD-hydrolase domain including promoter	HD-hydrolase	1199	328	36.8
STC 13	Cation export system protein	CESP	3114	1037	100.2
STC 90	Peptidase M29	M29	1131	376	41.5
STC 92	Na^+/H^+ antiporter	Na^+/H^+ antiporter	1638	545	59.5

Table 20. BLASTp analysis of putative stress-related proteins.

Protein	Highest Similarity Organism	% Identity to Closest Match	E-value
ABC SBD	*Bacteriovorax* sp. BAL6_X	43	$3e^{-168}$
RelA-SpoT	*B. marinus*	46	$4e^{-117}$
HD-hydrolase	*Bdellovibrio bacteriovorus*	27	$3e^{-37}$
CESP	*Hyalangium minutum*	59	0.0
M29	*Roseiflexus castenholzii*	50	$2e^{-118}$
Na^+/H^+ antiporter	*Conexibacter woesei*	57	$1e^{-175}$

5.1.1 ABC Transporter Substrate-binding Domain

PCR amplification of the gene encoding the ABC transporter substrate-binding domain from STC 1 fosmid DNA was successful (Figure 15). The gene was successfully cloned into the *Nde*I and *Hind*III sites of the pET21a vector, confirmed by enzymatic digestion (Figure 16) and sequencing.

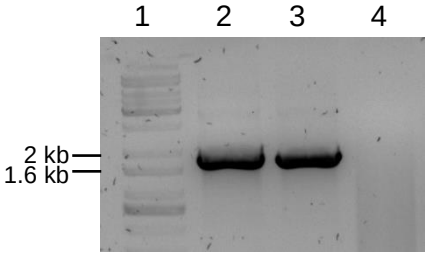


Figure 15. PCR amplification of the ABC SBD from STC 1 using iProof High-Fideliy DNA Polymerase (Bio-Rad). Lane 1: KAPA Universal DNA Ladder. Lane 2 – 3: ABC SBD amplicons. Lane 4: Negative control.

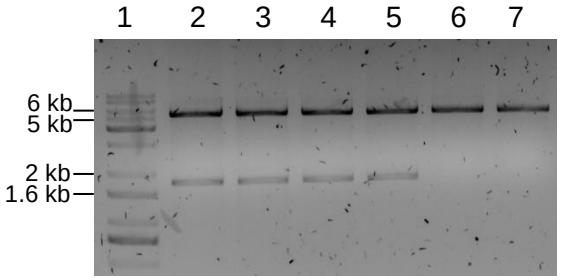


Figure 16. Restriction enzyme digestion of ABC SBD:pET21a clones with Fastdigest *Nde*I and *Hind*III. The vector backbone can be seen at 5.4 kb and the ABC SBD insert at 1.8 kb. Lane 1: KAPA Universal DNA Ladder. Lane 2 – 5: ABC SBD:pET21a clones. Lane 6 – 7: empty pET21a.

Substrate-binding domains (SBDs) are commonly found as part of uptake ABC-transporter systems and have been divided into six clusters (A – F) based on structure (BERNTSSON *et*

al. 2010). Cluster C SBDs, into which the ABC SBD identified in this study is classified, bind various substrates such as, di- and oligopeptides, arginine, cellobiose and nickel (BERNTSSON *et al.* 2010). Cluster C SBD proteins contain an extra domain, when compared to other clusters of SBDs, and are between 55 and 70 kDa in size, which is 10 – 40 kDa larger than SBDs in the other clusters (BERNTSSON *et al.* 2010).

In Gram-negative bacteria SBDs are located in the periplasmic space and are hydrophilic proteins (BIEMANS-OLDEHINKEL AND POOLMAN 2003). Secreted proteins generally contain signal peptides. Signal peptides are short, usually hydrophobic, stretches of amino acids that are usually cleaved by signal peptidases in the cellular membrane and the protein is released on the external side of the membrane, but some proteins may remain attached to the cellular membrane via these peptides (PUGSLEY 1993). As the STC 1 metagenomic DNA is grouped with Gram-negative bacteria, the signal peptide prediction for this group of bacteria was used on the SignalP server. This program provided a D-score cut-off of 0.570. This is the score used to distinguish signal peptides from non-signal peptides. The ABC SBD had a D-score of 0.562. A signal peptide was not predicted to be present (Figure 17). A putative binding protein (yliB) from *E. coli*, to which the ABC SBD had 25% amino acid identity with an E-value of $1.13e^{-29}$ on the TCDB (Transporter Classification Database), was predicted to contain a signal peptide (Figure 17). The ABC SBD should contain a signal peptide for localisation to the periplasmic space. A combination of database bias and the novelty of the ABC SBD sequence, due to its metagenomic origin, could be a possible explanation for the failure to predict a signal peptide.

The LipoP program predicted both signal peptidase (Sp) I and II cleavage sites (Figure 18). The SpII site is most likely the true signal peptide site, as this site corresponds to amino acid positions 24 – 25, which is the site at which the highest score was obtained from the SignalP prediction. However, lipid-anchored SBDs are found only in Gram-positive bacteria and Archaea (VAN DER HEIDE AND POOLMAN 2002). Furthermore, the Kyte and Doolittle hydrophobicity plot (Figure 19) shows that the majority of the ABC SBD is hydrophilic, as expected for SBDs. However, the N-terminal region of the protein is strongly hydrophobic compared to the rest of the protein, corresponding to the hydrophobic nature of signal peptides. The combination of the results in Figure 17, Figure 18 and Figure 19 indicate the probable presence of a signal peptide.

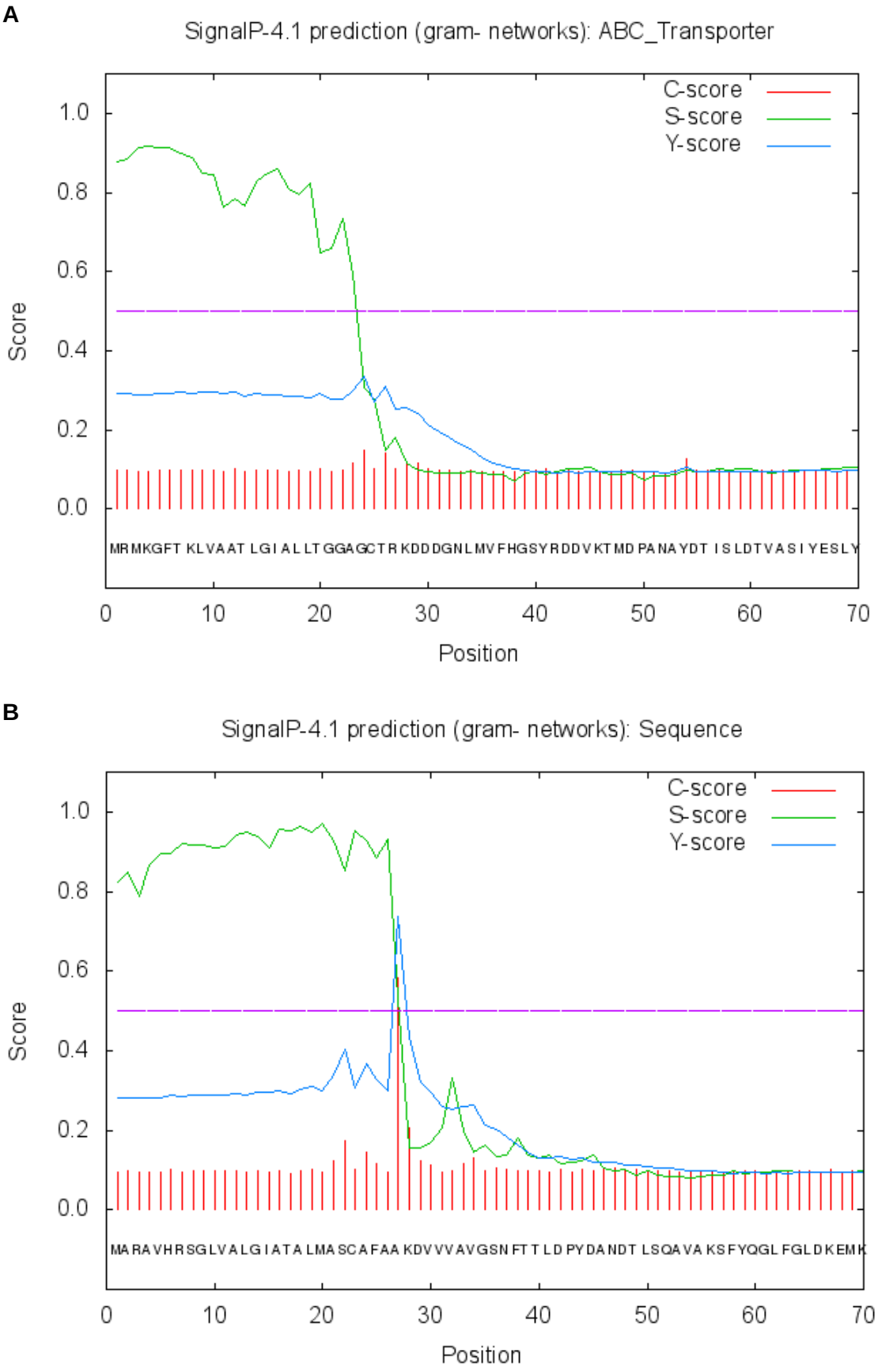


Figure 17. SignalP results for A) the ABC SBD and B) the *E. coli* SBD homologue, showing the absence of a signal peptide in A and the presence of a signal peptide in B. The steep gradient of the S-score (in combination with spikes in the C- and Y-scores) is indicative of the presence of a signal peptide.

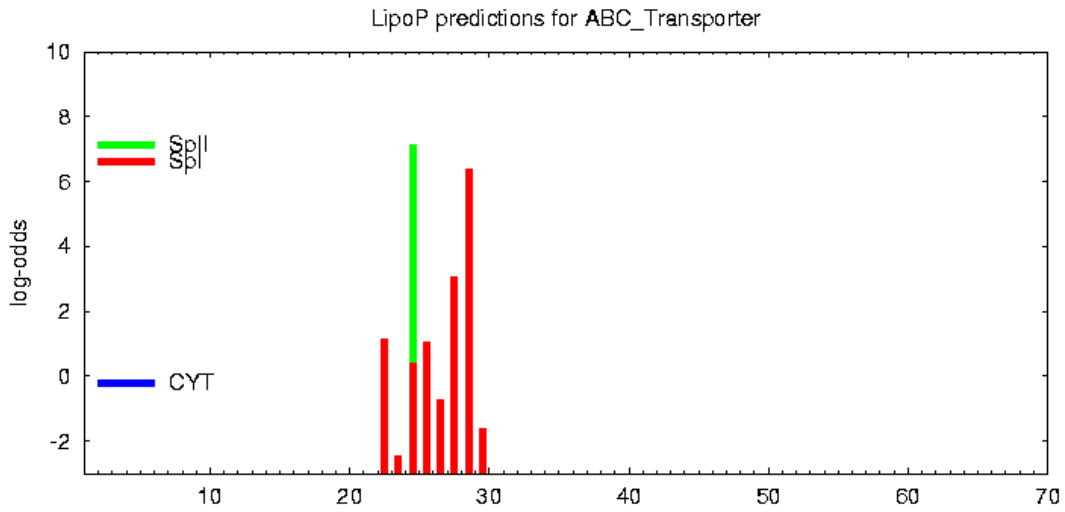


Figure 18. LipoP results for ABC SBD indicating the predicted signal peptidase (Sp) I and II cleavage sites.

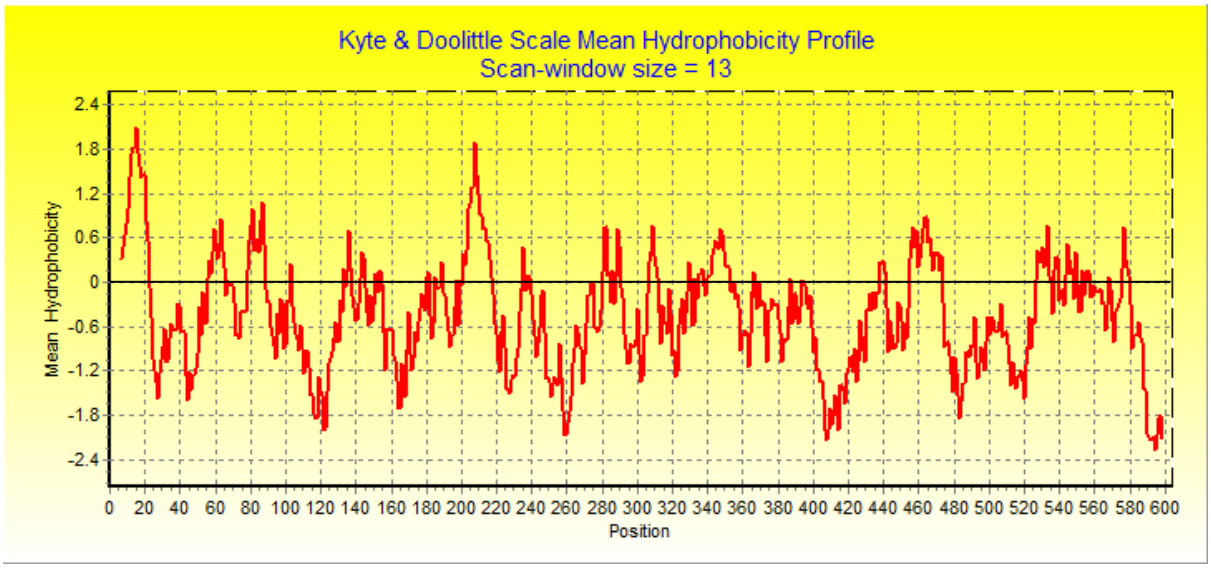


Figure 19. The Kyte and Doolittle scale was used to determine the hydrophobic character of the entire ABC SBD protein with a window size of 13. Regions with values above 0 are hydrophobic.

The ABC SBD gene contained 9.6% rare codons and four doublets and two triplets of consecutive rare codons (Table 21). Rare codons, particularly arginine codons AGA and AGG and codon clusters, can result in a decrease of recombinant protein production and quality (KANE 1995). The amount of tRNAs present in the cytoplasm is directly related to the codon abundance in the genome of the particular organism (KANE 1995). This results in stalling of the ribosome at rare codon positions during translation, causing in-frame skipping of codons or insertion of incorrect amino acids (KANE *et al.* 1992; KURLAND AND GALLANT 1996). This occurrence is exacerbated by consecutive rare codons and can result in premature termination of translation (SØRENSEN *et al.* 2003). The ABC SBD gene contained only two AGG codons, but contained 18 CCC proline codons, which is also a rare codon known to cause problems during heterologous protein translation (SØRENSEN AND MORTENSEN 2005).

Table 21. Rare codons present in the ABC SBD gene.

Amino Acid	Codon	Frequency
Arginine	CGA	3
	CGG	7
	AGG	2
	AGA	0
Glycine	GGA	3
	GGG	4
Isoleucine	AUA	0
Leucine	CUA	0
Proline	CCC	18
Threonine	ACG	21
Repeated and/or consecutive rare codons	ACG CCC	1
	CCC CGG	1
	CCC GGA	1
	ACG ACG	1
	GGA ACG GGG	1
	CCC GGG CCC	1

5.1.2 RelA-SpoT-like Hypothetical Protein and the HD-hydrolase Domain Protein

PCR amplification of both the RelA-SpoT-like hypothetical protein (using primers STC1-2F & STC1-2R for cloning into the *Nde*I and *Hind*III sites and primers STC1-2F & STC1-2.1R for cloning into the *Nde*I and *Eco*RI sites) and the HD-hydrolase genes from STC 1 fosmid DNA was successful (Figure 20). Initially, the RelA-SpoT gene was cloned individually into the *Nde*I and *Hind*III sites of the pET21a vector (Figure 21 A) and tested *in vivo* for salt-tolerance activity (see section 5.2). Thereafter, the RelA-SpoT gene was cloned into the *Nde*I and *Eco*RI sites and the HD-hydrolase (together with its native promoter) was cloned downstream of the RelA-SpoT gene into the *Eco*RI and *Sal*I sites. This was achieved by first cloning the RelA-SpoT gene into pET21a (Figure 21 B) and then cloning the HD-hydrolase into the recombinant RelA-SpoT:pET21a vector (Figure 22). Cloning was confirmed via colony PCR and sequencing.

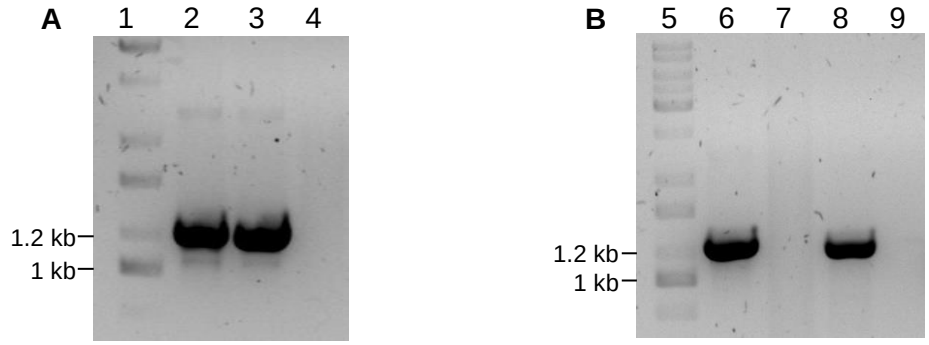


Figure 20. PCR amplification using iProof polymerase of RelA-SpoT using primers STC1-2F & STC1-2R (A) and primers STC1-2F and STC1-2.1R (B). The HD-hydrolase including its native promoter was also amplified successfully (B). Lanes 1 and 5: KAPA Universal DNA Ladder. Lanes 2, 3 and 6: RelA-SpoT amplicons. Lanes 4, 7 and 9: Negative controls. Lane 8: HD-hydrolase, including the native promoter, amplicon.

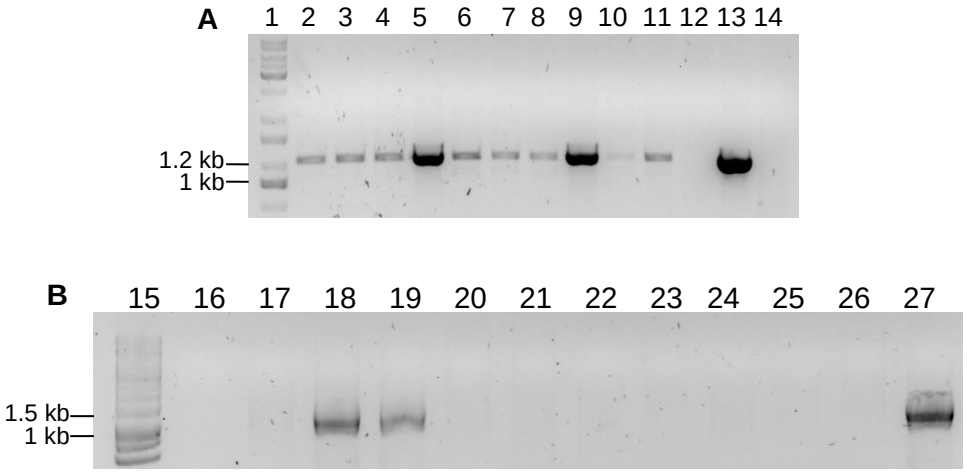


Figure 21. Colony PCR screening using DreamTaq DNA Polymerase (Thermo Fisher Scientific) for RelA-SpoT:pET21a recombinant plasmids transformed into *E. coli* JM109 cells. A) RelA-Spot cloned into *Nde*I and *Hind*III sites. Lane 1: KAPA Universal DNA Ladder. Lanes 2 – 12: JM109 transformants. Lane 13: Positive control (STC1 DNA). Lane 14: Negative control. B) RelA-SpoT cloned into *Nde*I and *Eco*RI sites. Lane 15: GeneRuler 1 kb DNA Ladder (Fermentas). Lane 16: Negative control. Lanes 17 – 26: JM109 transformants. Lane 27: Positive control (STC 1 DNA).

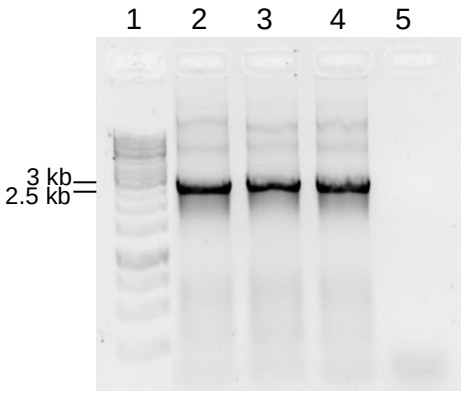


Figure 22. PCR amplification of RelA-SpoT and HD-hydrolase insert in pET21a. Primers used: STC1-2F and STC1-3R. The expected size of the insert is 2447 bp. Lane 1: GeneRuler 1 kb DNA Ladder. Lanes 2 – 4: Amplicons of RelA-SpoT:HD-hydrolase. Lane 5: Negative control.

RelA-SpoT proteins are responsible for the synthesis of a signalling alarmone, guanosine 3‘,5’-bis(diphosphate) ((p)ppGpp), which regulates the stringent response as a result of amino acid starvation in *E. coli* (BRAEKEN *et al.* 2006). Initially pppGpp is produced, but it rapidly converted into ppGpp, which then binds RNA polymerase and alters transcription to stress and starvation-related genes (MAGNUSSON *et al.* 2005). It is known that ppGpp levels increase in response to hypertonic shock induced by sodium chloride (HARSHMAN AND YAMAZAKI 1972). In *E. coli* and other Gram-negative bacteria two separate, but similar proteins, RelA and SpoT, are responsible for regulating the level of ppGpp in the cell in response to amino acid starvation and other stresses (MAGNUSSON *et al.* 2005). RelA is a synthetase associated with ribosomes and is involved mainly in the response to amino acid starvation (WENDRICH *et al.* 2002). SpoT contains both a synthetase and hydrolase domain and it is thought that this pathway to controlling ppGpp levels in the cell is activated by other stresses (GENTRY AND CASHEL 1996; BRAEKEN *et al.* 2006).

The conserved domain database (CDD) on the NCBI website did not predict the RelA-SpoT domain in the RelA-SpoT-like hypothetical protein identified here. However, the amino acid sequence showed between 42 – 46% identity to hypothetical proteins from five different *Bacteriovorax* species, all of which were predicted to contain the RelA-SpoT domain by the CDD search. Analysis on the SMART server of the amino acid sequences of the protein of interest, the top BLASTp match and the *E. coli* RelA protein showed that the RelA-SpoT domain was present in the protein of interest (Figure 23). What is also clear from Figure 23 is that both the *B. marinus* and the protein of interest are substantially smaller than the *E. coli* RelA protein and lack the HD, TGS and ACT domains. However, LEMOS *et al.* (2007) showed that these domains are not required for ppGpp synthetase activity by identifying two genes in *Streptococcus mutans* encoding proteins containing only the RelA-SpoT domain

capable of synthesising ppGpp in *S. mutans* and in *E. coli*. In fact, the HD domain, a hydrolase domain, is inactive in the RelA protein of *E. coli* (POTRYKUS AND CASHEL 2008). Instead, the SpoT protein in *E. coli* carries out hydrolysis of ppGpp through a functional HD domain and also contains a weakly active synthetase domain (GENTRY AND CASHEL 1996). Furthermore, it was shown that one of these shortened proteins, RelP, was responsible for synthesising ppGpp during exponential growth (LEMOS *et al.* 2007). The amino acid sequences of the top five BLASTp matches and the translated sequence of the *E. coli* RelA gene were aligned to the amino acid sequence of the protein of interest using Clustal Omega server and showed that the region where the RelA-SpoT domain is predicted is conserved across these species (Appendix C: Figure C1).

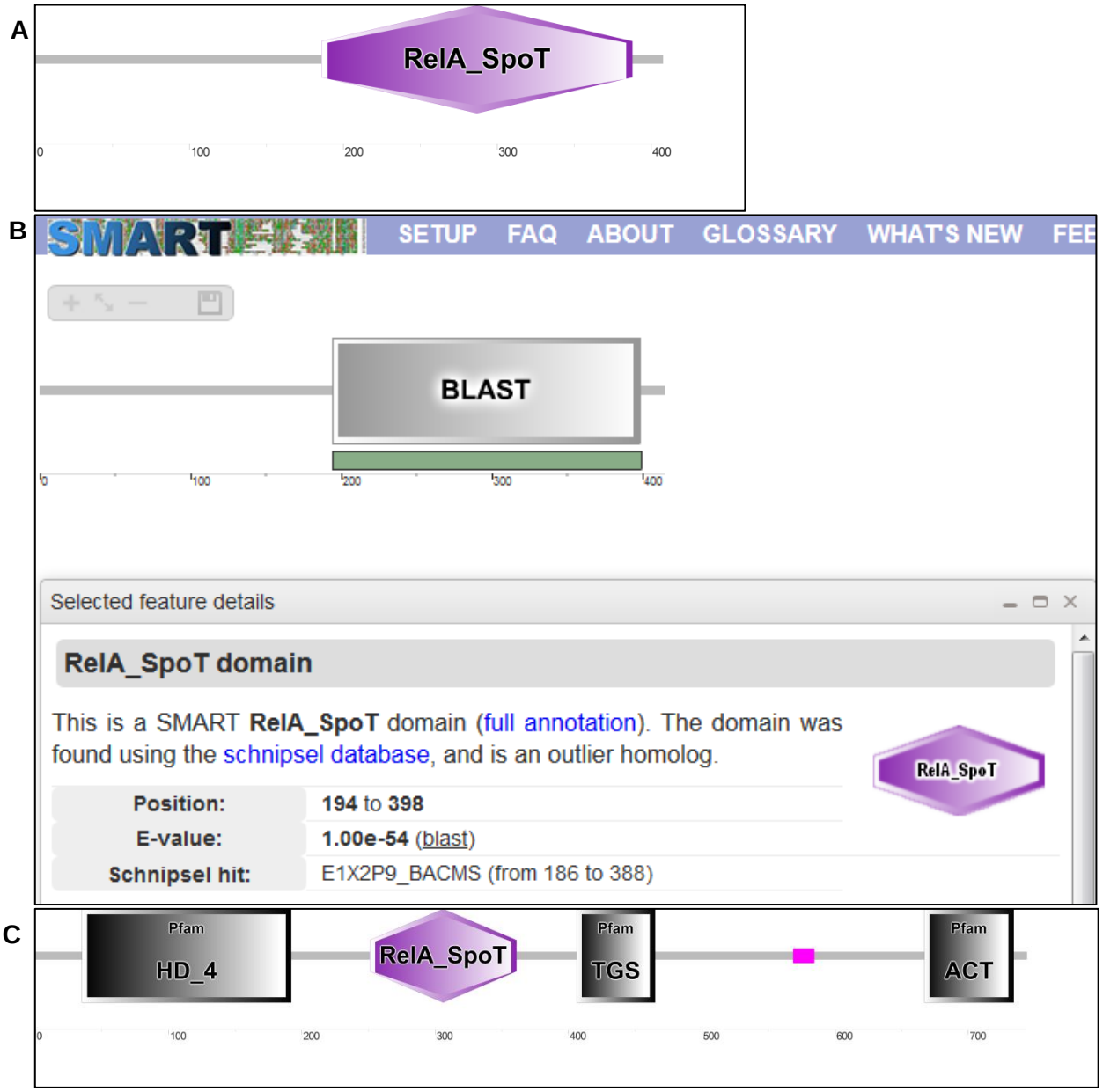


Figure 23. SMART domain prediction of the RelA-SpoT domain A) in the top BLASTp hit, *Bacteriovorax marinus* hypothetical protein, B) in the protein of interest and C) in the *E. coli* RelA protein.

HD-hydrolases are a group of divalent metal ion dependent phosphohydrolases with conserved histidine-apartate (H-D) residues in the catalytic site (ARAVIND AND KOONIN 1998) (Figure 24). This domain is present in proteins such as dGTPases, (p)ppGpp hydrolases and cyclic-nucleotide phosphodiesterases, all of which act on substrates containing phosphodiester bonds (ARAVIND AND KOONIN 1998). The HD-hydrolase domain protein identified in this study contains the conserved HD residues and conforms to the consensus sequence on either side of these residues (Figure 24 and Figure C2). Interestingly, the RelA protein of *E. coli* contains a phenylalanine and a proline residue instead of the H and D residues, rendering the hydrolase domain inactive (ARAVIND AND KOONIN 1998). Both the RelA-SpoT and HD-hydrolase domains are required for maintaining an optimal level of ppGpp in the cell, as high levels of ppGpp can inhibit cell growth (LEMOS *et al.* 2007).

RelA-SpoT homologues from the halophilic plant *Suaeda japonica* and *Arabidopsis thaliana* have been shown to confer salt-tolerance to *Saccharomyces cerevisiae* and *E. coli* (YAMADA *et al.* 2003; OCHI *et al.* 2012). The general stress-response elicited by the RelA-SpoT product, ppGpp, and the ability of the plant homologues to confer salt-tolerance to other species are good indicators that the RelA-SpoT-hypothetical protein and the HD-hydrolase, or a combination of these proteins, could be responsible for the observed salt-tolerance of STC 1.

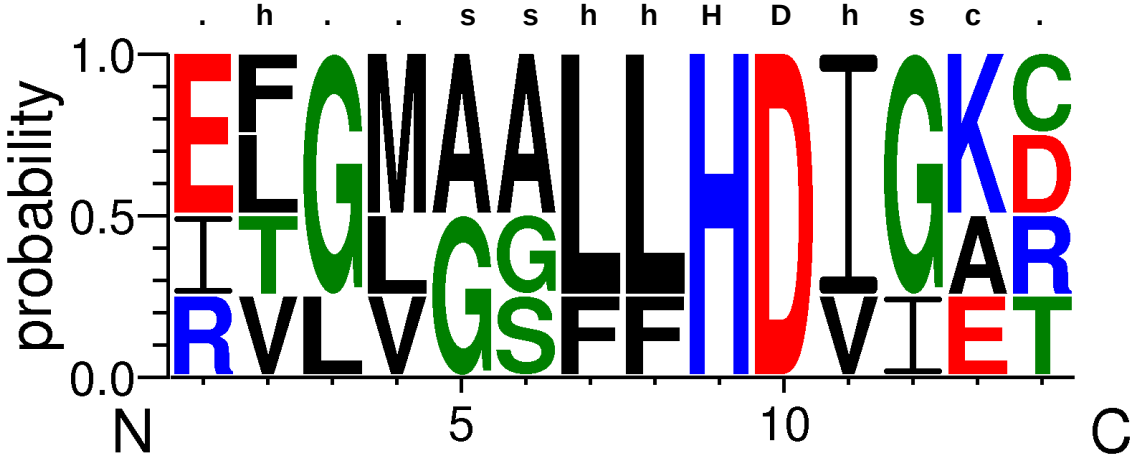


Figure 24. The conserved HD domain of HD-hydrolase proteins. The sequence logo was constructed using a multiple sequence alignment of the amino acid sequences of the protein of interest, the closest two BLASTp matches (*Bdellovibrio bacteriovorus* and *Hydrogenedentes* spp.) and the *E. coli* SpoT protein sequence. Residues are coloured according to chemistry: green – polar, blue – basic, red – acidic, black – hydrophobic. The consensus sequence is shown above the image, where: h – hydrophobic residues (A, C, F, I, L, M, V, W and Y), s – small residues (A, C, S, T, D, N, V, G and P), c – charged residues (D, E, R, K and H), H – histidine, D – aspartate.

The RelA-SpoT gene contained 9.9% rare codons and one doublet of consecutive rare codons (Table 22). The HD-hydrolase gene contained 9.1% rare codons and two doublets of consecutive rare codons (Table 23).

Table 22. Rare codons present in the RelA-SpoT gene.

Amino Acid	Codon	Frequency
Arginine	CGA	5
	CGG	8
	AGG	5
	AGA	0
Glycine	GGA	3
	GGG	3
Isoleucine	AUA	0
Leucine	CUA	0
Proline	CCC	6
Threonine	ACG	11
Repeated and/or consecutive rare codons	GGG CGA	1

Table 23. Rare codons present in the HD-hydrolase gene.

Amino Acid	Codon	Frequency
Arginine	CGA	1
	CGG	2
	AGG	0
	AGA	0
Glycine	GGA	1
	GGG	4
Isoleucine	AUA	0
Leucine	CUA	1
Proline	CCC	6
Threonine	ACG	15
Repeated and/or consecutive rare codons	CGG ACG CCC GGA	1 1

5.1.3 Cation Export System Protein

PCR amplification of the CESP gene was successful, but a number of non-specific amplicons were observed. Thus, the band of interest was excised and purified from the agarose gel (Figure 25). The gene was successfully cloned into the *Nde*I and *Hind*III sites of the pET21a vector, as confirmed by enzymatic digestion (not shown) and sequencing.

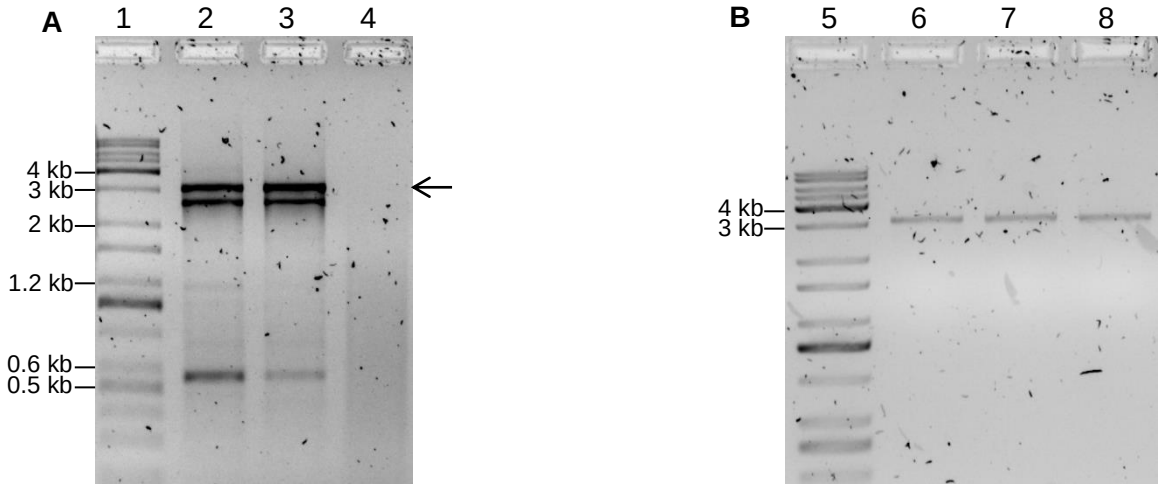


Figure 25. A) PCR amplification of CESP from STC 13 using iProof Polymerase. Lane 1: KAPA Universal DNA Ladder. Lanes 2 and 3: CESP amplicons. Lane 4: negative control. The arrow indicates the band of interest at the expected size of 3.1 kb. B) CESP amplicons after purification from the agarose gel. Lane 5: KAPA Universal DNA Ladder. Lanes 6 – 8: CESP amplicons.

The TCDB classified CESP as part of the resistance-nodulation-cell division (RND) superfamily. Members of this superfamily are integral membrane proteins that catalyse the export of a wide range of substrates across the cellular membrane, most probably via a proton (H^+) antiporter mechanism (TSENG *et al.* 1999). Members of this superfamily can be divided into eight phylogenetic families, of which families one, two and three belong to Gram-negative bacteria. The heavy metal exporters (HMEs) belong to family one (SAIER 2000; NIES 2003). Figure 26 shows the phylogenetic relationship of selected RND proteins from each family (1 – 8). RND proteins are found in bacteria, archaea and eukaryotes and their wide range of substrates, which include heavy metal ions, multiple drugs, oligosaccharides, organic solvents, fatty acids, phospholipids and cholesterol, are a possible reflection of their widespread occurrence through the Domains of life (TSENG *et al.* 1999).

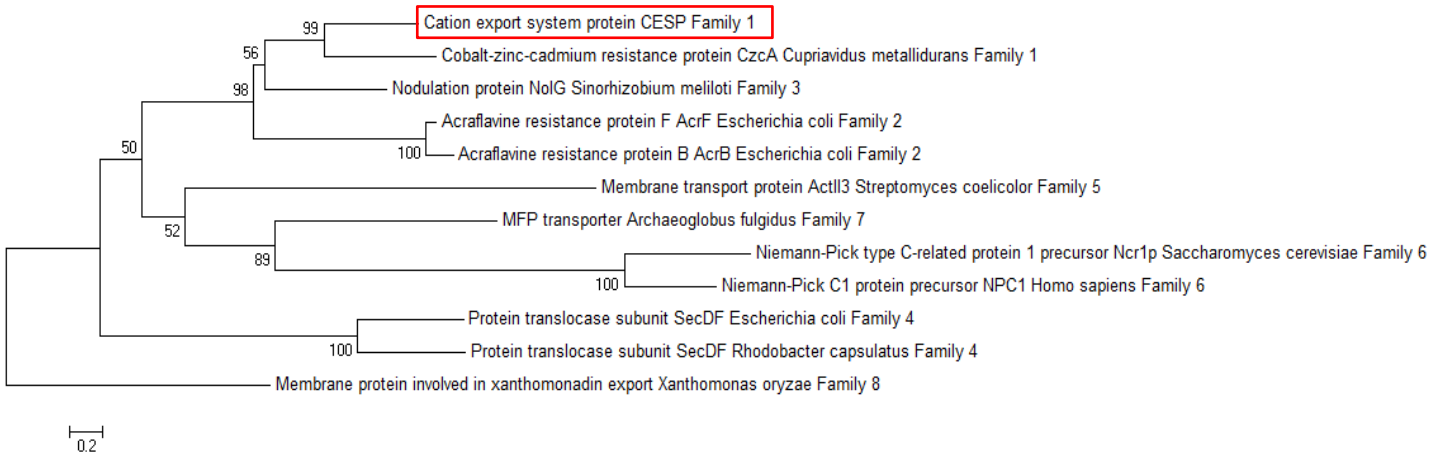


Figure 26. Neighbour-joining phylogeny of CESP and selected RND proteins from each of the eight families. The numbers at the nodes indicate the bootstrap percentages based on 1000 replicates (values below 50 were excluded). The bar represents 0.2 nucleotide substitutions per nucleotide position. The branch representing the RND protein from family 8 was used to root the tree. The tree was constructed using MEGA 5.2 software.

CESP grouped with these HMEs, as the closest BLASTp matches in the NCBI database and the TCDB were CzcA (cobalt-zinc-cadmium resistance) proteins, belonging to *Hyalangium minutum* and *Caulobacter crescentus*, respectively. CESP showed much higher amino acid identity (59%) to CzcA from *H. minutum* than to the same protein of *C. crescentus* (38%). HME-RND proteins can further be classified into sub-groups based on the signature motif present in the fourth transmembrane helix, which is used to predict the substrate of the exporter. These sub-groups are HME1, 2, 3a, 3b, 4 and 5 (NIES 2003). The signature motif present in CESP matched the sub-group HME3b consensus sequence at all residues, except one (NIES 2003). Members of the HME3b sub-group are monovalent cation exporters and could thus be involved in sodium (Na^+) stress response (NIES 2013). Figure 27 shows the alignment of various HME-RND proteins and the signature motifs of divalent (HME1 and 2) and monovalent (HME3b) cation exporters. Table 24, which was adapted from NIES (2003) and augmented with information from NIES *et al.* (2006), shows the signature motifs of the HME sub-groups.

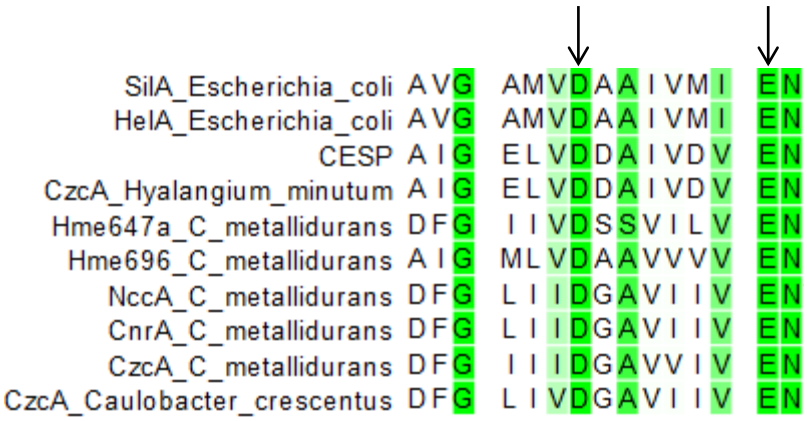


Figure 27. Multiple sequence alignment using Clustal Omega of the amino acid sequences of various HME-RND proteins and CESP. The region shown is the signature motif in the fourth transmembrane helix of the proteins. C_metallidurans: *Cupriavidus metallidurans*. Shading indicates 70% conservation, or greater, at a particular position. The arrows indicate conserved aspartate and glutamate residues that are essential for metal resistance (GOLDBERG *et al.* 1999). This figure should be interpreted with the aid of Table 24.

Table 24. Consensus sequences of HME signature motifs (adapted from NIES (2003)).

HME Sub-group	Signature Motif*	Predicted Substrate	Examples
HME1	DFG-DGA-VEN	Zn^{2+}, Co^{2+}, Cd^{2+}	CzcA, CztA
HME2	DFG-DGA-VEN	Ni^{2+}, Co^{2+}	CnrA, NccA
HME3a	GFG-D(G,S,A)(S,A)-(V,M)EN	Divalent cations	Hme647a
HME3b	(A,g)(I,L)G-D(G,A,s)A-VEN	Monovalent cations	Hme696
HME4	A(I,V)G-DA(A,s)-(V,I)(E,d)N	Cu^{+}, Ag^{+}	SilA, CusA
HME5	AIG-DDX-(M,V)EN	Ni^{2+}	All7631

*A dash indicates any three, mostly hydrophobic, amino acids. Parentheses indicate alternative amino acids for the same position. Residues in lower-case indicate rare amino acid choices and X indicates any amino acid.

CESP and its closest homologue, CzcA from *Hyalangium minutum*, differ from the consensus sequences of HME3b at position eight, where either an alanine (A), glycine (G) or serine (S) is expected, but there is an aspartate (D) instead. None of the other sub-families have a D occupying this position. This amino acid substitution of a small, neutral residue with an acidic residue could have a significant effect on the function or substrate specificity of the protein due to the different chemistry of the residues. The substrate specificity of HME3b RND proteins has not been fully investigated and it is not uncommon for RND metal transporters to non-specifically export various substrates (CONROY *et al.* 2010). Thus, CESP was considered to be a good candidate for conferring tolerance to elevated NaCl levels, as Na^{+} ions could be extruded from the cytoplasm by this protein.

CESP has 10 transmembrane helices predicted (Figure 28 A). The Kyte and Doolittle hydrophobicity plot correlates with the positions of the transmembrane regions, which are hydrophobic, and the periplasmic regions, which are hydrophilic (Figure 28 B). The CESP gene contained 9.1% rare codons and nine doublets of consecutive rare codons (Table 25).

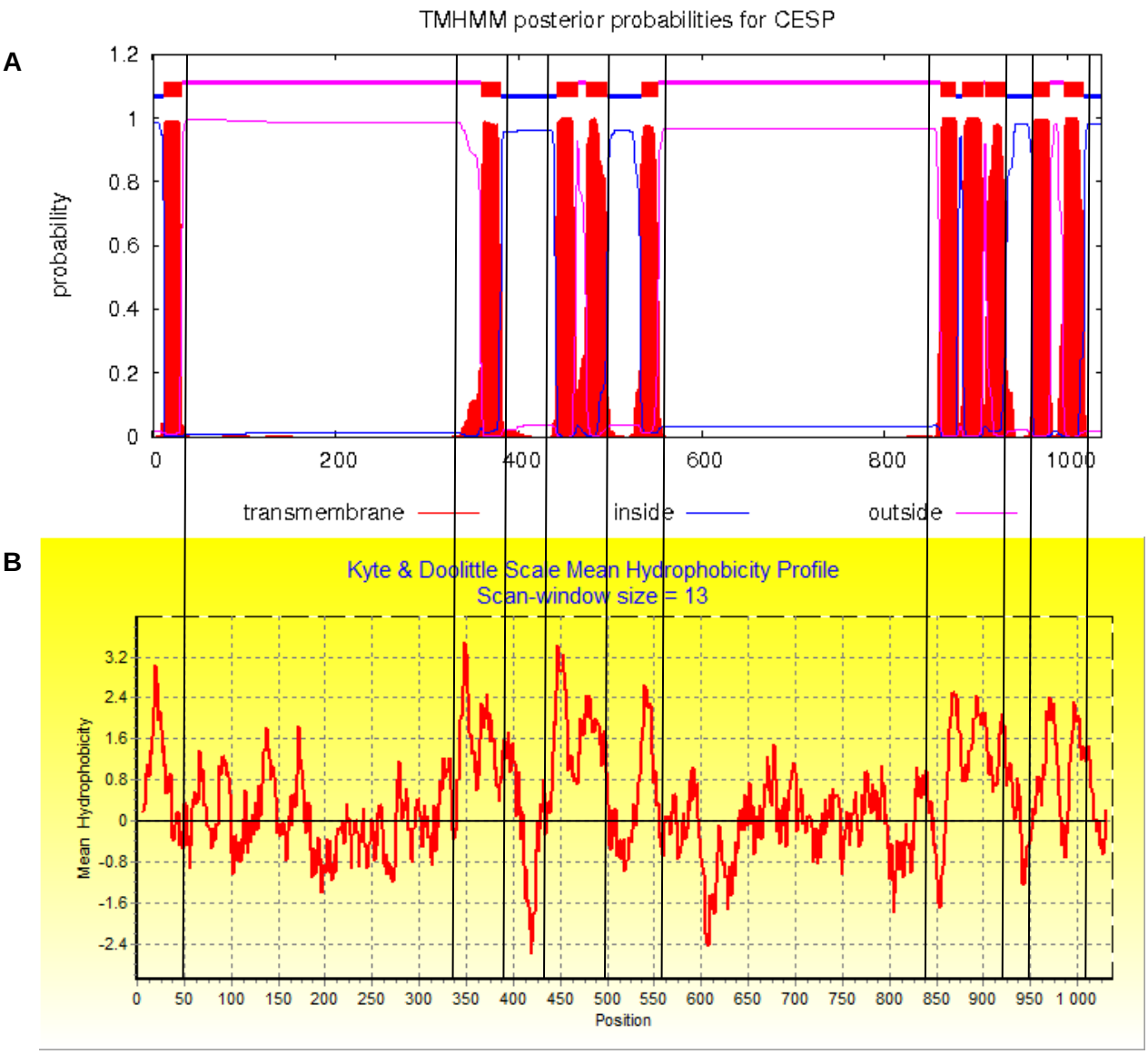


Figure 28. A) TMHMM results showing the predicted transmembrane helices (red) of CESP and the perisplasmic (pink line) and cytoplasmic regions (blue line) of the protein. B) The Kyte and Doolittle hydrophobicity plot, with a window size of 13, shows the hydrophobic regions of the protein. Regions above 0 are hydrophobic. The vertical black lines illustrate how the hydrophobic regions of CESP (B) correspond to the location of the transmembrane helices (A).

Table 25. Rare codons present in the CESP gene.

Amino Acid	Codon	Frequency
Arginine	CGA	0
	CGG	20
	AGG	4
	AGA	0
Glycine	GGA	4
	GGG	23
Isoleucine	AUA	0
Leucine	CUA	0
Proline	CCC	37
Threonine	ACG	6
Repeated and/or consecutive rare codons	GGG CCC	1
	CCC CCC	2
	CGG CGG	1
	CGG CCC	1
	GGA CGG	1
	GGG CGG	1
	CGG GGG	1
	ACG GGG	1

5.1.4 Peptidase M29

The peptidase M29 gene was successfully amplified via PCR from STC 90 (Figure 29) and cloned into the *Nde*I and *Hind*III sites of the pET21a vector, as confirmed via colony PCR (data not shown) and sequencing. The M29 gene contained a GTG start codon, an alternative to the common ATG start codon. The *Nde*I site on the pET21a vector contains an ATG codon. Genes targeted for over-expression, under the control of the IPTG-inducible promoter, are cloned into the *Nde*I site as it is the ideal distance downstream of the T7 promoter and the ribosome-binding site. This allows for optimal over-expression of cloned genes. In order to achieve this with the M29 gene, site-directed mutagenesis was performed by inserting an A at the position corresponding to the first G in the GTG start codon in the forward primer used to amplify the gene (STC90-1F), which results in the substitution of the initial methionine for a valine. The final PCR amplicons of the gene thus contained an ATG start codon and could be cloned into the *Nde*I site.

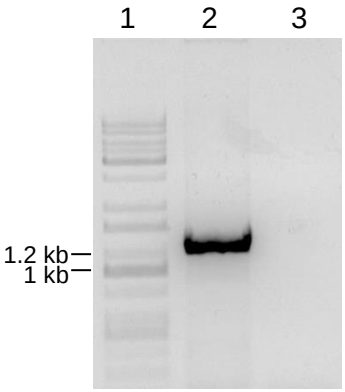


Figure 29. PCR amplification of the M29 gene from STC 90 using iProof Polymerase. Lane 1: KAPA Universal DNA Ladder. Lane 2: M29 amplicon. Lane 3: Negative control.

Members of the peptidase M29 family are thermophilic metalloproteases (also known as the T (thermophilic) family) and are aminopeptidases, which means they cleave the amino acids from the N-terminus of proteins (MOTOSHIMA *et al.* 1997; BERTIN *et al.* 2005). The limited data on M29 family members shows that these peptidases are active in a pH range of 7 to 9 and temperatures between 55°C and 70°C (FERNANDEZ-ESPLA AND RUL 1999; BERTIN *et al.* 2005). The enzymes require a metal ion, Zn^{2+}, for activity and are inhibited by bestatin (BURLEY *et al.* 1991; BERTIN *et al.* 2005). Some bacterial aminopeptidases are secreted (GONZALES AND ROBERT-BAUDOUY 1996), but M29 was not predicted to contain a signal peptide by either SignalP or LipoP (data not shown) and is thus expected to be maintained in the cytoplasm.

The C-terminus of these enzymes is highly conserved and contains the putative Zn^{2+} binding domain as well as the bestatin ligation residues (Figure 30). The peptidase M29 identified here does contain the Zn^{2+} binding domain, but contains an isoleucine, instead of a threonine in the bestatin ligation region. The two closest BLASTp matches on the NCBI non-redundant protein sequences database, *Roseiflexus castenholzii* and *Thermobrachium celere*, contain an isoleucine and methionine at this position, respectively. Furthermore, these three proteins are 40 to 50 amino acids shorter than those M29 family members described thus far (BERTIN *et al.* 2005). The M29 gene contained 9.5% rare codons with four doublets and one triplet of consecutive rare codons (Table 26).

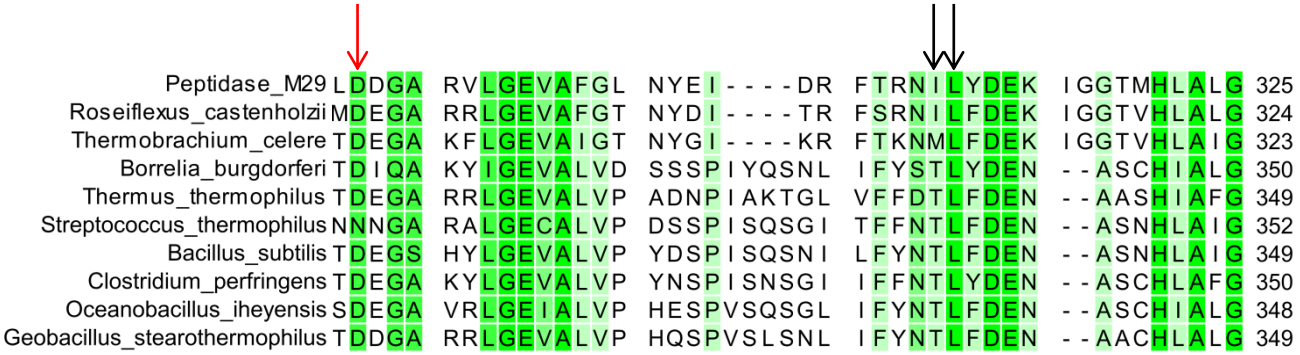


Figure 30. Multiple sequence alignment using Clustal Omega of the amino acid sequences of members of the M29 superfamily of aminopeptidases. A highly conserved portion of the C-terminal end of the proteins is shown. The red arrow indicates the putative Zn^{2+} binding residue. The black arrows indicate the putative bestatin ligation residues. Shading indicates 70% conservation, or greater, at a particular position.

Table 26. Rare codons present in the M29 gene.

Amino Acid	Codon	Frequency
Arginine	CGA	3
	CGG	15
	AGG	0
	AGA	0
Glycine	GGA	1
	GGG	6
Isoleucine	AUA	0
Leucine	CUA	0
Proline	CCC	2
Threonine	ACG	10
Repeated and/or consecutive rare codons	GGG CGG	1
	CGA CGG	2
	GGG CCC	1
	ACG CGG CGA	1

5.1.5 Na^+/H^+ Antiporter

The Na^+/H^+ antiporter was successfully amplified via PCR (Figure 31) and cloned into the *Nde*I and *Hind*III sites of the pET21a vector, as confirmed via colony PCR (data not shown) and sequencing. This gene also contained a GTG start codon and thus the same site-directed mutagenesis approach was used as with the M29 gene to change the codon to an ATG codon for cloning into the *Nde*I site.

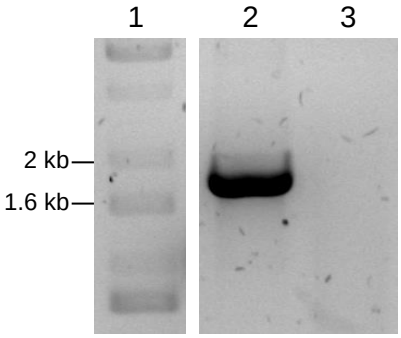


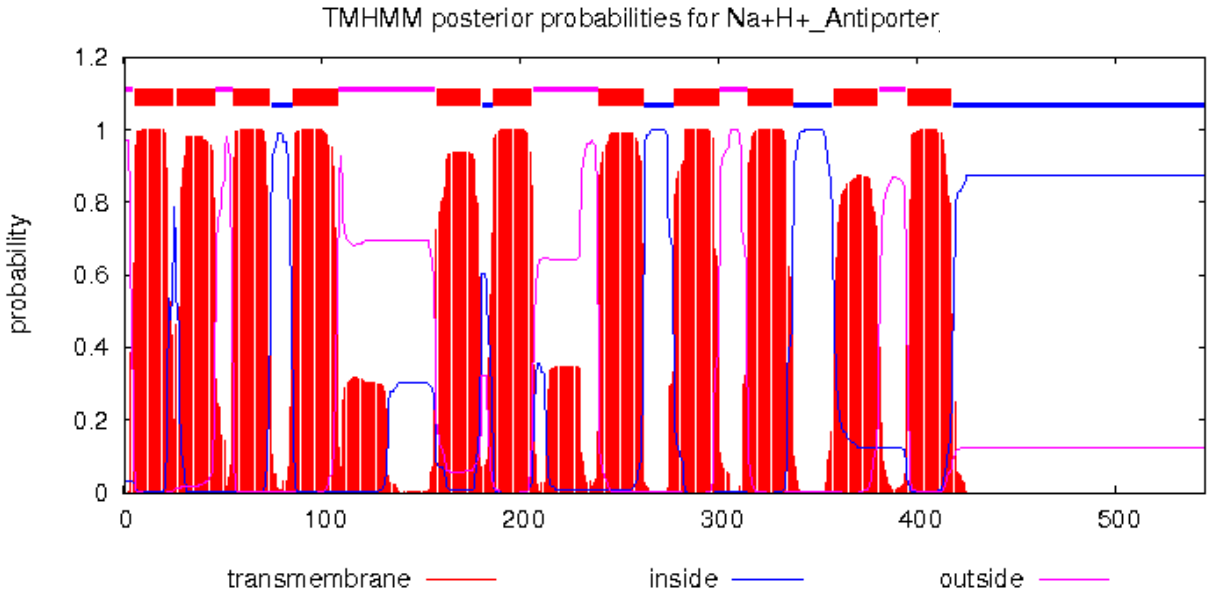
Figure 31. PCR amplification of the Na^+/H^+ gene from STC 92 using iProof Polymerase. Lane 1: KAPA Universal DNA Ladder. Lane 2: Na^+/H^+ amplicon. Lane 3: Negative control.

Na^+/H^+ antiporters occur ubiquitously across the three main domains of life; bacteria, archaea and eukaryotes (SAIER 2000), and are involved in pH and Na^+ homeostasis of the cell (PADAN *et al.* 2001). The Na^+/H^+ antiporter identified in this study was classified as a NhaP-type antiporter, of the family CPA1 (SAIER 2000) in both the NCBI non-redundant protein sequences database and the TCDB. CPA1 anitporters typically contain 10 to 12 transmembrane helices and substrates include Na^+/H^+, Na^+ and/-or K^+/H^+ (SAIER 2000). However, the antiporter identified in this study was predicted to contain 13 transmembrane helices (TMHs) (Figure 32 A), although two of these regions were predicted with much lower probability than the other 11 TMHs. The Kyte and Doolittle hydrophobicity plot (Figure 32 B) showed that these two regions (TMH V and VIII) were hydrophobic and were thus most probably situated in the cytoplasmic membrane. The amino acid sequence of the Na^+/H^+ antiporter was aligned to other members of the NhaP family, which included the type NhaP protein from *Pseudomonas aeruginosa,* the closest match on the TCDB *Aphanothece halophytica* and the NhaP sequence from the cyanobacterium *Synechocystis* sp. PCC 6803 (Figure 33). The aspartate residue, A^{138}, in NhaP of *Synechocystis* has been shown to be essential in transporter activity (HAMADA *et al.* 2001) and is conserved in the Na^+/H^+ antiporter identified on STC 92. The long C-terminal tail is also important in the maintenance of the transporter activity of NhaP transporters (Figure 32 A) (HAMADA *et al.* 2001; WADITEE *et al.* 2001).

Various Na^+/H^+ antiporters have been cloned from organisms adapted to hypersaline environments (halophilic/ -tolerant and non-halophilic organisms) and shown to confer combined salt- and alkali-tolerance to *E. coli* KNabc cells, which lack three major Na^+/H^+ antiporters (NhaA, NhaB and ChaA) and are thus more sensitive to high salt concentrations (YANG *et al.* 2006; WANG *et al.* 2013; ZHANG *et al.* 2014). It has also been shown that bacterial and plant Na^+/H^+ antiporters confer salt-tolerance and combined salt- and alkali-tolerance to *Arabidopsis thaliana* (SHI *et al.* 2003; ZHONG *et al.* 2012). The cation exporter genes are also present in multiple copies in halophilic organisms and are considered to be

important in the survival of these organisms in hypersaline, alkaline environments (VENTOSA *et al.* 1998). Thus, the Na^+/H^+ antiporter was considered to be the most probable cause for the salt-tolerance observed for STC 92. The Na^+/H^+ antiporter gene contained 13% rare codons, with seven doublets and one triplet of consecutive rare codons (Table 27).

A



B

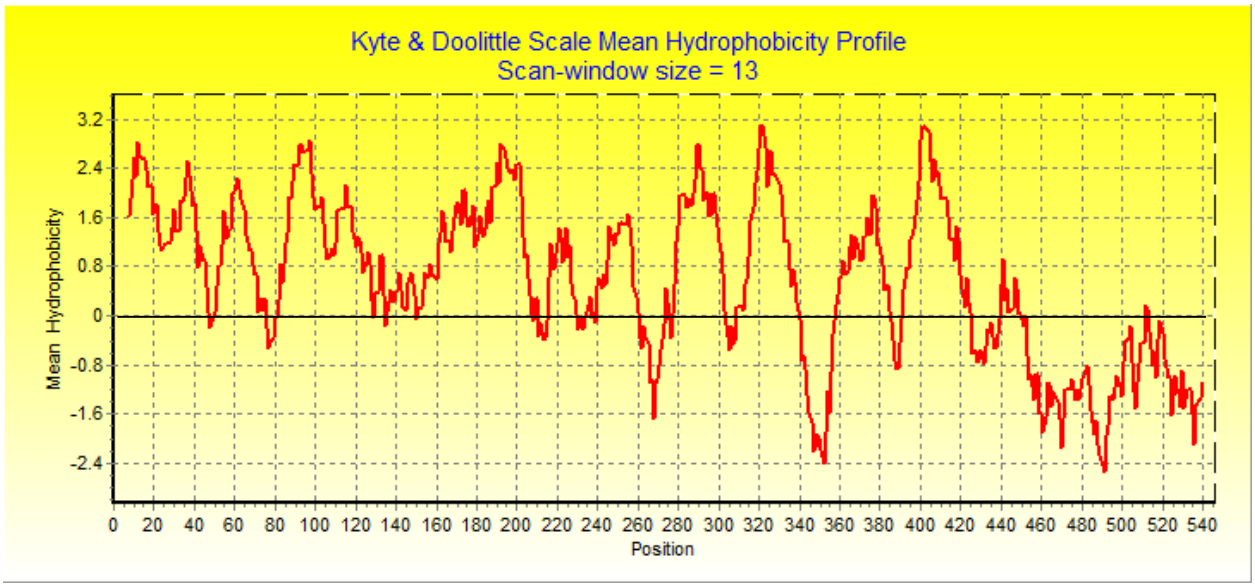


Figure 32. A) TMHMM results showing the predicted transmembrane helices (red) of Na^+/H^+ antiporter and the perisplasmic (pink line) and cytoplasmic regions (blue line) of the protein. B) The Kyte and Doolittle hydrophobicity plot, with a window size of 13, shows the hydrophobic regions of the protein. Regions above 0 are hydrophobic.

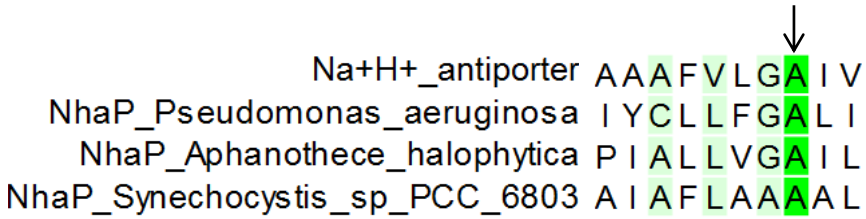


Figure 33. Multiple sequence alignment using Clustal Omega of the amino acid sequences of members of the NhaP group of Na^+/H^+ antiporters. The region containing the conserved aspartate residue is shown. The residue is indicated with an arrow. Shading indicates 75% conservation, or greater, at a particular position.

Table 27. Rare codons present in the Na^+/H^+ antiporter gene.

Amino Acid	Codon	Frequency
Arginine	CGA	2
	CGG	14
	AGG	5
	AGA	0
Glycine	GGA	4
	GGG	19
Isoleucine	AUA	5
Leucine	CUA	3
Proline	CCC	11
Threonine	ACG	8
Repeated and/or consecutive rare codons	AUA CCC	1
	AUA GGG	1
	GGG ACG	1
	CCC CCC	1
	CUA CCC	1
	ACG ACG	1
	CCC GGG	1
	CGG CGG CGG	1

5.2 *In vivo* Testing of Putative Stress-tolerance Proteins

E. coli BL21(DE3) and Rosetta(DE3)pLysS transformants used for *in vivo* testing of stress-tolerance are listed in Table 28.

Table 28. *E. coli* BL21(DE) and Rosetta(DE3)pLysS transformants used to test stress-tolerance.

***E. coli* Transformant**	**Stress Conditions Tested**
BL21:RelA-SpoT	5% NaCl
Rosetta:RelA-SpoT	5% NaCl
BL21:ABC SBD	5% NaCl
Rosetta:ABC SBD	5% NaCl
BL21:CESP	5% NaCl
Rosetta:CESP	5% NaCl
Rosetta:pET21a	5% NaCl
BL21:RelA-SpoT:HD-hydrolase	5% NaCl, 3.5% NaCl
BL21:M29	5% NaCl, 3.5% NaCl, 46°C
BL21:Na^+/H^+	5% NaCl, 3.5% NaCl, pH 9, 5% NaCl-pH 9
BL21:pET21a	All

5.2.1 Heterologous Protein Expression

Expression of the stress-related proteins was confirmed via SDS-PAGE analysis during all stress-tolerance experiments (Figure 34 and Figure 35). Over-expression of RelA-SpoT, ABC SBD and M29 was readily detected after IPTG induction in all expression strains as the visualisation of a large band in the SDS-PAGE gel at the expected size of the protein (Figure 34 and Figure 35). Expression of HD-hydrolase could not be detected visually. This gene was cloned with its native promoter and therefore was not over-expressed, as it was not under the control of the IPTG-inducible promoter. The expression of HD-hydrolase would thus be dependent on the ability of the *E. coli* host to recognise the native promoter. If this protein was indeed responsible for conferring salt-tolerance, together with RelA-SpoT, it would have been expressed from its native promoter on the metagenomic DNA during the salt-screening of the metagenomic library. It would thus be safe to assume that the native promoter should be recognised by the *E. coli* BL21(DE3) expression strain and the HD-hydrolase should be expressed. Expression of CESP and the Na^+/H^+ antiporter could not be detected via SDS-PAGE analysis in either of the expression strains. Expression optimisation

of these two proteins was attempted by increasing the concentration of IPTG used for inducing expression from 0.4 mM to 0.8 mM, incubating induced cultures for 4 hrs, 8 hrs and overnight and reducing the incubation temperature during induction from 37°C to 30°C. However, expression could still not be detected (data not shown).

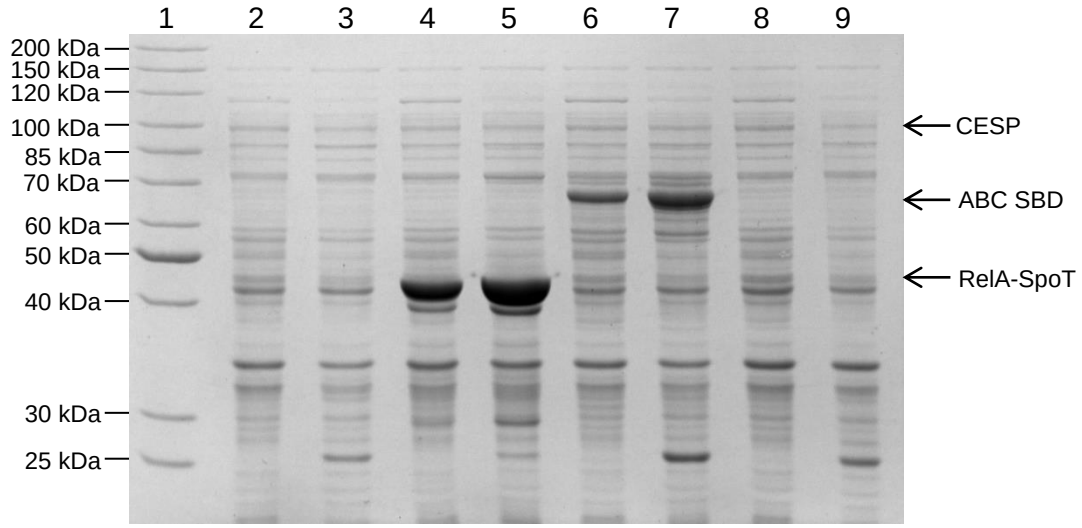


Figure 34. 10% SDS-PAGE analysis of proteins expressed in *E. coli* BL21(DE3) and Rosetta(DE3)pLysS cells. Total protein extract was analysed in each lane. Lane 1: PageRuler Unstained Protein Ladder (Thermo Fisher Scientific). Lane 2: BL21:pET21a. Lane 3: Rosetta:pET21a. Lane 4: BL21:RelA-SpoT. Lane 5: Rosetta:RelA-SpoT. Lane 6: BL21:ABC SBD. Lane 7: Rosetta:ABC SBD. Lane 8: BL21:CESP. Lane 9: Rosetta:CESP. Arrows on the right indicate the expected position of each protein.

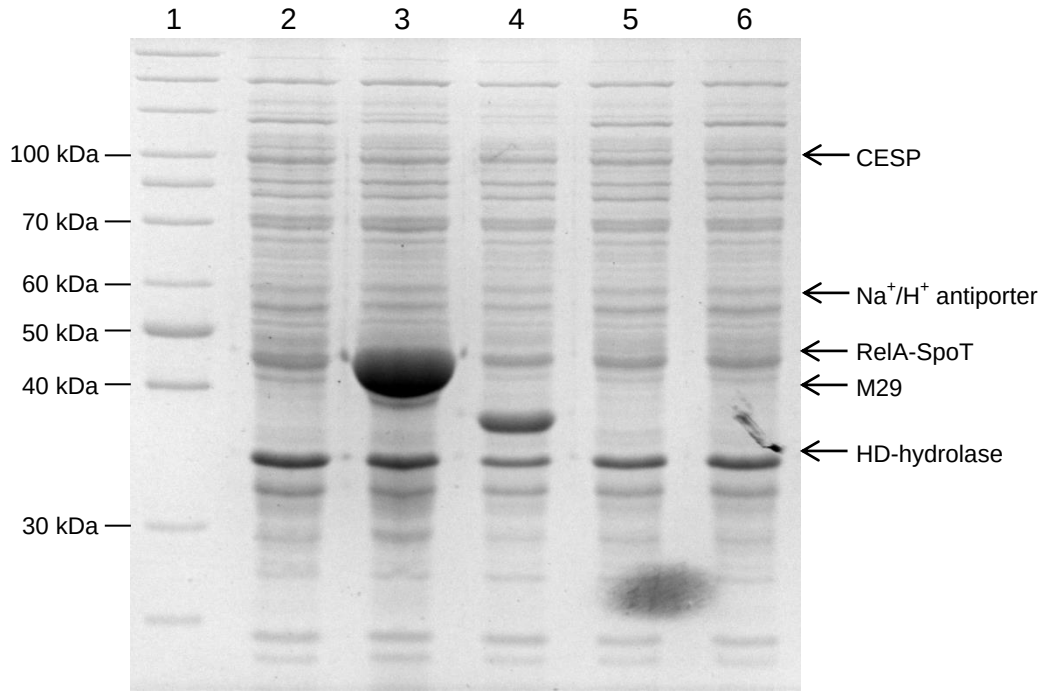


Figure 35. 10% SDS-PAGE analysis of proteins expressed in *E. coli* BL21(DE3) cells. Total protein extract was analysed in each lane. Lane 1: PageRuler Unstained Protein Ladder. Lane 2: BL21:pET21a. Lane 3: BL21:RelA-SpoT:HD-hydrolase. Lane 4: BL21:M29. Lane 5: BL21:CESP. Lane 6: BL21: Na^+/H^+ antiporter. Arrows on the right indicate the expected position of each protein.

5.2.2 *In vivo* Testing for Stress-tolerance

Initially RelA-SpoT, ABC SBD and CESP expressed in *E. coli* BL21(DE3) and Rosetta(DE3)pLysS cells were tested for salt-tolerance activity by spread-plating induced cultures on LB agar containing 5% NaCl. However, no growth was observed on any of the stress plates after 48 hrs incubation at 37°C. The ability of the thermophilic peptidase M29 to confer heat-tolerance to *E. coli* BL21(DE3) was tested by incubating serial dilutions plated on LB agar supplemented with ampicillin and 0.4 mM IPTG at 46°C. No growth was observed on the BL21:pET21a or BL21:M29 plates.

Thereafter, tolerance of the RelA-SpoT:HD-hydrolase, M29, CESP, Na^+/H^+ antiporter and pET21a *E. coli* BL21(DE3) cultures to 5% NaCl was tested in liquid culture (Figure 36). The tolerance of the BL21:Na^+/H^+ antiporter to pH 9 and 5% NaCl-pH 9 was tested in liquid culture (Figure 37). All stress culture growth curves were plotted on the same y-axis scale as the control cultures in LB broth containing no stress (Figure 38). None of the cultures containing stress-related genes showed significantly increased growth, compared to the *E. coli* BL21(DE3) host harbouring an empty pET21a vector, in liquid stress cultures.

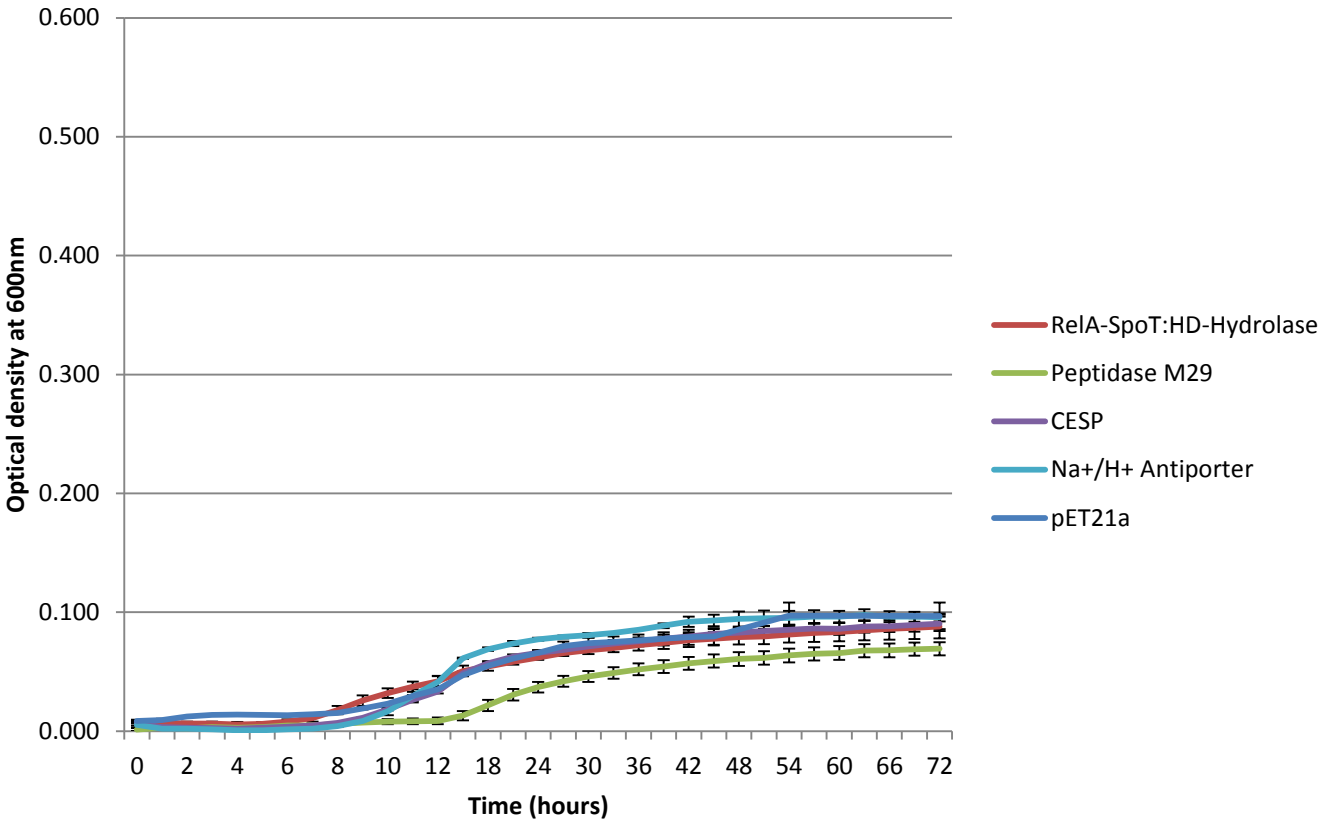


Figure 36. Growth curves of *E. coli* BL21(DE3) cultures containing stress-related genes in LB broth supplemented with 5% NaCl and ampicillin. Results are shown as the average of triplicate experiments, with error bars showing the standard error of the mean (SEM).

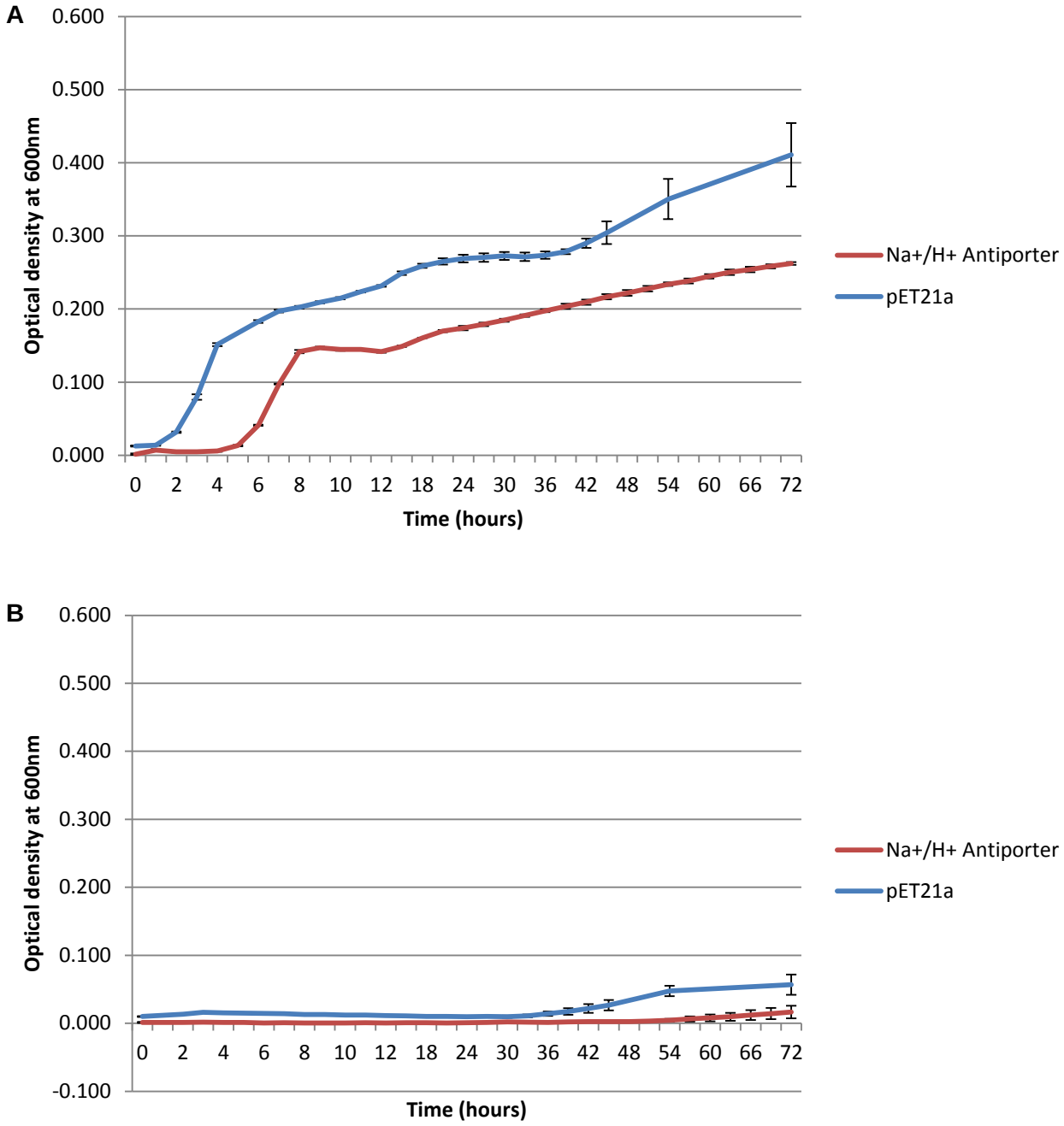


Figure 37. Growth curves of *E. coli* BL21(DE3) cultures containing stress-related genes in A) LB broth adjusted to pH 9 and supplemented with ampicillin and B) LB broth adjusted to pH 9 and supplemented with 5% NaCl and ampicillin. Results are shown as the average of triplicate experiments, with error bars showing the SEM.

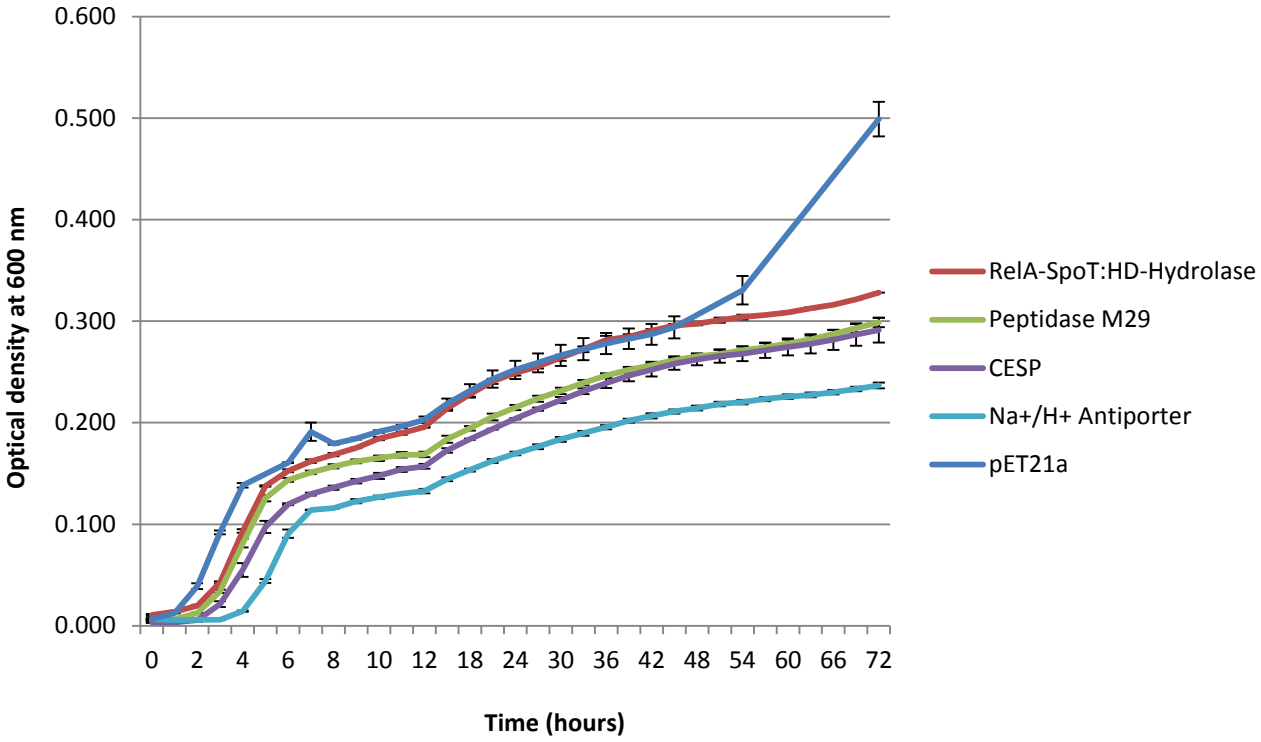


Figure 38. Growth curves of *E. coli* BL21(DE3) cultures containing stress-related genes in LB broth supplemented with ampicillin. Results are shown as the average of triplicate experiments, with error bars showing the SEM.

As no increased tolerance of the stress-related genes to 5% NaCl was observed on solid, or in liquid media, the NaCl concentration was decreased. The RelA-SpoT:HD-hydrolase, M29 and Na^+/H^+ antiporter *E. coli* BL21(DE3) cultures were tested for salt-tolerance on LB agar plates containing 3.5% NaCl. Serial dilutions were plated out on the stress plates as well as corresponding control LB agar plates. The survival rate of each culture was determined (see section 2.8). Experiments were performed at least twice, in triplicate (Figure 39).

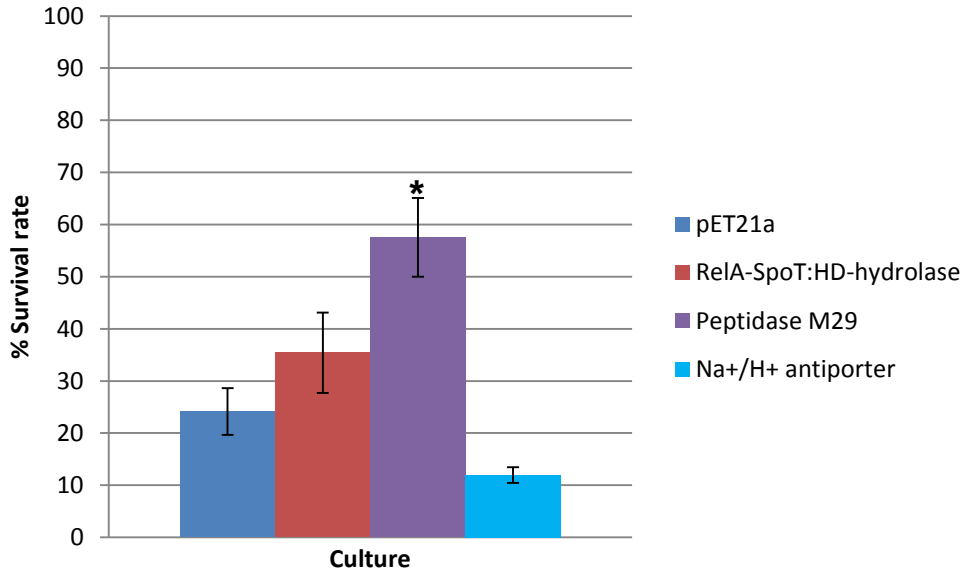


Figure 39. Survival rates of *E. coli* BL21(DE3) cultures containing stress-related genes on LB agar supplemented with 3.5% NaCl, ampicillin and 0.4 mM IPTG. Error bars show the SEM. Asterisks show cultures with a statistically significant survival rate (*P*-value <0.05) compared to the control (BL21:pET21a), as determined using Student's unpaired t-test.

The *E. coli* BL21(DE3) transformants expressing peptidase M29 showed statistically significant higher survival rates on 3.5% NaCl LB agar plates than the *E. coli* BL21(DE3) host harbouring an empty pET21a vector (P-value <0.05, using Student's unpaired t-test) (Figure 39 and Table 29). Transformants expressing RelA-SpoT and HD-hydrolase showed higher survival rates compared to the control (BL21:pET21a), but these were not statistically significant. Transformants expressing the Na^+/H^+ antiporter showed lower survival rates compared to the control, but these were not significantly different from the control (Table 29). Cultures expressing peptidase M29 also showed a statistically significant higher survival rate than cultures expressing the Na^+/H^+ antiporter, but the survival rate of the other cultures were not significantly different from each other (Appendix C: Table C1).

Table 29. Statistical significance of survival rates determined using Student's unpaired t-test.

***E. coli* BL21(DE3) Culture**	***P*-value***
RelA-SpoT:HD-hydrolase	0.24
Peptidase M29	0.0051
Na^+/H^+ antiporter	0.070

**P*-values <0.05 show significant differences in survival rates compared to the host.

5.3 Discussion

In this study, six novel genes considered to be good candidates for the conveyance of salt-tolerance to *E. coli* were successfully cloned in the pET21a expression vector. The proteins encoded by these genes were classified and characterised through the use of various bioinformatic tools. Three of the six proteins, ABC SBD, RelA-SpoT and peptidase M29, were successfully over-expressed in *E. coli* Rosetta(DE3)pLysS and/ or BL21(DE3) cells. Interestingly, all three proteins were predicted to be hydrophilic through bioinformatic analysis. The HD-hydrolase was not cloned under the control of the inducible IPTG promoter and it was thus difficult to determine, using SDS-PAGE and Coomassie staining, whether this protein was expressed.

The peptidase M29 was shown to confer salt-tolerance to *E. coli* BL21(DE3) on solid media. The *E. coli* BL21(DE3) expressing peptidase M29 showed a significantly higher survival rate on LB agar plates containing 3.5% NaCl, compared to *E. coli* BL21:pET21a. Bioinformatic analysis of the peptidase M29 identified in this study showed that the protein contained an isoleucine, instead of a threonine, in an otherwise well-conserved position (Figure 30). This substitution of a polar amino acid with a non-polar residue could have both functional and structural consequences, which could be investigated in future studies through comparisons to homologues containing a threonine at this position. In this study, a putative novel function (salt-tolerance) has been described for this aminopeptidase of the M29 superfamily. The mechanism by which the peptidase M29 conferred salt-tolerance to *E. coli* is unknown. Hyperosmotic stress reduces the cytoplasmic volume, leading to macromolecular crowding that can exacerbate protein misfolding and lead to the formation of insoluble protein aggregates, which are detrimental to cellular functioning (DIAMANT *et al.* 2001; WOOD 2011). It is possible that the peptidase M29 reduced the prevalence of protein aggregates during hyperosmotic stress thereby decreasing the detrimental effects of these aggregates, which allowed the cells to efficiently respond to the salt-induced stress. *E. coli* BL21(DE3) cultures expressing the peptidase M29 in liquid culture supplemented with 5% NaCl did not have increased growth rates compared to *E. coli* BL21:pET21a (Figure 36). The optical density of all cultures was extremely low in these experiments, as OD_{600nm} of > 2 are routinely obtained. This could be as a result of the small volumes and inoculums (OD_{600nm} < 0.01) used in the 96 well plates, which could increase the lag phase during which the strains could be more susceptible to stress. Evidence of this increased lag phase is potentially shown in Figure 38, where the *E. coli* BL21:pET21a culture shows increased growth after 54 hrs. This hypothesis should be investigated in future studies, to determine the optimal amount of inoculum. Further studies are required to characterise and classify the M29 family of

aminopeptidases in greater detail, as metagenomic studies provides more sequence and functional data on these proteins.

Two genes, encoding RelA-SpoT and HD-hydrolase proteins, were cloned into the same vector as these genes are involved in the same pathway to produce and hydrolyse ppGpp, a transcriptional regulator of the general stress response sigma factor, RpoS, in *E. coli* (BATTESTI *et al.* 2011). Co-expression of these genes in the same cell was attempted in this way, as the genes were separated by a relatively large distance on STC 1 (Appendix B: Figure B2 and Table B1) and thus did not form part of an operon, as might be expected.

The other proteins that were successfully expressed did not show significantly greater salt-tolerance compared to the BL21:pET21a control. There could be numerous explanations for this observation. Firstly, the proteins may not confer salt-tolerance. This is more likely for ABC SBD, as RelA-SpoT homologues have been shown to confer salt-tolerance to *E. coli* and other organisms (VAN DER BIEZEN *et al.* 2000; YAMADA *et al.* 2003; OCHI *et al.* 2012). Secondly, the salt-tolerant phenotype of the STCs observed in Chapter 3 could be a polygenic effect. Therefore multiple genes may have to be sub-cloned to confer salt-tolerance. An alternative explanation could be that *E. coli* BL21(DE3) is unable to produce functionally active forms of these heterologous proteins due to codon usage bias. These proteins all contained >9% rare codons and *E. coli* BL21(DE3) does not contain genes to supply rare codon tRNAs, as *E. coli* Rosetta(DE3)pLysS does (GUSTAFSSON *et al.* 2004). This is supported by the SDS-PAGE analysis (Figure 34), which shows better expression of the same proteins in *E. coli* Rosetta(DE3)pLysS than BL21(DE3), suggesting that the six rare codons encoded on the pLysSRARE plasmid aid expression of the target proteins. The difference in survival rate under stress between *E. coli* BL21(DE3) and Rosetta(DE3)pLysS has previously been observed by ANDERSON (2012), where a bacterial water hypersensitivity (WHy) protein showed higher survival rates at 3.5% NaCl compared to the BL21:pET21 control, but the values were not significantly different. When *in vivo* assays were performed in an *E. coli* strain that provides rare codons, such as Rosetta(DE3)pLysS, statistically significant salt-tolerance was observed in cultures expressing the WHy protein. Despite the high percentage of rare codons and the occurrence of several consecutive rare codons in each protein, expression could still be detected. However, codon bias can result in missense amino acid insertions causing a reduction in activity of the expressed protein, or frameshift events that result in premature termination of translation and result in a truncated protein being produced (KURLAND AND GALLANT 1996). The pLysSRARE plasmid harboured by the *E. coli* Rosetta(DE3)pLysS strain used in this study encodes the majority of the necessary rare tRNAs that correspond to the codon usage of the target genes (SØRENSEN AND

MORTENSEN 2005). However, more rare tRNAs can be provided by using the *E. coli* Rosetta2(DE3)pLysS strain that contains the pLysSRARE2 plasmid (Novagen). Assuming the protein produced is free of amino acid substitutions and is full-length, the formation of inclusion bodies, which are aggregates of misfolded proteins in the cytoplasm, is common during over-expression of heterologous proteins (HANNIG AND MAKRIDES 1998). Due to the methods used in this study, it was not possible to distinguish between improper folded proteins in inclusion bodies and correctly folded and functional proteins. This situation is clearly disadvantageous when attempting to demonstrate *in vivo* function of the heterologous protein.

Over-expression of the two hydrophobic, integral membrane proteins, CESP and Na^+/H^+ antiporter, in *E. coli* BL21(DE3) or Rosetta(DE3)pLysS could not be detected via SDS-PAGE analysis, despite attempts to optimise the expression parameters. Neither protein conferred stress-tolerance to these *E. coli* strains. The over-expression of heterologous membrane proteins in *E. coli* is known to be difficult, due to potential toxic effects of the proteins, molecular crowding in membranes, improper detection of signal peptides and the lack of appropriate chaperones to facilitate correct folding (CHEN 2012). Studies that have shown phenotypic activity of Na^+/H^+ antiporters do not attempt to over-express the protein, but simply allow the host to express the protein from its native promoter (ZHANG *et al.* 2014). This is advantageous, as the production of the heterologous protein does not impose such a large energy requirement on the cell. Reduced expression levels of the membrane protein should also facilitate in more precise folding and correct localisation of the protein, as protein aggregates and inclusion bodies are less likely to form (BANEYX 1999). However, the host would have to be able to recognise the foreign promoter.

The novel genes and their corresponding proteins identified, characterised, cloned and expressed in this study all harbour some potential to confer stress-tolerance to *E. coli*. A peptidase M29 was shown to confer salt-tolerance to *E. coli* BL21(DE3). In order to further demonstrate, and potentially elucidate the mechanism of salt-tolerance provided by the peptidase M29 *in vivo*, heterologous complementation studies could be performed in the osmosensitive *E. coli* strain MKH13. The other putative stress-related proteins can also be tested for salt-tolerant phenotypes in appropriate mutant strains. For example, in the *E. coli* KNabc strain the Na^+/H^+ antiporters have been knocked-out, which would therefore be ideal to study CESP and the novel Na^+/H^+ antiporter identified here. Furthermore, *E. coli* strains in which the *relA* and *spoT* genes are knocked-out have been constructed to study RelA-SpoT homologues (GENTRY AND CASHEL 1996; WENDRICH *et al.* 2002). Lastly, the osmosensitive

E. coli strain MKH13 can be used as a system to identify and study additional novel osmotic stress-tolerance genes from metagenomes (CULLIGAN *et al.* 2013).

Chapter 6: General Discussion and Conclusions

Metagenomics enables researchers to study the genetics of as-yet-uncultured microorganisms. A sequencing-based screening approach can provide vast amounts of data about a microbial community as a whole and about individual cells within that community. However, the power of this approach is limited by sequence data currently present in databases, as a large proportion of genes cannot be assigned a putative function. Nonetheless, genes that do contain homologues in databases can provide insight into the evolutionary relationships of the organism from which a segment of DNA was obtained, as well as information about the potential ecological role the organism may fulfil. Functional-based screening allows for the identification of entirely novel genes to which functions can be assigned based on the parameters used during the initial screening procedure and subsequent studies to characterise the gene product. This approach is particularly effective when screening small insert libraries or when using transposon mutagenesis in conjunction with screening of large insert libraries. Functional characterisation of genes identified from a metagenomic library can assign putative ecological roles for the organism from which the genes were obtained. Both of the above-mentioned culture-independent approaches are useful in exploring the genetic resource of uncultured microorganisms not only for evolutionary and ecologically important information, but also for the discovery of novel genes with potentially valuable applications.

The aim of this study, to identify novel stress-tolerance genes from a large-insert Namib Desert soil metagenomic library, was achieved by identifying 12 salt-tolerant clones (STCs) through functional screening; sequencing the clones using Ion Torrent technology and using bioinformatic tools to identify novel salt-tolerance genes. This hybrid approach between functional- and sequencing-based methods provided information in two ways: functional screening under hyperosmotic stress showed microorganisms are adapted to the extreme osmotic challenges of the Namib Desert soil, whereas sequencing of these clones showed that a variety of different methods are employed by these microorganisms to withstand abiotic stress. Genes could thus be identified that have potential industrial applications or contribute to the existing knowledge of stress-adaptation in microorganisms. Furthermore, phylogenetic analysis of these sequences provided data about the probable identity of the organism from which the DNA originated. Functional screening of the Namib Desert soil metagenome also identified three heat-tolerant clones, but no UVB radiation- or oxidative stress-tolerant clones were identified.

Phylogenetic analysis of each clone revealed that a variety of bacteria, Gram-positive and Gram-negative, were present in the soil and sequences grouped into phyla that have been observed in Namib Desert soil, namely Deltaproteobacteria, Fibrobacteres, Cyanobacteria, Actinobacteria and Planctomyces (MAKHALANYANE *et al.* 2013). Sequence analyses showed that these organisms applied various strategies to resist not only hyperosmotic stress, but also other abiotic stresses, particularly oxidative stress. Genes encoding proteins that were part of multi-unit transport systems or membrane channel proteins involved in osmotic stress-adaptation strategies were present on STC 1, 9, 12, 18, 13, 88, 90 and 92. Reductases involved in preventing oxidative damage to macromolecules, such as DNA and proteins, were present on STC 88, 91 and 92. A phage-shock protein, which forms part of a general stress response operon, was identified on STC 90. Two enzymes involved in the production of the bacterial alarmone, ppGpp, were identified on STC 1 and 18. Production of the alarmone is induced upon amino acid starvation and other stresses and activates the stringent response in bacteria, which is expected due to the oligotrophic nature of Namib Desert soil. A penicillin-binding protein and a putative bacteriocin were identified on STC 89, which are both probably necessary to survive in the competitive soil microbial communities. The metagenomic inserts sequenced in this study contained a high percentage of hypothetical proteins, ranging from 31% to 59%. The dependence of sequence-based screening on gene annotations in databases is highlighted when this observation is considered with the fact that osmotic stress-related genes could not be identified on all STCs. Furthermore, the low amino acid identity (often below 60%) of proteins encoded on the metagenomic inserts to homologues illustrates the extreme novelty of the Namib Desert soil metagenomic sequences and highlights the unique and untapped microbial communities of this environment.

The 12 STCs identified in this study showed variable resistance to an increased concentration of NaCl in growth media. Four clones (STC 1, 9, 12 and 18) containing almost exactly the same segment of DNA, most probably from individuals of the same species, showed variable performance under hyperosmotic conditions. It was noted, through bioinformatic analysis, that the difference might be the presence of a hypothetical protein present on the two clones that showed increased salt-tolerance (STC 1 and 18) that was absent on the two clones with reduced salt-tolerance (STC 9 and 12). This hypothetical protein, containing a RelA-SpoT-like domain, did not confer salt-tolerance to *E. coli*, despite the fact that homologous proteins have been shown to confer salt-tolerance to *E. coli* (YAMADA *et al.* 2003). A possible explanation for this is that the salt-tolerant phenotype was most probably a synergistic effect of multiple genes present on the metagenomic insert of these clones. *E. coli* BL21(DE3) simultaneously expressing the RelA-SpoT-like protein and

HD-hydrolase protein, with complementary functions, found on the same metagenomic insert, did not result in significantly higher salt-tolerance compared to the control culture, suggesting that other genes on the insert may be needed to confer salt-tolerance. Individual expression of other potential stress-response genes did not confer salt-tolerance to *E. coli* BL21(DE3), which could again be explained by the synergistic effect hypothesis. Two genes, encoding membrane transport proteins, were not successfully expressed, but it is hypothesised that if these were to be expressed, particularly in an osmosensitive *E. coli* strain or a strain lacking the homologous transport protein genes, salt-tolerance would be conferred to these strains. Another possible explanation for the failure of these proteins to confer stress-tolerance to *E. coli* BL21(DE3) or Rosetta(DE3)pLysS is that the heterologous proteins might be incorrectly folded and present as inclusion bodies and are thus inactive (SØRENSEN AND MORTENSEN 2005). The process for identifying the gene, or genes, responsible for the salt-tolerant phenotype could be improved through the use of transposon mutagenesis, digesting the fosmid and sub-cloning the fragments or sub-cloning every gene present individually. The identification and successful sub-cloning of six potential stress-tolerance genes from a Namib Desert soil metagenomic library in this study allows for future research to investigate the functions and characteristics of these proteins in various organisms. This could expand basic knowledge of the mechanisms microorganisms employ to survive in the harsh desert environment and could lead to application in other fields or industry.

In this study, a novel metalloprotease of the thermophilic M29 superfamily, peptidase M29, was shown to confer salt-tolerance to *E. coli* BL21(DE3). The peptidase M29 has potential application in various industrial processes. Proteases are included in formulations of detergents and contact lens solutions and are involved in processes such as cheese production and meat and leather hide processing (SARETHY *et al.* 2011). Activity of a protease at alkaline pH, high temperatures, high salt or organic solvent concentrations or a combination of these is required to be applicable in most commercial or industrial processes (GUPTA *et al.* 2002). Peptidase M29 was isolated from a hot, saline environment. It should therefore be active at high temperatures (45 – 60°C) and high salt concentrations. The substrate specificity and activity of this enzyme should be tested over a range of temperatures, pH levels and salt concentrations, to determine the optimal parameters for protease activity. An ideal commercial operation could then be chosen in which to apply this protease, based on the biochemical properties of the enzyme, which should match the conditions (particularly the pH) of the commercial process as closely as possible and should not be inhibited by any chemicals present in the solutions throughout the process (RAO *et al.* 1998). This particular peptidase M29 is an intracellular enzyme, which is not ideal when the

protein must be produced on an industrial scale, as it complicates the purification process (RAO *et al.* 1998). Therefore, the peptidase M29 could be engineered to contain a signal peptide, recognised by the relevant expression host, at the N-terminus which would allow the protease to be secreted from the cell.

Two membrane cation transporters were identified in this study, a Na^+/H^+ antiporter and a monovalent cation export system protein (CESP). Ectopic expression of Na^+/H^+ antiporters in *Arabidopsis thaliana* and *Nicotiana tabacum* resulted in an increase in salt- and alkali-tolerance of these plants (SHI *et al.* 2003; ZHONG *et al.* 2012). In the study by ZHONG *et al.* (2012), the transporter was of bacterial origin and localised to the tonoplast, the membrane surrounding the vacuole in plant cells, where intracellular Na^+ and pH levels are regulated (RODRÍGUEZ-ROSALES *et al.* 2009). Increasing salt- and/or alkali-tolerance in plants, particularly crops, could allow previously non-arable land to now be arable, thus increasing crop production and contributing to the attempts to alleviate food shortages across the globe. The novel Na^+/H^+ antiporter and CESP could be tested for their predicted function *in vivo* and *in vitro* using bacterial systems and thereafter be ectopically expressed in a plant system to determine if these proteins confer salt- and/ or alkali-tolerance to plants.

The RelA-SpoT-like protein and the HD-hydrolase identified in this study contribute to the growing list of enzymes involved in ppGpp synthesis. The production of this molecule is induced by various stresses in both bacteria and plants where it regulates expression of various genes, including up-regulation of general stress-response transcription factors and the sigma factor, RpoS (GIVENS *et al.* 2004; MAGNUSSON *et al.* 2005; BRAEKEN *et al.* 2006). The RelA-SpoT proteins encoded by most bacteria and plants are approximately 700 amino acids in length and contain four distinct functional domains. The RelA-SpoT-like protein identified in this study was 415 amino acids in length and contained only a single functional domain, the RelA-SpoT synthetase domain. These truncated RelA-SpoT-like proteins seem to be less prevalent than the larger homologues. Two such truncated RelA-SpoT-like proteins have been identified in *Streptococcus mutans* and knock-out studies of these two genes and the gene encoding the full-length RelA-SpoT protein showed that these proteins can functionally substitute each other under various conditions to ensure ppGpp is produced (LEMOS *et al.* 2007). Functional redundancy of genes is common in nature and the fact that it is present for ppGpp production is an indication of the importance of this molecule for cellular functioning, particularly under stress. The identification of the truncated protein in a Namib Desert soil metagenomic library could indicate the importance of ppGpp as a stress-response regulator of bacteria in this environment. Researchers investigating ppGpp metabolism and RelA-SpoT proteins should attempt to identify truncated versions of these proteins in the genome of the organism under investigation. This would indicate whether this

redundancy is ubiquitous throughout nature or only evolved in certain phyla or organisms, which, in turn, could highlight the conditions under which ppGpp is necessary or the most beneficial and thus expanding our knowledge of stress-adaptation strategies on a cellular level.

This study is the first functional metagenomics investigation of a Namib Desert soil metagenome. Through sequencing of a portion of the metagenome it provided insights into the possible mechanisms employed by bacteria to withstand abiotic stresses such as high temperature, high salinity and oxidative stress in this environment. A suite of bioinformatic tools were used to classify and characterise novel potential stress-related genes and the encoded proteins. Subsequent cloning and heterologous protein expression identified a peptidase M29 that conferred salt-tolerance to *E. coli* BL21(DE3). Other proteins that were investigated did not confer stress-tolerance to *E. coli* BL21(DE3) or Rosetta(DE3)pLysS. However, stress-tolerance is, in most cases, a synergistic effect involving multiple genes and pathways. Therefore, to advance understanding of the mechanisms employed by microorganisms to tolerate the numerous, often simultaneous, abiotic stresses of environments such as the Namib Desert, metagenomic data should be linked to metatranscriptomic and metaproteomic data to obtain a more holistic view of the microbial community. Finally, only by analysing a stress-related gene, or protein, as part of a whole, in its native system, will we be able to understand the range of its effects and thus apply it more successfully in foreign environments and systems.

References

Aakvik, T., K. F. Degnes, R. Dahlsrud, F. Schmidt, R. Dam *et al.*, 2009. A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. FEMS Microbiology Letters 296**:** 149-158.

Aklujkar, M., S. Haveman, R. DiDonato, O. Chertkov, C. Han *et al.*, 2012. The genome of *Pelobacter carbinolicus* reveals surprising metabolic capabilities and physiological features. BMC Genomics 13**:** 690.

Almeida, B., I. Vaz-Moreira, P. Schumann, O. C. Nunes, G. Carvalho *et al.*, 2013. *Patulibacter medicamentivorans* sp. nov., isolated from activated sludge of a wastewater treatment plant. International Journal of Systematic and Evolutionary Microbiology 63**:** 2588-2593.

Amann, R. I., W. Ludwig and K. H. Schleifer, 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiological Reviews 59**:** 143-169.

Amanullah, A., C. J. Hewitt, A. W. Nienow, C. Lee, M. Chartrain *et al.*, 2003. Measurement of strain-dependent toxicity in the indene bioconversion using multiparameter flow cytometry. Biotechnology and Bioengineering 81**:** 405-420.

Anderson, D. E., 2012. Metagenome sequencing and *in silico* gene discovery: from genetic potential to function, pp. 249 in *Department of Biotechnology*. University of the Western Cape, University of the Western Cape., PhD thesis.

Anderson, I., E. Saunders, A. Lapidus, M. Nolan, S. Lucas *et al.*, 2012. Genome sequence of the thermophilic sulfate-reducing ocean bacterium *Thermodesulfatator indicus* type strain (CIR29812 T). Standards in Genomic Sciences 6**:** 155-164.

Angelov, A., M. Mientus, S. Liebl and W. Liebl, 2009. A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. Systematic and Applied Microbiology 32**:** 177-185.

Aravind, L., and E. V. Koonin, 1998. The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends in Biochemical Sciences 23**:** 469-472.

Baneyx, F., 1999. Recombinant protein expression in *Escherichia coli*. Current Opinion in Biotechnology 10**:** 411-421.

Baqué, M., E. Viaggiu, G. Scalzi and D. Billi, 2013. Endurance of the endolithic desert cyanobacterium *Chroococcidiopsis* under UVC radiation. Extremophiles 17**:** 161-169.

Barrow, C. J., 1992. World atlas of desertification (United Nations environment programme), edited by N. Middleton and D. S. G. Thomas. Land Degradation & Development 3**:** 249.

Battesti, A., N. Majdalani and S. Gottesman, 2011. The RpoS-mediated general stress response in *Escherichia coli*. Annual Review of Microbiology 65**:** 189-213.

Bauermeister, A., E. Bentchikou, R. Moeller and P. Rettberg, 2009. Roles of PprA, IrrE, and RecA in the resistance of *Deinococcus radiodurans* to germicidal and environmentally relevant UV radiation. Archives of Microbiology 191**:** 913-918.

Bauermeister, A., R. Moeller, G. Reitz, S. Sommer and P. Rettberg, 2011. Effect of relative humidity on *Deinococcus radiodurans*' resistance to prolonged desiccation, heat, ionizing, germicidal, and environmentally relevant UV radiation. Microbial Ecology 61**:** 715-722.

Berntsson, R. P. A., S. H. J. Smits, L. Schmitt, D.-J. Slotboom and B. Poolman, 2010. A structural classification of substrate-binding proteins. FEBS Letters 584**:** 2606-2617.

Bertin, P. B., S. P. Lozzi, J. K. Howell, G. Restrepo-Cadavid, D. Neves *et al.*, 2005. The thermophilic, homohexameric aminopeptidase of *Borrelia burgdorferi* is a member of the M29 family of metallopeptidases. Infection and Immunity 73**:** 2253-2261.

Biemans-Oldehinkel, E., and B. Poolman, 2003. On the role of the two extracytoplasmic substrate-binding domains in the ABC transporter OpuA. The EMBO Journal 22**:** 5983-5993.

Billi, D., 2009. Subcellular integrities in *Chroococcidiopsis* sp. CCMEE 029 survivors after prolonged desiccation revealed by molecular probes and genome stability assays. Extremophiles 13**:** 49-57.

Billi, D., and M. Potts, 2002. Life and death of dried prokaryotes. Research in Microbiology 153**:** 7-12.

Bond, D. R., D. E. Holmes, L. M. Tender and D. R. Lovley, 2002. Electrode-reducing microorganisms that harvest energy from marine sediments. Science 295**:** 483-485.

Bonfield, J. K., and A. Whitwham, 2010. Gap5—editing the billion fragment sequence assembly. Bioinformatics 26**:** 1699-1703.

Bouveret, E., and A. Battesti, 2011. The stringent response, pp. 231-250 in *Bacterial Stress Responses*, edited by G. Storz and R. Hengge. American Society for Microbiology Press.

Braeken, K., M. Moris, R. Daniels, J. Vanderleyden and J. Michiels, 2006. New horizons for (p)ppGpp in bacterial and plant physiology. Trends in Microbiology 14**:** 45-54.

Bunthof, C. J., S. van den Braak, P. Breeuwer, F. M. Rombouts and T. Abee, 1999. Rapid fluorescence assessment of the viability of stressed *Lactococcus lactis*. Applied and Environmental Microbiology 65**:** 3681-3689.

Burley, S. K., P. R. David and W. N. Lipscomb, 1991. Leucine aminopeptidase: bestatin inhibition and a model for enzyme-catalyzed peptide hydrolysis. Proceedings of the National Academy of Sciences 88**:** 6916-6920.

Cadet, J., T. Douki, D. Gasparutto and J.-L. Ravanat, 2005. Radiation-induced damage to cellular DNA: measurement and biological role. Radiation Physics and Chemistry 72**:** 293-299.

Caruso, T., Y. Chan, D. C. Lacap, M. C. Lau, C. P. McKay *et al.*, 2011. Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. The ISME Journal 5**:** 1406-1413.

Cary, S. C., J. E. Barrett, I. R. McDonald and D. A. Cowan, 2010. On the rocks: the microbiology of Antarctic Dry Valley soils. Nature Reviews Microbiology 8**:** 129-138..

Cava, F., A. Hidalgo and J. Berenguer, 2009. *Thermus thermophilus* as biological model. Extremophiles 13**:** 213-231.

Chan, Y., D. C. Lacap, M. C. Y. Lau, K. Y. Ha, K. A. Warren-Rhodes *et al.*, 2012. Hypolithic microbial communities: between a rock and a hard place. Environmental Microbiology 14**:** 2272-2282.

Chen, R., 2012. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. Biotechnology Advances 30**:** 1102-1107.

Chevreux, B., T. Wetter and S. Suhai, 1999. Genome sequence assembly using trace signals and additional sequence information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics 99**:** 45-56.

Conroy, O., E.-H. Kim, M. M. McEvoy and C. Rensing, 2010. Differing ability to transport nonmetal substrates by two RND-type metal exporters. FEMS Microbiology Letters 308**:** 115-122.

Cotter, P. D., R. P. Ross and C. Hill, 2013. Bacteriocins - a viable alternative to antibiotics? Nature Reviews Microbiology 11**:** 95-105.

Couto, G. H., A. Glogauer, H. Faoro, L. S. Chubatsu, E. M. Souza *et al.*, 2010. Isolation of a novel lipase from a metagenomic library derived from mangrove sediment from the south Brazilian coast. Genetics and Molecular Research 9**:** 514-523.

Cowan, D., Q. Meyer, W. Stafford, S. Muyanga, R. Cameron *et al.*, 2005. Metagenomic gene discovery: past, present and future. Trends in Biotechnology 23**:** 321-329.

Cowan, D. A., N. Khan, S. B. Pointing and S. C. Cary, 2010. Diverse hypolithic refuge communities in the McMurdo Dry Valleys. Antarctic Science 22**:** 714-720.

Cowan, D. A., N. J. Russell, A. Mamais and D. M. Sheppard, 2002. Antarctic Dry Valley mineral soils contain unexpectedly high levels of microbial biomass. Extremophiles 6**:** 431-436.

Cox, M. M., and J. R. Battista, 2005. *Deinococcus radiodurans* -- the consummate survivor. Nature Reviews Microbiology 3**:** 882-892.

Craig, J. W., F.-Y. Chang, J. H. Kim, S. C. Obiajulu and S. F. Brady, 2010. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range

cosmid environmental DNA libraries in diverse Proteobacteria. Applied and Environmental Microbiology 76**:** 1633-1641.

Crooks, G. E., G. Hon, J.-M. Chandonia and S. E. Brenner, 2004. WebLogo: a sequence logo generator. Genome Research 14**:** 1188-1190.

Crowe, J. H., J. F. Carpenter and L. M. Crowe, 1998. The role of vitrification in anhydrobiosis. Annual Review of Physiology 60**:** 73-103.

Crowe, J. H., L. M. Crowe, J. E. Carpenter, S. Petrelski, F. A. Hoekstra *et al.*, 2010. Anhydrobiosis: cellular adaptation to extreme dehydration, pp. 1445-1477 in *Comprehensive Physiology*. John Wiley & Sons, Inc.

Csonka, L. N., 1989. Physiological and genetic responses of bacteria to osmotic stress. Microbiological Reviews 53**:** 121-147.

Culligan, E. P., R. D. Sleator, J. R. Marchesi and C. Hill, 2012. Functional metagenomics reveals novel salt tolerance loci from the human gut microbiome. The ISME Journal 6**:** 1916-1925.

Culligan, E. P., R. D. Sleator, J. R. Marchesi and C. Hill, 2013. Functional environmental screening of a metagenomic library identifies *stlA*; a unique salt tolerance locus from the human gut microbiome. PLoS ONE 8**:** e82985.

da Silva, T. L., A. Reis, C. A. Kent, M. Kosseva, J. C. Roseiro *et al.*, 2005. Stress-induced physiological responses to starvation periods as well as glucose and lactose pulses in *Bacillus licheniformis* CCMI 1034 continuous aerobic fermentation processes as measured by multi-parameter flow cytometry. Biochemical Engineering Journal 24**:** 31-41.

Da Silveira, M. G., and T. Abee, 2009. Activity of ethanol-stressed *Oenococcus oeni* cells: a flow cytometric approach. Journal of Applied Microbiology 106**:** 1690-1696.

Daly, M. J., 2009. A new perspective on radiation resistance based on *Deinococcus radiodurans*. Nature Reviews Microbiology 7: 237-245.

Daly, M. J., 2012. Death by protein damage in irradiated cells. DNA Repair 11**:** 12-21.

Daly, M. J., E. K. Gaidamakova, V. Y. Matrosova, J. G. Kiang, R. Fukumoto *et al.*, 2010. Small-molecule antioxidant proteome-shields in *Deinococcus radiodurans*. PLoS ONE 5**:** e12570.

Daly, M. J., E. K. Gaidamakova, V. Y. Matrosova, A. Vasilenko, M. Zhai *et al.*, 2007. Protein oxidation implicated as the primary determinant of bacterial radioresistance. PLoS Biology 5**:** e92.

Daniel, R., 2004. The soil metagenome – a rich resource for the discovery of novel natural products. Current Opinion in Biotechnology 15**:** 199-204.

Darwin, A. J., 2005. The phage-shock-protein response. Molecular Microbiology 57**:** 621-628.

Diamant, S., N. Eliahu, D. Rosenthal and P. Goloubinoff, 2001. Chemical chaperones regulate molecular chaperones *in-vitro* and in cells under combined salt and heat stresses. Journal of Biological Chemistry 276**:** 39586-39591.

Dolhi, J., D. Maxwell and R. Morgan-Kiss, 2013. The Antarctic *Chlamydomonas raudensis*: an emerging model for cold adaptation of photosynthesis. Extremophiles 17**:** 711-722.

Eckardt, F. D., K. Soderberg, L. J. Coop, A. A. Muller, K. J. Vickery *et al.*, 2013. The nature of moisture at Gobabeb, in the central Namib Desert. Journal of Arid Environments 93: 7-19.

Ekkers, D. M., M. S. Cretoiu, A. M. Kielak and J. D. Elsas, 2012. The great screen anomaly—a new frontier in product discovery through functional metagenomics. Applied Microbiology and Biotechnology 93**:** 1005-1020.

Empadinhas, N., and M. S. da Costa, 2006. Diversity and biosynthesis of compatible solutes in hyper/thermophiles. International Microbiology 9**:** 199-206.

Entcheva, P., W. Liebl, A. Johann, T. Hartsch and W. R. Streit, 2001. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. Applied and Environmental Microbiology 67**:** 89-99.

Fernandez-Espla, M. D., and F. Rul, 1999. PepS from *Streptococcus thermophilus*. European Journal of Biochemistry 263**:** 502-510.

Ferreira, A. C., M. F. Nobre, E. Moore, F. A. Rainey, J. R. Battista *et al.*, 1999. Characterization and radiation resistance of new isolates of *Rubrobacter radiotolerans* and *Rubrobacter xylanophilus*. Extremophiles 3**:** 235-238.

Ferrera, I., and A.-L. Reysenbach, 2001. Thermophiles in *eLS*. John Wiley & Sons, Ltd. doi: 10.1002/9780470015902.a0000406.

Forsberg, C. W., K. J. Cheng and B. A. White, 1997. Polysaccharide degradation in the rumen and large intestine, pp. 319-379 in *Gastrointestinal Microbiology*, edited by R. Mackie and B. White. Springer US.

França, M. B., A. D. Panek and E. C. A. Eleutherio, 2007. Oxidative stress and its effects during dehydration. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 146**:** 621-631.

Fredrickson, J. K., S.-m. W. Li, E. K. Gaidamakova, V. Y. Matrosova, M. Zhai *et al.*, 2008. Protein oxidation: key to bacterial desiccation resistance? The ISME Journal 2**:** 393-403.

Gabor, E. M., W. B. L. Alkema and D. B. Janssen, 2004. Quantifying the accessibility of the metagenome by random expression cloning techniques. Environmental Microbiology 6**:** 879-886.

Gaillard, J. L., P. Berche and P. Sansonetti, 1986. Transposon mutagenesis as a tool to study the role of hemolysin in the virulence of *Listeria monocytogenes*. Infection and Immunity 52**:** 50-55.

Gao, Q., and F. Garcia-Pichel, 2011. Microbial ultraviolet sunscreens. Nature Reviews Microbiology 9: 791-802.

Garcia-Pichel, F., and R. W. Castenholz, 1991. Characterization and biological implications of scytonemin, a Cyanobacterial sheath pigment. Journal of Phycology 27**:** 395-409.

Garcia-Pichel, F., and R. W. Castenholz, 1993. Occurrence of UV-absorbing, mycosporine-like compounds among Cyanobacterial isolates and an estimate of their screening capacity. Applied and Environmental Microbiology 59**:** 163-169.

Garcia-Pichel, F., A. López-Cortés and U. Nübel, 2001. Phylogenetic and morphological diversity of Cyanobacteria in soil desert crusts from the Colorado Plateau. Applied and Environmental Microbiology 67**:** 1902-1910.

Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins *et al.*, 2005. Protein identification and analysis tools on the ExPASy server, pp. 571-607 in *The Proteomics Protocol Handbook*, edited by J. M. Walker. Humana Press.

Gentry, D. R., and M. Cashel, 1996. Mutational analysis of the *Escherichia coli spoT* gene identifies distinct but overlapping regions involved in ppGpp synthesis and degradation. Molecular Microbiology 19**:** 1373-1384.

Gerday, C., 2011. Life at the extremes of temperature, pp. 425 - 444 in *Bacterial Stress Responses*, edited by G. Storz and R. Hengge. American Society for Microbiology Press.

Givens, R. M., M.-H. Lin, D. J. Taylor, U. Mechold, J. O. Berry *et al.*, 2004. Inducible expression, enzymatic activity, and origin of higher plant homologues of bacterial RelA/SpoT stress proteins in *Nicotiana tabacum*. Journal of Biological Chemistry 279**:** 7495-7504.

Goldberg, M., T. Pribyl, S. Juhnke and D. H. Nies, 1999. Energetics and topology of CzcA, a cation/proton antiporter of the resistance-nodulation-cell division protein family. Journal of Biological Chemistry 274**:** 26065-26070.

Goldstein, R. A., 2007. Amino-acid interactions in psychrophiles, mesophiles, thermophiles, and hyperthermophiles: insights from the quasi-chemical approximation. Protein Science 16**:** 1887-1895.

Gonzales, T., and J. Robert-Baudouy, 1996. Bacterial aminopeptidases: properties and functions. FEMS Microbiology Reviews 18**:** 319-344.

Gottschal, J. C., and R. A. Prins, 1991. Thermophiles: a life at elevated temperatures. Trends in Ecology & Evolution 6**:** 157-162.

Green, M. R., and J. Sambrook, 2012. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Gundlapally, S. R., and F. Garcia-Pichel, 2006. The community and phylogenetic diversity of biological soil crusts in the Colorado Plateau studied by molecular fingerprinting and intensive cultivation. Microbial Ecology 52**:** 345-357.

Gupta, R., Q. Beg and P. Lorenz, 2002. Bacterial alkaline proteases: molecular approaches and industrial applications. Applied Microbiology and Biotechnology 59**:** 15-32.

Gustafsson, C., S. Govindarajan and J. Minshull, 2004. Codon bias and heterologous protein expression. Trends in Biotechnology 22**:** 346-353.

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

Hamada, A., T. Hibino, T. Nakamura and T. Takabe, 2001. Na^+/H^+ antiporter from *Synechocystis* species PCC 6803, homologous to SOS1, contains an aspartic residue and long C-terminal tail important for the carrier activity. Plant Physiology 125**:** 437-446.

Handelsman, J., 2011. Metamicrobiology: analyzing microbial behavior at the community level, pp. 419 - 424 in *Bacterial Stress Responses*, edited by G. Storz and R. Hengge. American Society for Microbiology Press.

Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy and R. M. Goodman, 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chemistry & Biology 5**:** R245-R249.

Hannig, G., and S. C. Makrides, 1998. Strategies for optimizing heterologous protein expression in *Escherichia coli*. Trends in Biotechnology 16**:** 54-60.

Harmsen, H. J. M., B. L. M. Van Kuijk, C. M. Plugge, A. D. L. Akkermans, W. M. De Vos *et al.*, 1998. *Syntrophobacter fumaroxidans* sp. nov., a syntrophic propionate-degrading sulfate-reducing bacterium. International Journal of Systematic Bacteriology 48**:** 1383-1387.

Harrington, C., A. Del Casale, J. Kennedy, H. Neve, B. E. Picton *et al.*, 2012. Evidence of bacteriophage-mediated horizontal transfer of bacterial 16S rRNA genes in the viral metagenome of the marine sponge *Hymeniacidon perlevis*. Microbiology 158**:** 2789-2795.

Harshman, R. B., and H. Yamazaki, 1972. MS I accumulation induced by sodium chloride. Biochemistry 11**:** 615-618.

He, P., C. Deng, B. Liu, L. Zeng, W. Zhao *et al.*, 2013. Characterization of a bifunctional enzyme with (p)ppGpp-hydrolase/synthase activity in *Leptospira interrogans*. FEMS Microbiology Letters 348**:** 133-142.

Heath, C., X. P. Hu, S. C. Cary and D. Cowan, 2009. Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from Antarctic desert soil. Applied and Environmental Microbiology 75**:** 4657-4659.

Hengge-Aronis, R., 2002. Signal transduction and regulatory mechanisms involved in control of the σ^S (RpoS) subunit of RNA polymerase. Microbiology and Molecular Biology Reviews 66**:** 373-395.

Hengge, R., 2011. The general stress response in Gram-negative bacteria, pp. 251 - 289 in *Bacterial Stress Responses*, edited by G. Storz and R. Hengge. American Society for Microbiology Press.

Hogg, T., U. Mechold, H. Malke, M. Cashel and R. Hilgenfeld, 2004. Conformational antagonism between opposing active sites in a bifunctional RelA/SpoT homolog modulates (p)ppGpp metabolism during the stringent response. Cell 117**:** 57-68.

Hu, X. P., C. Heath, M. P. Taylor, M. Tuffin and D. Cowan, 2012. A novel, extremely alkaliphilic and cold-active esterase from Antarctic desert soil. Extremophiles 16**:** 79-86.

Jaenicke, R., and R. Sterner, 2013. Life at high temperatures, pp. 337-374 in *The Prokaryotes*, edited by E. Rosenberg, E. DeLong, S. Lory, E. Stackebrandt and F. Thompson. Springer Berlin Heidelberg.

Juncker, A. S., H. Willenbrock, G. von Heijne, S. Brunak, H. Nielsen *et al.*, 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Science 12**:** 1652-1662.

Kane, J. F., 1995. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. Current Opinion in Biotechnology 6**:** 494-500.

Kane, J. F., B. N. Violand, D. F. Curran, N. R. Staten, K. L. Duffin *et al.*, 1992. Novel in-frame two codon translational hop during synthesis of bovine placental lactogen in a recombinant strain of *Escherichia coli*. Nucleic Acids Research 20**:** 6707-6712.

Karlin, S., J. Mrázek and A. M. Campbell, 1997. Compositional biases of bacterial genomes and evolutionary implications. Journal of Bacteriology 179**:** 3899-3913.

Kempf, B., and E. Bremer, 1998. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. Archives of Microbiology 170**:** 319-330.

Krämer, R., 2010. Bacterial stimulus perception and signal transduction: response to osmotic stress. The Chemical Record 10**:** 217-229.

Krisko, A., and M. Radman, 2010. Protein damage and death by radiation in *Escherichia coli* and *Deinococcus radiodurans*. Proceedings of the National Academy of Sciences 107**:** 14373-14377.

Kurland, C., and J. Gallant, 1996. Errors of heterologous protein expression. Current Opinion in Biotechnology 7**:** 489-493.

Lapidus, A., O. Chertkov, M. Nolan, S. Lucas, N. Hammon *et al.*, 2011. Genome sequence of the moderately thermophilic halophile *Flexistipes sinusarabici* strain (MAS10 T). Standards in Genomic Sciences 5**:** 86-96.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan *et al.*, 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23**:** 2947-2948.

Lemos, J. A., V. K. Lin, M. M. Nascimento, J. Abranches and R. A. Burne, 2007. Three gene products govern (p)ppGpp production by *Streptococcus mutans*. Molecular Microbiology 65**:** 1568-1581.

Lewin, A., A. Wentzel and S. Valla, 2013. Metagenomics of microbial life in extreme temperature environments. Current Opinion in Biotechnology 24**:** 516-525.

Lezin, G., Y. Kosaka, H. J. Yost, M. R. Kuehn and L. Brunelli, 2011. A one-step miniprep for the isolation of plasmid DNA and Lambda phage particles. PLoS ONE 6**:** e23457.

Lim, B., and C. A. Gross, 2011. Cellular response to heat shock and cold shock, pp. 93 - 114 in *Bacterial Stress Responses*, edited by G. Storz and R. Hengge. American Society for Microbiology Press.

Lorenz, P., and J. Eck, 2005. Metagenomics and industrial applications. Nature Reviews Microbiology 3**:** 510-516.

Los, D. A., I. Suzuki, V. V. Zinchenko and N. Murata, 2008. Stress reponses in *Synechocystis*: regulated genes and regulatory systems, pp. 117-157 in *The Cyanobacteria: Molecular Biology, Genomics and Evolution*, edited by A. Herrero and E. Flores. Caister Academic Press.

Magnusson, L. U., A. Farewell and T. Nyström, 2005. ppGpp: a global regulator in *Escherichia coli*. Trends in Microbiology 13**:** 236-242.

Majerník, A., G. Gottschalk and R. Daniel, 2001. Screening of environmental DNA libraries for the presence of genes conferring Na+(Li+)/H+antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. Journal of Bacteriology 183**:** 6645-6653.

Makhalanyane, T. P., A. Valverde, D. C. Lacap, S. B. Pointing, M. I. Tuffin *et al.*, 2013. Evidence of species recruitment and development of hot desert hypolithic communities. Environmental Microbiology Reports 5**:** 219-224.

Margheri, M. C., S. Ventura, J. Kaštovský and J. Komárek, 2008. The taxonomic validation of the Cyanobacterial genus *Halothece*. Phycologia 47**:** 477-486.

Mayer, M. P., and B. Bukau, 2005. Hsp70 chaperones: cellular functions and molecular mechanism. Cellular and Molecular Life Sciences 62**:** 670-684.

McLaggan, D., J. Naprstek, E. T. Buurman and W. Epstein, 1994. Interdependence of K+ and glutamate accumulation during osmotic adaptation of *Escherichia coli*. Journal of Biological Chemistry 269**:** 1911-1917.

Medema, M. H., K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski *et al.*, 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Research 39**:** W339-W346.

Minton, K. W., 1996. Repair of ionizing-radiation damage in the radiation resistant bacterium *Deinococcus radiodurans*. Mutation Research/DNA Repair 363**:** 1-7.

Model, P., G. Jovanovic and J. Dworkin, 1997. The *Escherichia coli* phage-shock-protein (psp) operon. Molecular Microbiology 24**:** 255-261.

Moe, P., and P. Blount, 2005. Assessment of potential stimuli for mechano-dependent gating of MscL: Effects of pressure, tension, and lipid headgroups. Biochemistry 44**:** 12239-12244.

Mohamed, Y. M., M. A. Ghazy, A. Sayed, A. Ouf, H. El-Dorry *et al.*, 2013. Isolation and characterization of a heavy metal-resistant, thermophilic esterase from a Red Sea Brine Pool. Scientific Reports 3**:** 3358.

Monciardini, P., L. Cavaletti, P. Schumann, M. Rohde and S. Donadio, 2003. *Conexibacter woesei* gen. nov., sp. nov., a novel representative of a deep evolutionary line of descent within the class Actinobacteria. International Journal of Systematic and Evolutionary Microbiology 53**:** 569-576.

Monnet, V., 2003. Bacterial oligopeptide-binding proteins. Cellular and Molecular Life Sciences 60**:** 2100-2114.

Motoshima, H., E. Minagawa, F. Tsukasaki and S. Kaminogawa, 1997. Cloning of genes of the aminopeptidase T family from *Thermus thermophilus* HB8 and *Bacillus stearothermophilus* NCIB8924: apparent similarity to the leucyl aminopeptidase family. Bioscience, Biotechnology, and Biochemistry 61**:** 1710-1717.

Muñoz-López, M., and J. L. García-Pérez, 2010. DNA transposons: nature and applications in genomics. Current Genomics 11**:** 115-128.

Nies, D. H., 2003. Efflux-mediated heavy metal resistance in prokaryotes. FEMS Microbiology Reviews 27**:** 313-339.

Nies, D. H., 2013. RND-efflux pumps for metal cations, pp. 79-121 in *Microbial Efflux Pumps: Current Research*, edited by E. W. Yu, Q. Zhang, M. H. Brown and S. Baugh. Horizon Scientific Press.

Nies, D. H., G. Rehbein, T. Hoffmann, C. Baumann and C. Grosse, 2006. Paralogs of genes encoding metal resistance proteins in *Cupriavidus metallidurans* strain CH34. Journal of Molecular Microbiology and Biotechnology 11**:** 82-93.

Nilsen, R. K., T. Torsvik and T. Lien, 1996. *Desulfotomaculum thermocisternum* sp. nov., a sulfate reducer isolated from a hot North Sea oil reservoir. International Journal of Systematic Bacteriology 46**:** 397-402.

Nishida, S., and G. Nakagawara, 1964. Isolation of toxigenic strains of *Clostridium novyi* from soil. Journal of Bacteriology 88**:** 1636-1640.

Ochi, K., T. Nishizawa, T. Inaoka, A. Yamada, K. Hashimoto *et al.*, 2012. Heterologous expression of a plant RelA-SpoT homologue results in increased stress tolerance in *Saccharomyces cerevisiae* by accumulation of the bacterial alarmone ppGpp. Microbiology 158**:** 2213-2224.

Overbeek, R., R. Olson, G. D. Pusch, G. J. Olsen, J. J. Davis *et al.*, 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Research 42**:** D206-D214.

Padan, E., E. Bibi, M. Ito and T. A. Krulwich, 2005. Alkaline pH homeostasis in bacteria: new insights. Biochimica et Biophysica Acta - Biomembranes 1717**:** 67-88.

Padan, E., and T. A. Krulwich, 2000. Sodium stress, pp. 117-130 in *Bacterial Stress Responses*, edited by G. Storz and R. Hengge. American Society for Microbiology Press.

Padan, E., M. Venturi, Y. Gerchman and N. Dover, 2001. Na^+/H^+ antiporters. Biochimica et Biophysica Acta - Bioenergetics 1505**:** 144-157.

Papadimitriou, K., H. Pratsinis, G. Nebe-von-Caron, D. Kletsas and E. Tsakalidou, 2007. Acid tolerance of *Streptococcus macedonicus* as assessed by flow cytometry and single-cell sorting. Applied and Environmental Microbiology 73**:** 465-476.

Petersen, T., S. Brunak, G. von Heijne and H. Nielsen, 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 8**:** 785-786.

Pleitner, A., Y. Zhai, R. Winter, L. Ruan, L. M. McMullen *et al.*, 2012. Compatible solutes contribute to heat resistance and ribosome stability in *Escherichia coli* AW1.7. Biochimica et Biophysica Acta - Proteins and Proteomics 1824**:** 1351-1357.

Podar, M., and A.-L. Reysenbach, 2006. New opportunities revealed by biotechnological explorations of extremophiles. Current Opinion in Biotechnology 17**:** 250-255.

Pointing, S. B., and J. Belnap, 2012. Microbial colonization and controls in dryland systems. Nature Reviews Microbiology 10**:** 551-562.

Potrykus, K., and M. Cashel, 2008. (p)ppGpp: still magical? Annual Review of Microbiology 62**:** 35-51.

Potts, M., 1994. Desiccation tolerance of prokaryotes. Microbiological Reviews 58**:** 755-805.

Poxton, I. R., L. C. Hunter, R. Brown, H. G. Lough and J. K. Miller, 1997. Clostridia and equine grass sickness. Reviews in Medical Microbiology 8**:** S52.

Prestel, E., S. Salamitou and M. S. DuBow, 2008. An examination of the bacteriophages and bacteria of the Namib Desert. Journal of Microbiology 46**:** 364-372.

Pugsley, A. P., 1993. The complete general secretory pathway in Gram-negative bacteria. Microbiology Reviews 57**:** 50 - 108.

Rainey, F. A., K. Ray, M. Ferreira, B. Z. Gatz, M. F. Nobre *et al.*, 2005. Extensive diversity of ionizing-radiation-resistant bacteria recovered from Sonoran Desert soil and description of nine new species of the genus *Deinococcus* obtained from a single soil sample. Applied and Environmental Microbiology 71**:** 5225-5235.

Rajeev, L., U. N. da Rocha, N. Klitgord, E. G. Luning, J. Fortney *et al.*, 2013. Dynamic Cyanobacterial response to hydration and dehydration in a desert biological soil crust. The ISME Journal 7: 2178-2191.

Rao, M. B., A. M. Tanksale, M. S. Ghatge and V. V. Deshpande, 1998. Molecular and biotechnological aspects of microbial proteases. Microbiology and Molecular Biology Reviews 62**:** 597-635.

Reed, C. J., H. Lewis, E. Trejo, V. Winston and C. Evilia, 2013. Protein adaptations in archaeal extremophiles. Archaea 2013**:** 14.

Richter, K., M. Haslbeck and J. Buchner, 2010. The heat shock response: life on the verge of death. Molecular Cell 40**:** 253-266.

Rocha, E. R., and C. J. Smith, 1999. Role of the alkyl hydroperoxide reductase (ahpCF) gene in oxidative stress defense of the obligate anaerobe *Bacteroides fragilis*. Journal of Bacteriology 181**:** 5701-5710.

Rodrigues, D. F., and J. M. Tiedje, 2008. Coping with our cold planet. Applied and Environmental Microbiology 74**:** 1677-1686.

Rodríguez-Rosales, M. P., F. J. Gálvez, R. Huertas, M. N. Aranda, M. Baghour *et al.*, 2009. Plant NHX cation/proton antiporters. Plant Signaling & Behavior 4**:** 265-276.

Rodríguez, A. C., and D. Stock, 2002. Crystal structure of reverse gyrase: insights into the positive supercoiling of DNA. The EMBO Journal 21**:** 418-426.

Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman *et al.*, 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Applied and Environmental Microbiology 66**:** 2541-2547.

Rothschild, L. J., and R. L. Mancinelli, 2001. Life in extreme environments. Nature 409**:** 1092-1101.

Saier, M. H., 2000. A functional-phylogenetic classification system for transmembrane solute transporters. Microbiology and Molecular Biology Reviews 64**:** 354-411.

Saier, M. H., V. S. Reddy, D. G. Tamang and Å. Västermark, 2014. The transporter classification database. Nucleic Acids Research 42**:** D251-D258.

Salzberg, S., A. Delcher, S. Kasif and O. White, 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Research 26**:** 544-548.

Sambrook, J., E. Fritsch and T. Maniatis, 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sarethy, I. P., Y. Saxena, A. Kapoor, M. Sharma, S. K. Sharma *et al.*, 2011. Alkaliphilic bacteria: applications in industrial biotechnology. Journal of Industrial Microbiology & Biotechnology 38**:** 769-790.

Schlesner, H., C. Rensmann, B. J. Tindall, D. Gade, R. Rabus *et al.*, 2004. Taxonomic heterogeneity within the Planctomycetales as derived by DNA–DNA hybridization, description of *Rhodopirellula baltica* gen. nov., sp. nov., transfer of *Pirellula marina* to the genus *Blastopirellula* gen. nov. as *Blastopirellula marina* comb. nov. and emended description of the genus *Pirellula*. International Journal of Systematic and Evolutionary Microbiology 54**:** 1567-1580.

Schmieder, R., and R. Edwards, 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27**:** 863-864.

Schultz, J., F. Milpetz, P. Bork and C. P. Ponting, 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. Proceedings of the National Academy of Sciences 95**:** 5857-5864.

Selzer, J., F. Hofmann, G. Rex, M. Wilm, M. Mann *et al.*, 1996. *Clostridium novyi* α-toxin-catalyzed incorporation of GlcNAc into Rho subfamily proteins. Journal of Biological Chemistry 271**:** 25173-25177.

Shabala, L., J. Bowman, J. Brown, T. Ross, T. McMeekin *et al.*, 2009. Ion transport and osmotic adjustment in *Escherichia coli* in response to ionic and non-ionic osmotica. Environmental Microbiology 11**:** 137-148.

Shi, H., B.-h. Lee, S.-J. Wu and J.-K. Zhu, 2003. Overexpression of a plasma membrane Na^+/H^+ antiporter gene improves salt tolerance in *Arabidopsis thaliana*. Nature Biotechnology 21**:** 81-85.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus *et al.*, 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7**:** 539-544.

Simon, C., and R. Daniel, 2009. Achievements and new knowledge unraveled by metagenomic approaches. Applied Microbiology and Biotechnology 85**:** 265-276.

Simon, C., J. Herath, S. Rockstroh and R. Daniel, 2009. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. Applied and Environmental Microbiology 75**:** 2964-2968.

Slade, D., and M. Radman, 2011. Oxidative stress resistance in *Deinococcus radiodurans*. Microbiology and Molecular Biology Reviews 75**:** 133-191.

Sleator, R. D., and C. Hill, 2002. Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. FEMS Microbiology Reviews 26**:** 49-71.

Solovyev, V., and A. Salamov, 2011. Automatic annotation of microbial genomes and metagenomic sequences, pp. 61-78 in *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*, edited by R. W. Li. Nova Science Publishers.

Sonne-Hansen, J., and B. K. Ahring, 1999. *Thermodesulfobacterium hveragerdense* sp. nov., and *Thermodesulfovibrio islandicus* sp. nov., two thermophilic sulfate reducing bacteria isolated from a Icelandic hot spring. Systematic and Applied Microbiology 22**:** 559-564.

Sørensen, H. P., and K. K. Mortensen, 2005. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. Journal of Biotechnology 115**:** 113-128.

Sørensen, H. P., H. U. Sperling-Petersen and K. K. Mortensen, 2003. Production of recombinant thermostable proteins expressed in *Escherichia coli*: completion of protein synthesis is the bottleneck. Journal of Chromatography B 786**:** 207-214.

Sotgia, F., U. Martinez-Outschoorn and M. Lisanti, 2011. Mitochondrial oxidative stress drives tumor progression and metastasis: should we use antioxidants as a key component of cancer treatment and prevention? BMC Medicine 9**:** 62-66.

Stolp, H., and M. P. Starr, 1963. *Bdellovibrio bacteriovorus* gen. et sp. n., a predatory, ectoparasitic, and bacteriolytic microorganism. Antonie van Leeuwenhoek 29**:** 217-248.

Stomeo, F., A. Valverde, S. Pointing, C. McKay, K. Warren-Rhodes *et al.*, 2013. Hypolithic and soil microbial community assembly along an aridity gradient in the Namib Desert. Extremophiles 17**:** 329-337.

Storz, G., and R. Hengge, 2011. *Bacterial Stress Responses (2nd Edition)*. American Society for Microbiology Press.

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

Taupp, M., K. Mewis and S. J. Hallam, 2011. The art and design of functional metagenomic screens. Current Opinion in Biotechnology 22**:** 465-472.

Terron-Gonzalez, L., C. Medina, M. C. Limon-Mortes and E. Santero, 2013. Heterologous viral expression systems in fosmid vectors increase the functional analysis potential of metagenomic libraries. Scientific Reports 3**:** 1107-1114.

Tracy, B. P., S. M. Gaida and E. T. Papoutsakis, 2010. Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. Current Opinion in Biotechnology 21**:** 85-99.

Trivedi, S., S. R. Rao and H. S. Gehlot, 2005. Nucleic acid stability in thermophilic prokaryotes: a review. Journal of Cell and Molecular Biology 4**:** 61-69.

Troeschel, S. C., S. Thies, O. Link, C. I. Real, K. Knops *et al.*, 2012. Novel broad host range shuttle vectors for expression in *Escherichia coli, Bacillus subtilis* and *Pseudomonas putida*. Journal of Biotechnology 161**:** 71-79.

Tseng, T.-T., K. S. Gratwick, J. Kollman, D. Park, D. H. Nies *et al.*, 1999. The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. Journal of Molecular Microbiology and Biotechnology 1**:** 107 - 125.

Tuffin, M., D. Anderson, C. Heath and D. A. Cowan, 2009. Metagenomic gene discovery: how far have we moved into novel sequence space? Biotechnology Journal 4**:** 1671-1683.

Turgay, K., 2011. Role of proteolysis and chaperones in stress response and regulation, pp. 75 - 90 in *Bacterial Stress Responses*, edited by G. Storz and R. Hengge. American Society for Microbiology Press.

Uchiyama, T., T. Abe, T. Ikemura and K. Watanabe, 2005. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nature Biotechnology 23**:** 88-93.

Uchiyama, T., and K. Miyazaki, 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. Current Opinion in Biotechnology 20**:** 616-622.

Uchiyama, T., and K. Miyazaki, 2010. Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. Applied and Environmental Microbiology 76**:** 7029–7035.

Ullman, I., and B. Budel, 2003. Biological soil crusts of Africa, pp. 107 - 118 in *Biological Soil Crusts: Structure, Function and Management*, edited by J. Belnap and O. L. Lange. Springer-Verlag.

Ulusu, N. N., and E. F. Tezcan, 2001. Cold shock proteins. Turkish Journal of Medical Science 31**:** 283-290.

Vabulas, R. M., S. Raychaudhuri, M. Hayer-Hartl and F. U. Hartl, 2010. Protein folding in the cytoplasm and the heat shock response. Cold Spring Harbor Perspectives in Biology 2**:** a004390.

van der Biezen, E. A., J. Sun, M. J. Coleman, M. J. Bibb and J. D. G. Jones, 2000. Arabidopsis RelA/SpoT homologs implicate (p)ppGpp in plant signaling. Proceedings of the National Academy of Sciences 97**:** 3747-3752.

van der Heide, T., and B. Poolman, 2002. ABC transporters: one, two or four extracytoplasmic substrate-binding sites? EMBO Reports 3**:** 938-943.

Varin, T., C. Lovejoy, A. D. Jungblut, W. F. Vincent and J. Corbeil, 2012. Metagenomic analysis of stress genes in microbial mat communities from Antarctica and the High Arctic. Applied and Environmental Microbiology 78**:** 549-559.

Ventosa, A., J. J. Nieto and A. Oren, 1998. Biology of moderately halophilic aerobic bacteria. Microbiology and Molecular Biology Reviews 62**:** 504-544.

Verde, C., G. di Prisco, D. Giordano, R. Russo, D. Anderson *et al.*, 2012. Antarctic psychrophiles: models for understanding the molecular basis of survival at low temperature and responses to climate change. Biodiversity 13**:** 249-256.

Waditee, R., T. Hibino, Y. Tanaka, T. Nakamura, A. Incharoensakdi *et al.*, 2001. Halotolerant Cyanobacterium *Aphanothece halophytica c*ontains a Na^+/H^+ antiporter, homologous to eukaryotic ones, with novel ion specificity affected by C-terminal tail. Journal of Biological Chemistry 276**:** 36931-36938.

Wang, G.-Y.-S., E. Graziani, B. Waters, W. Pan, X. Li *et al.*, 2000. Novel natural products from soil DNA libraries in a *Streptomycete* host. Organic Letters 2**:** 2401-2404.

Wang, X., F. Xu and S. Chen, 2013. Metagenomic cloning and characterization of Na^+/H^+ antiporter genes taken from sediments in Chaerhan Salt Lake in China. Biotechnology Letters 35**:** 619-624.

Warren-Rhodes, K., K. Rhodes, S. Pointing, S. Ewing, D. Lacap *et al.*, 2006. Hypolithic Cyanobacteria, dry limit of photosynthesis, and microbial ecology in the hyperarid Atacama Desert. Microbial Ecology 52**:** 389-398.

Warren-Rhodes, K. A., C. P. McKay, L. N. Boyle, M. R. Wing, E. M. Kiekebusch *et al.*, 2013. Physical ecology of hypolithic communities in the central Namib Desert: the role of fog, rain, rock habitat and light. Journal of Geophysical Research: Biogeosciences 118**:** 1451-1460.

Weber, H., T. Polen, J. Heuveling, V. F. Wendisch and R. Hengge, 2005. Genome-wide analysis of the general stress response network in *Escherichia coli*: σ^S-dependent genes, promoters, and sigma factor selectivity. Journal of Bacteriology 187**:** 1591-1603.

Welsh, D. T., and R. A. Herbert, 1999. Osmotically induced intracellular trehalose, but not glycine betaine accumulation promotes desiccation tolerance in *Escherichia coli*. FEMS Microbiology Letters 174**:** 57-63.

Wendrich, T. M., G. Blaha, D. N. Wilson, M. A. Marahiel and K. H. Nierhaus, 2002. Dissection of the mechanism for the stringent factor RelA. Molecular Cell 10**:** 779-788.

Williamson, L. L., B. R. Borlee, P. D. Schloss, C. Guan, H. K. Allen *et al.*, 2005. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. Applied and Environmental Microbiology 71**:** 6335-6344.

Woebken, D., H. Teeling, P. Wecker, A. Dumitriu, I. Kostadinov *et al.*, 2007. Fosmids of novel marine Planctomycetes from the Namibian and Oregon coast upwelling systems and their cross-comparison with planctomycete genomes. The ISME Journal 1**:** 419-435.

Wood, J. M., 2011. Osmotic stress, pp. 133 - 156 in *Bacterial Stress Responses*, edited by G. Storz and R. Hengge. American Society for Microbiology Press.

Wu, D., G. Jospin and J. A. Eisen, 2013. Systematic identification of gene families for use as "markers" for phylogenetic and phylogeny-driven ecological studies of bacteria and archaea and their major subgroups. PLoS ONE 8**:** e77033.

Wu, D., M. Wu, A. Halpern, D. B. Rusch, S. Yooseph *et al.*, 2011. Stalking the fourth domain in metagenomic data: searching for, discovering, and interpreting novel, deep branches in marker gene phylogenetic trees. PLoS ONE 6**:** e18011.

Yakovchuk, P., E. Protozanova and M. D. Frank-Kamenetskii, 2006. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. Nucleic Acids Research 34**:** 564-574.

Yamada, A., K. Tsutsumi, S. Tanimoto and Y. Ozeki, 2003. Plant RelA/SpoT homolog confers salt tolerance in *Escherichia coli* and *Saccharomyces cerevisiae*. Plant and Cell Physiology 44**:** 3-9.

Yang, L. F., J. Q. Jiang, B. S. Zhao, B. Zhang, D. Q. Feng *et al.*, 2006. A Na^+/H^+ antiporter gene of the moderately halophilic bacterium *Halobacillus dabanensis* D-8T: cloning and molecular characterization. FEMS Microbiology Letters 255**:** 89-95.

Yeager, C. M., J. L. Kornosky, D. C. Housman, E. E. Grote, J. Belnap *et al.*, 2004. Diazotrophic community structure and function in two successional stages of biological soil crusts from the Colorado Plateau and Chihuahuan Desert. Applied and Environmental Microbiology 70**:** 973-983.

Yun, J., and S. Ryu, 2005. Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. Microbial Cell Factories 4**:** 8-12.

Zhang, H., Z. Wang, L. Wang, R. Mu, Z. Zou *et al.*, 2014. Cloning and identification of a novel NhaD-type Na^+/H^+ antiporter from metagenomic DNA of the halophilic bacteria in soil samples around Daban Salt Lake. Extremophiles 18**:** 89-98.

Zhong, N.-Q., L.-B. Han, X.-M. Wu, L.-L. Wang, F. Wang *et al.*, 2012. Ectopic expression of a bacterium NhaD-type Na^+/H^+ antiporter leads to increased tolerance to combined salt/alkali stresses. Journal of Integrative Plant Biology 54**:** 412-421.

Zhu, W., A. Lomsadze and M. Borodovsky, 2010. *Ab initio* gene identification in metagenomic sequences. Nucleic Acids Research 38**:** e132.

Appendix A

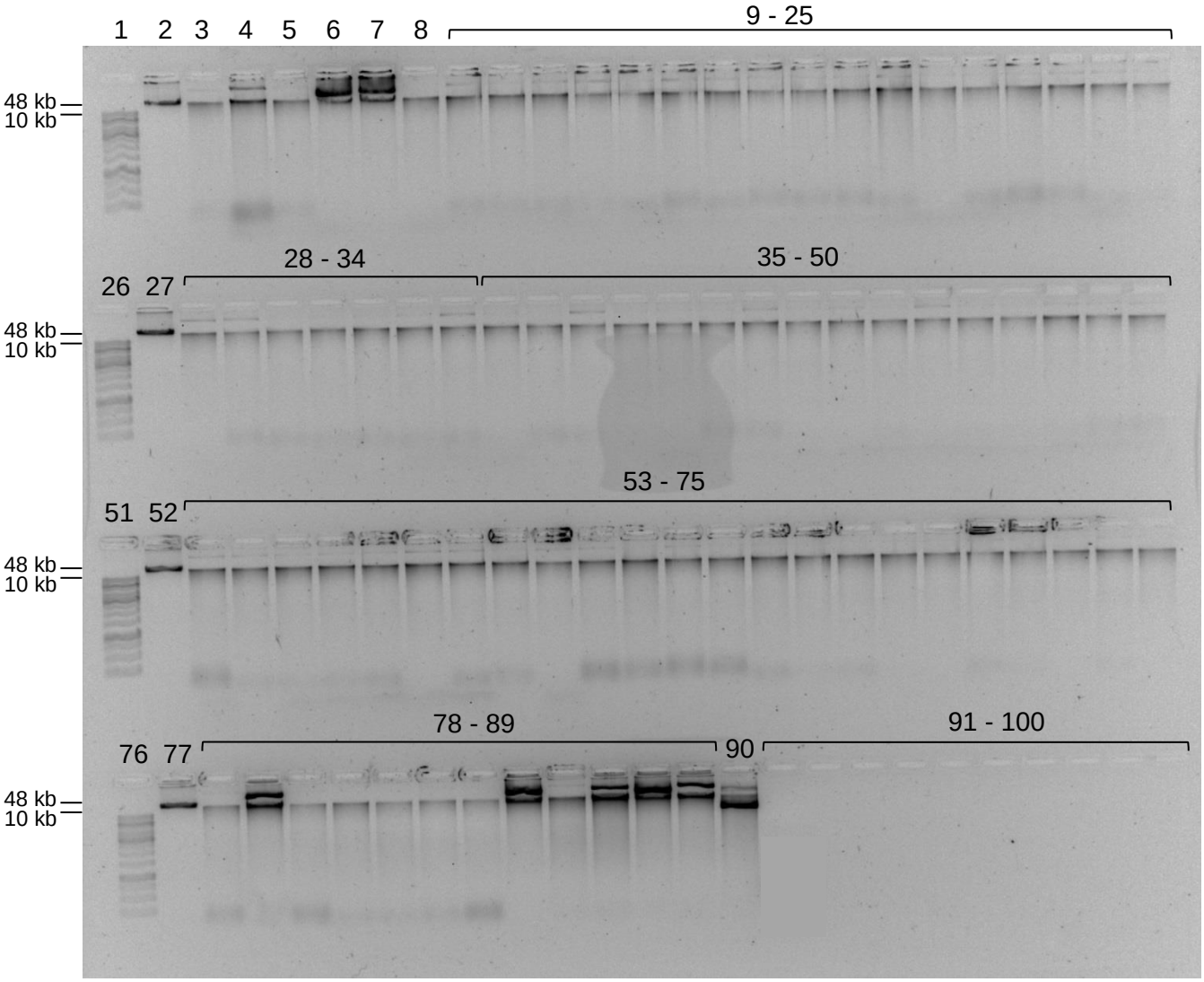


Figure A1. Extraction of fosmid DNA from STCs identified from the Namib Desert soil metagenomic library. Lanes 1, 26, 51 and 76: GeneRuler 1 kb DNA Ladder (Fermentas). Lanes 2, 27, 52 and 77: Lambda DNA (50 ng/µL). Lane 3: STC 1. Lanes 4 – 8: STC 9 – 13. Lanes 9 – 25: STC 16 – 32. Lanes 28 – 34: STC 33 – 39. Lanes 35 – 50: STC 42 – 57. Lanes 53 – 75: STC 58 – 80. Lanes 78 – 89: STC 81 – 92. Lane 90: pCC1FOS. Lanes 91 – 100: Empty.

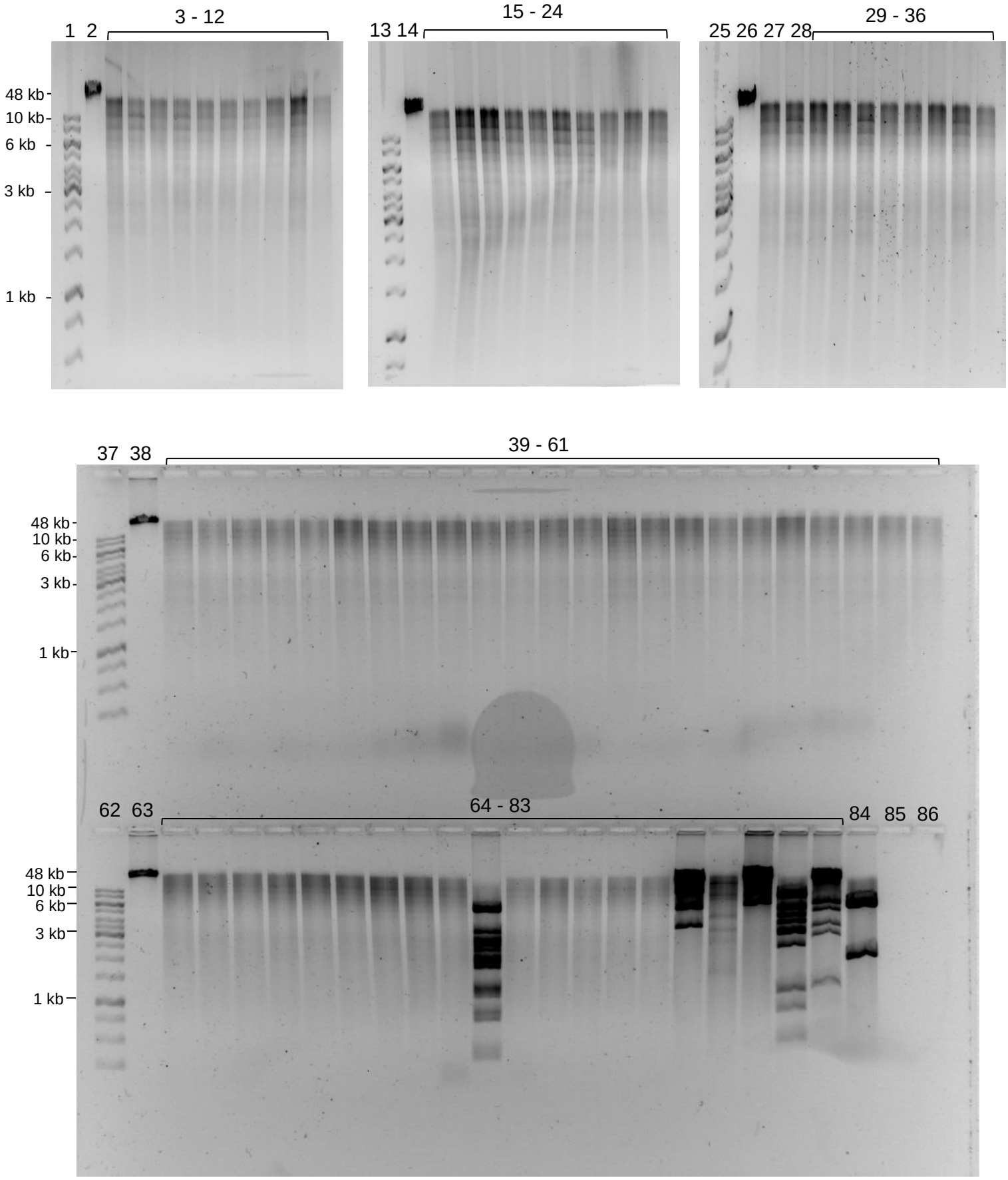


Figure A2. *Nco*I restriction digestion analysis of STC and pCC1FOS DNA. Lanes 1, 13, 25, 37 and 62: GeneRuler 1 kb DNA Ladder (Fermentas). Lanes 2, 14, 26, 38 and 63: Lambda DNA (50 ng/µL). Lanes 3 – 12: STC 18 – 27. Lanes 15 – 24: STC 28 – 37. Lanes 27 and 28: STC 38 and 39. Lanes 29 – 36: STC 42 – 49. Lanes 39 – 61: STC 50 – 72. Lanes 64 – 83: STC 73 – 92. Lane 84: pCC1FOS. Lanes 85 and 86: Empty.

Appendix B

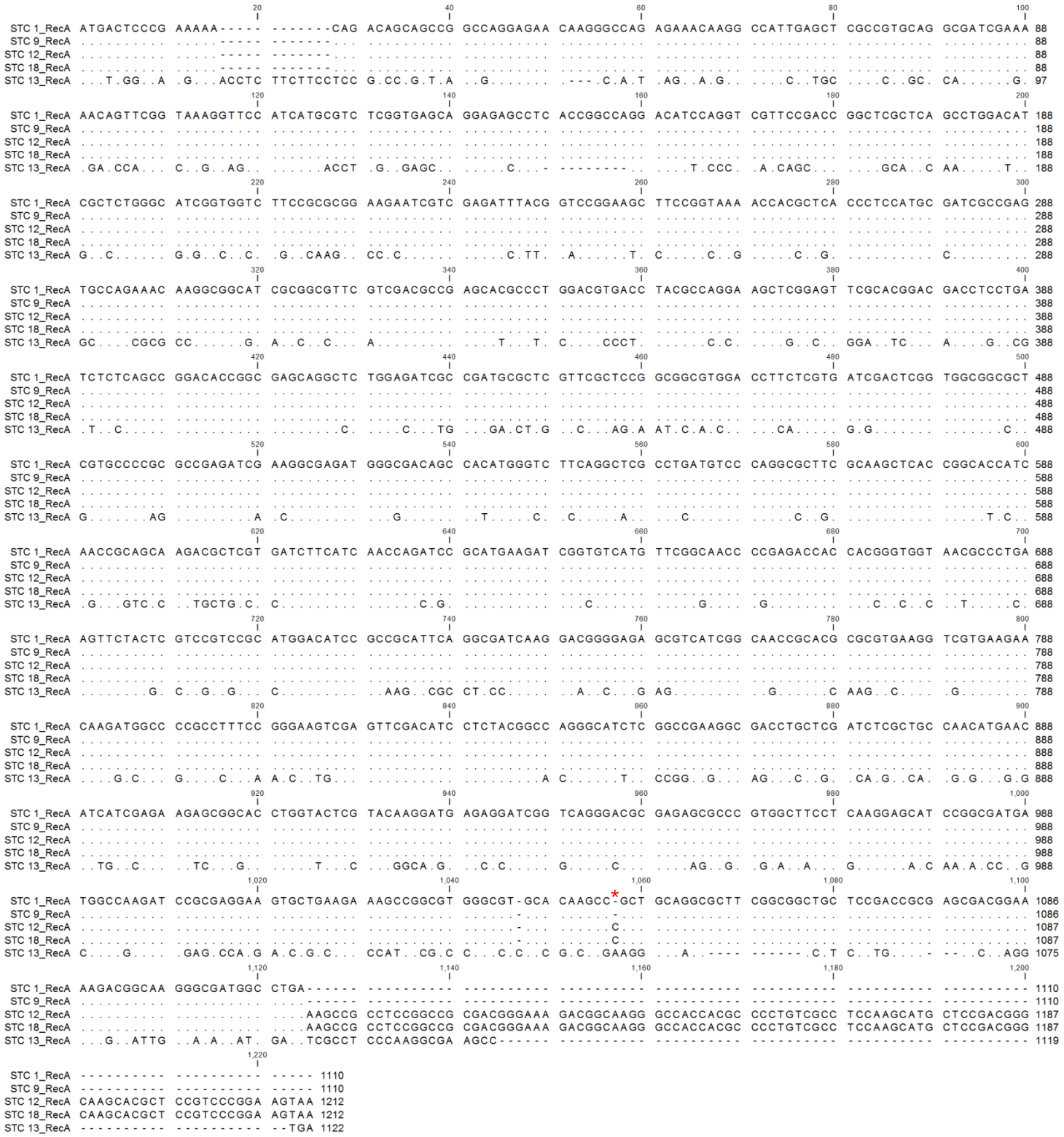


Figure B1. Sequence alignment of RecA genes from STC 1, 9, 12 and 18 showing the high identity of these sequences, using CLC Main Workbench. The STC 13 RecA gene is included to show the dissimilarity between it and the first four highly similar sequences. The alignment is dot-matched, where STC 1_RecA is used as a reference for the other sequences. A dot indicates an identical nucleotide to the reference sequence. The insertion of a cytosine (C) in the STC 12 and 18 RecA sequences is indicated by the red asterisks.

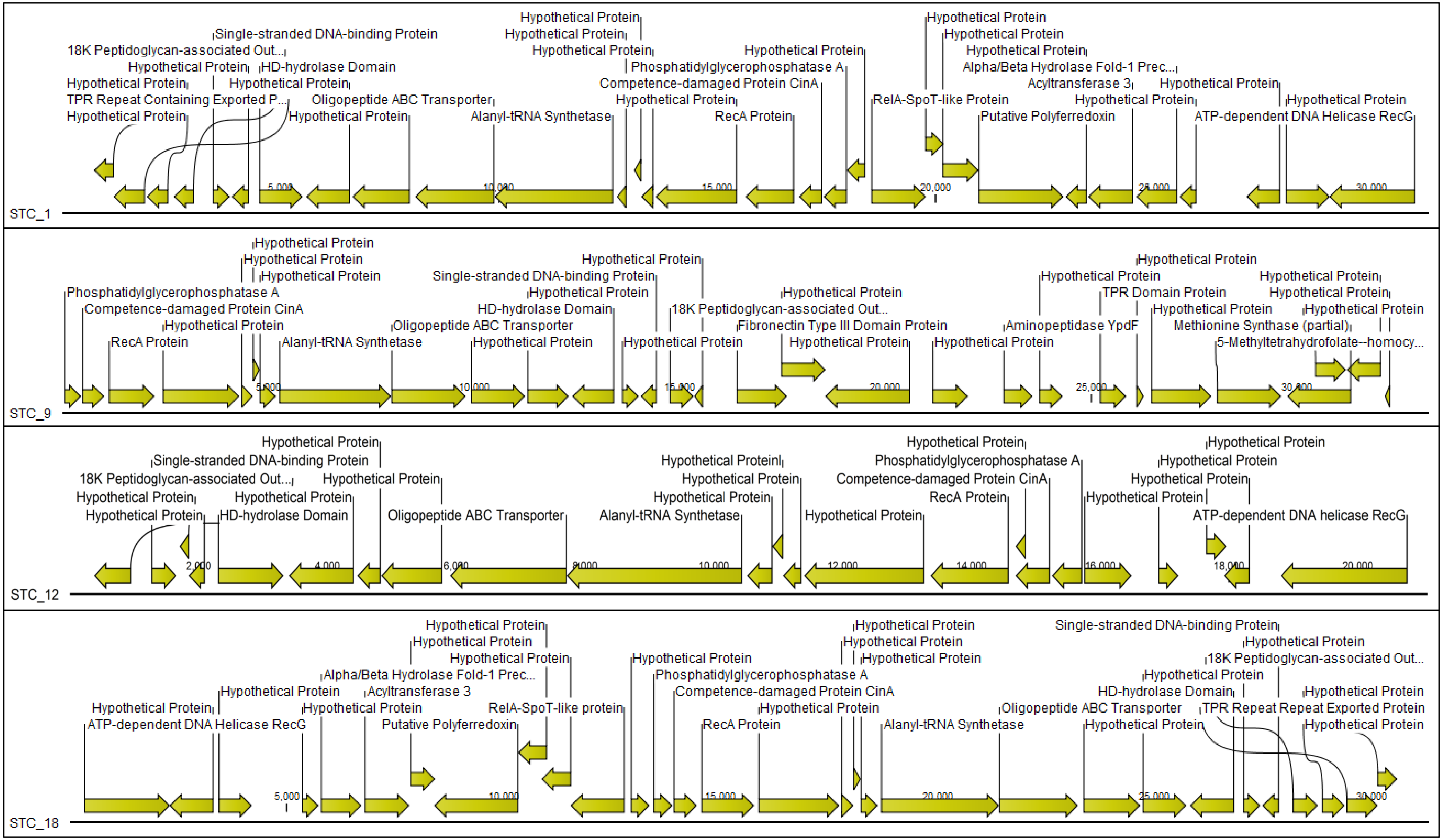


Figure B2. Graphic representation of the genetic organisation of the metagenomic inserts of STC 1, 9, 12 and 18.

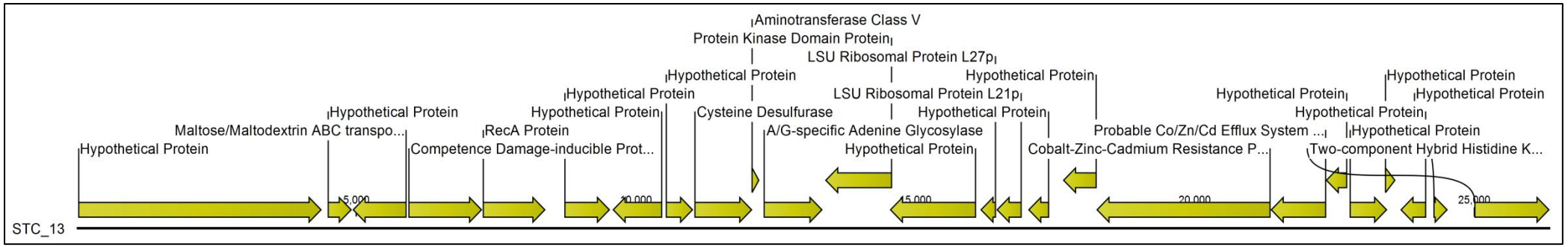


Figure B3. Graphic representation of the genetic organisation of the metagenomic insert of STC 13.

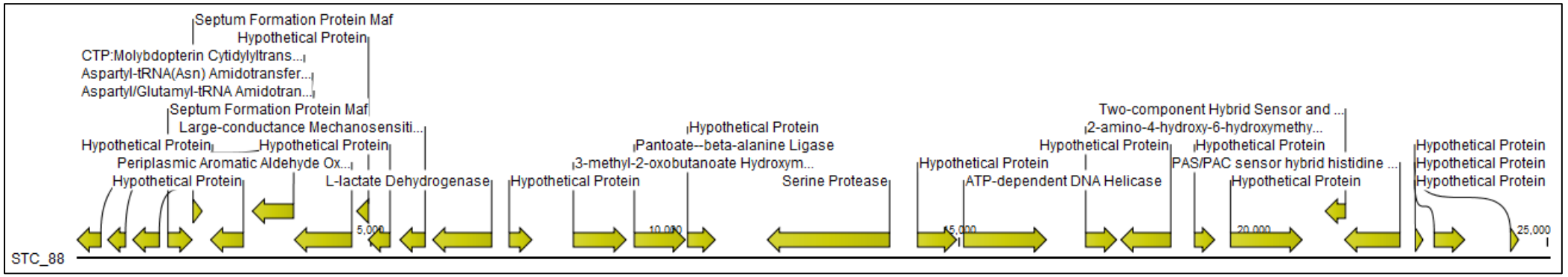


Figure B4. Graphic representation of the genetic organisation of the metagenomic insert of STC 88.

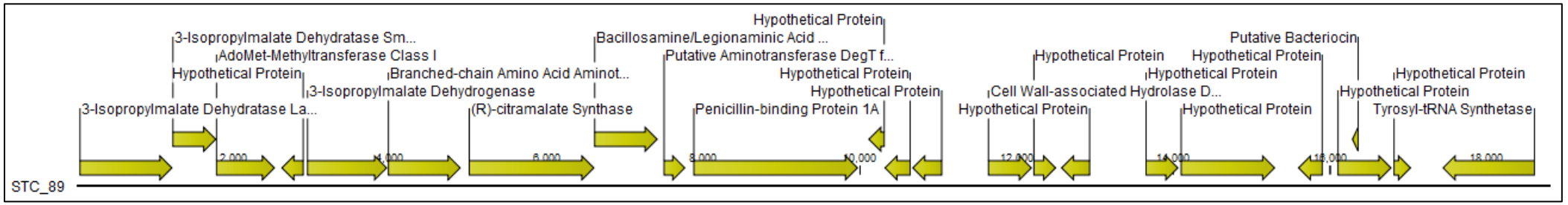


Figure B5. Graphic representation of the genetic organisation of the metagenomic insert of STC 89.

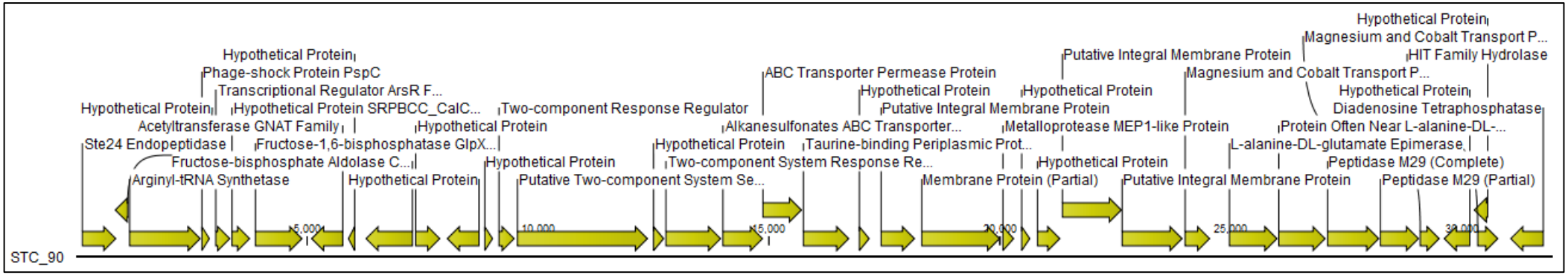


Figure B6. Graphic representation of the genetic organisation of the metagenomic insert of STC 90.

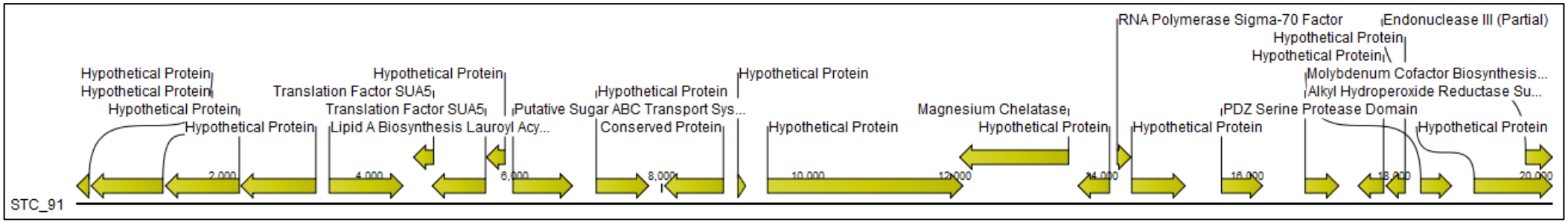


Figure B7. Graphic representation of the genetic organisation of the metagenomic insert of STC 91.

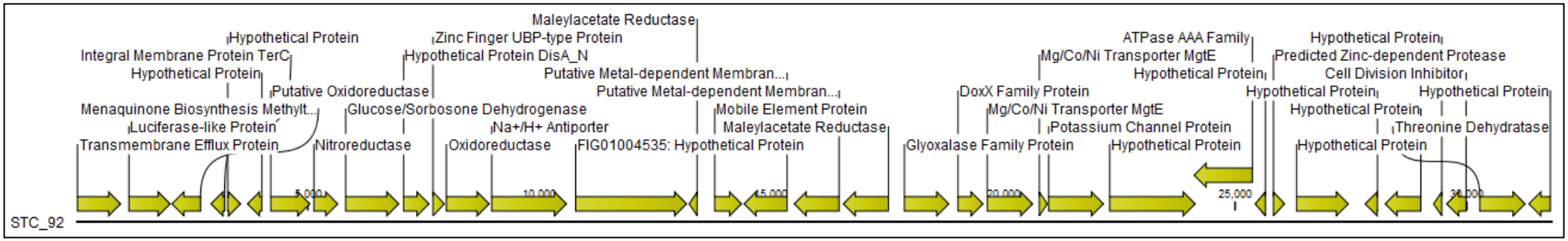


Figure B8. Graphic representation of the genetic organisation of the metagenomic insert of STC 92.

Table B1. List of genes present on the metagenomic insert of STC 1.

Start	Stop	Strand	Gene/ Function	KEGG/ TCDB*
1182	727	-	Hypothetical protein	
1895	1179	-	TPR repeat containing exported protein	
2428	1940	-	Hypothetical protein	
3135	2559	-	18K peptidoglycan-associated outer membrane lipoprotein; OmpA/MotB precursor	
3445	3831	+	Single-stranded DNA-binding protein	
4278	3889	-	Hypothetical protein	
4510	5496	+	HD-hydrolase domain	
6592	5591	-	Hypothetical protein	
7962	6652	-	Hypothetical protein	
9895	8090	-	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	TC 3.A.1.5.1
12621	9907	-	Alanyl-tRNA synthetase	EC 6.1.1.7
12924	12706	-	Hypothetical protein	
13266	13096	-	Hypothetical protein	
13544	13266	-	Hypothetical protein	
15452	13596	-	Hypothetical protein	
16765	15656	-	RecA protein	
17408	16878	-	Competence-damaged protein CinA	
17971	17441	-	Phosphatidylglycerophosphatase A	EC 3.1.3.27
18393	17968	-	Hypothetical protein (4-hydroxybenzoyl-CoA thioesterase (4HBT))	
18531	19778	+	Hypothetical protein (RelA-SpoT like domain)	
19765	20187	+	Hypothetical protein	

20157	20987	+	Hypothetical protein	
20984	22924	+	Putative polyferredoxin	
23470	22988	-	Hypothetical protein (Haloacid dehalogenase-like hydrolases (HAD-like superfamily))	
24530	23502	-	Acyltransferase 3	
25539	24607	-	Alpha/beta hydrolase fold-1 precursor	
25979	25596	-	Hypothetical protein	
27898	27131	-	Hypothetical protein	
28022	29023	+	Hypothetical protein	
30992	29025	-	ATP-dependent DNA helicase RecG	EC 3.6.1.-

*KEGG: Kyoto Encyclopedia of Genes and Genomes (EC); TCDB: Transporter Classification Database (TC). Red text shows genes targeted for sub-cloning. Grey shading shows the gene product used to construct the phylogenetic tree of the STC.

Table B2. List of genes present on the metagenomic insert of STC 9.

Start	Stop	Strand	Gene/ Function	KEGG/ TCDB*
55	453	+	Phosphatidylglycerophosphatase A	EC 3.1.3.27
486	1016	+	Competence-damaged protein CinA	
1129	2238	+	RecA protein	
2442	4298	+	Hypothetical protein	
4350	4628	+	Hypothetical protein	
4628	4798	+	Hypothetical protein	
4795	5184	+	Hypothetical protein	
5269	7983	+	Alanyl-tRNA synthetase	EC 6.1.1.7
7995	9800	+	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	TC 3.A.1.5.1
9928	11238	+	Hypothetical protein	
11298	12299	+	Hypothetical protein	
13410	12394	-	HD-hydrolase domain	
13597	14001	+	Hypothetical protein	
14445	14059	-	Single-stranded DNA-binding protein	
14756	15331	+	18K peptidoglycan-associated outer membrane lipoprotein; OmpA/MotB precursor	
15569	15354	-	Hypothetical protein	
16382	17605	+	Fibronectin type III domain protein	
17464	18540	+	Hypothetical protein	
20600	18537	-	Hypothetical protein	
21138	21998	+	Hypothetical protein	
22868	23569	+	Aminopeptidase YpdF (MP-, MA-, MS-, AP-, NP- specific)	

23731	24300	+	Hypothetical protein	
25207	25845	+	TPR domain protein	
26096	26278	+	Hypothetical protein	
26452	27915	+	Hypothetical protein (putatively related to sulfatases)	
28050	29612	+	5-methyltetrahydrofolate--homocysteine methyltransferase	EC 2.1.1.13
31315	29768	-	Methinonine synthase I (partial)	
30440	31186	+	Hypothetical protein	
32042	31212	-	Hypothetical protein	
32131	32265	+	Hypothetical protein	

*KEGG: Kyoto Encyclopedia of Genes and Genomes (EC); TCDB: Transporter Classification Database (TC).

Table B3. List of genes present on the metagenomic insert of STC 12.

Start	Stop	Strand	Gene/ Function	KEGG/ TCDB*
953	378	-	18K peptidoglycan-associated outer membrane lipoprotein; OmpA/MotB precursor	
1263	1649	+	Single-stranded DNA-binding protein	
1853	1707	-	Hypothetical protein	
2095	1850	-	Hypothetical protein	
2297	3313	+	HD-hydrolase domain	
4409	3408	-	Hypothetical protein	
4828	4469	-	Hypothetical protein	
5780	4839	-	Hypothetical protein	
7718	5909	-	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	TC 3.A.1.5.1
10438	7724	-	Alanyl-tRNA synthetase	EC 6.1.1.7
10912	10523	-	Hypothetical protein	
11079	10909	-	Hypothetical protein	
11357	11079	-	Hypothetical protein	
13265	11409	-	Hypothetical protein	
14579	13368	-	RecA protein	
15222	14692	-	Competence-damaged protein CinA	
15725	15255	-	Phosphatidylglycerophosphatase A	EC 3.1.3.27
15751	16485	+	Hypothetical protein	
16902	17210	+	Hypothetical protein	
17647	17958	+	Hypothetical protein	
18325	17927	-	Hypothetical protein	

20775	18808	-	ATP-dependent DNA helicase RecG	EC 3.6.1.-

*KEGG: Kyoto Encyclopedia of Genes and Genomes (EC); TCDB: Transporter Classification Database (TC).

Table B4. List of genes present on the metagenomic insert of STC 18.

Start	Stop	Strand	Gene/ Function	KEGG/ TCDB*
330	2297	+	ATP-dependent DNA helicase RecG	EC 3.6.1.-
3309	2299	-	Hypothetical protein	
3424	4191	+	Hypothetical protein	
5343	5726	+	Hypothetical protein	
5783	6715	+	Alpha/beta hydrolase fold-1 precursor	
6792	7820	+	Acyltransferase 3	
7852	8415	+	Hypothetical protein (Haloacid dehalogenase-like hydrolases (HAD-like superfamily))	
10339	8399	-	Putative polyferredoxin	
10998	10226	-	Hypothetical protein	
11558	10875	-	Hypothetical protein	
12792	11545	-	Hypothetical protein (RelA-SpoT-like protein)	
12930	13355	+	Hypothetical protein (4-hydroxybenzoyl-CoA thioesterase (4HBT))	
13452	13883	+	Phosphatidylglycerophosphatase A	EC 3.1.3.27
13916	14446	+	Competence-damaged protein CinA	
14559	15770	+	RecA protein	
15873	17729	+	Hypothetical protein	
17781	18059	+	Hypothetical protein	
18059	18229	+	Hypothetical protein	
18226	18615	+	Hypothetical protein	
18700	21416	+	Alanyl-tRNA synthetase	EC 6.1.1.7
21422	23233	+	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	TC 3.A.1.5.1

23361	24671	+	Hypothetical protein
24731	25732	+	Hypothetical protein
26843	25827	-	HD-hydrolase domain
27048	27434	+	Hypothetical protein
27878	27492	-	Single-stranded DNA-binding protein
28188	28763	+	18K peptidoglycan-associated outer membrane lipoprotein; OmpA/MotB precursor
28867	29382	+	Hypothetical protein
29427	30143	+	TPR repeat containing exported protein
30140	30595	+	Hypothetical protein

*KEGG: Kyoto Encyclopedia of Genes and Genomes (EC); TCDB: Transporter Classification Database (TC).

Table B5. List of genes present on the metagenomic insert of STC 13.

Start	Stop	Strand	Gene/ Function	KEGG/ TCDB*
26	4381	+	Hypothetical protein	
4489	4920	+	Hypothetical protein	
5902	4937	-	Maltose/maltodextrin ABC transporter, permease protein MalF	
5933	7251	+	Competence damage-inducible protein CinA	
7263	8384	+	RecA protein	
8723	9538	+	Hypothetical protein	
10475	9588	-	Hypothetical protein	
10537	11025	+	Hypothetical protein	
11049	12083	+	Cysteine desulfurase	EC 2.8.1.7
12071	12211	+	Aminotransferase class V	
12288	13340	+	A/G-specific adenine glycosylase	EC 3.2.2.-
14586	13390	-	Protein kinase domain protein	
16088	14547	-	Hypothetical protein	
16445	16167	-	LSU ribosomal protein L27p	
16904	16455	-	LSU ribosomal protein L21p	
17395	17021	-	Hypothetical protein	
18246	17644	-	Hypothetical protein	
21356	18242	-	Cation efflux system protein; Cobalt-zinc-cadmium resistance protein CzcA	TC 2.A.6.1.2
22350	21358	-	Probable Co/Zn/Cd efflux system membrane fusion protein	
22727	22344	-	Hypothetical protein	
22769	23440	+	Hypothetical protein	

23401	23589	+	Hypothetical protein
24140	23679	-	Hypothetical protein
24268	24522	+	Hypothetical protein
24999	16344	+	Two-component hybrid histidine kinase

*KEGG: Kyoto Encyclopedia of Genes and Genomes (EC); TCDB: Transporter Classification Database (TC). Red text shows genes targeted for sub-cloning. Grey shading shows the gene product used to construct the phylogenetic tree of the STC.

Table B6. **List of genes present on the metagenomic insert of STC 88.**

Start	Stop	Strand	Gene/ Function	KEGG*
422	1	-	Aspartyl-tRNA(Asn) amidotransferase subunit A	EC 6.3.5.6
841	518	-	Aspartyl-tRNA(Asn) amidotransferase subunit C	EC 6.3.5.6
1415	945	-	Hypothetical protein	
1543	1971	+	Septum formation protein Maf	
1968	2141	+	Septum formation protein Maf	
2843	2268	-	Hypothetical protein	
3696	2977	-	CTP:molybdopterin cytidylyltransferase	
4685	3693	-	Periplasmic aromatic aldehyde oxidoreductase, FAD binding subunit YagS	
4969	4751	-	Hypothetical protein	
5336	4947	-	Hypothetical protein	
5932	5486	-	Large-conductance mechanosensitive channel	
7063	6044	-	L-lactate dehydrogenase	EC 1.1.1.27
7343	7747	+	Hypothetical protein	
8434	9348	+	3-methyl-2-oxobutanoate hydroxymethyltransferase	EC 2.1.2.11
9469	10353	+	Pantoate--beta-alanine ligase	EC 6.3.2.1
10374	10862	+	Hypothetical protein	
13828	11732	-	Serine protease, subtilase family	
14284	14964	+	Hypothetical protein	
15066	16490	+	ATP-dependent DNA helicase	
17141	17683	+	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	EC 2.7.6.3
18608	17742	-	Hypothetical protein	

18989	19351	+	Hypothetical protein
19602	20837	+	Hypothetical protein
21576	21214	-	Two-component hybrid sensor and regulator
22553	21543	-	PAS/PAC sensor hybrid kinase
22743	22892	+	Hypothetical protein
23062	23601	+	Hypothetical protein
24362	24517	+	Hypothetical protein

*KEGG: Kyoto Encyclopedia of Genes and Genomes (EC). Grey shading shows the gene product used to construct the phylogenetic tree of the STC.

Table B7. List of genes present on the metagenomic insert of STC 89.

Start	Stop	Strand	Gene/ Function	KEGG*
35	1222	+	3-isopropylmalate dehydratase large subunit	EC 4.2.1.33
1222	1782	+	3-isopropylmalate dehydratase small subunit	EC 4.2.1.33
1779	2531	+	AdoMet-methyltransferase class I	
2900	2613	-	Hypothetical protein	
2941	3963	+	3-isopropylmalate dehydrogenase	EC 1.1.1.85
3973	4902	+	Branched-chain amino acid aminotransferase	EC 2.6.1.42
5004	6608	+	(R)-citramalate synthase	EC 2.3.1.182
6605	7417	+	Bacillosamine/Legionaminic acid biosynthesis aminotransferase PglE	
7491	7769	+	Putative aminotransferase, DegT family	
7871	9973	+	Penicillin-binding protein 1A	
10315	10106	-	Hypothetical protein	
10647	10312	-	Hypothetical protein	
11050	10667	-	Hypothetical protein	
11633	12193	+	Cell wall-associated hydrolase domain protein	
12212	12502	+	Hypothetical protein	
12937	12566	-	Hypothetical protein	
13644	14063	+	Hypothetical protein	
14093	15298	+	Hypothetical protein	
15907	15590	-	Hypothetical protein	
16095	16784	+	Hypothetical protein	
16365	16279	-	Putative bacteriocin	

16809	17033	+	Hypothetical protein	
20810	19629	-	Tyrosyl-tRNA synthetase	EC 6.1.1.1

*KEGG: Kyoto Encyclopedia of Genes and Genomes (EC). Grey shading shows the gene product used to construct the phylogenetic tree of the STC.

Table B8. List of genes present on the metagenomic insert of STC 90.

Start	Stop	Strand	Gene/ Function	KEGG*
125	871	+	Ste24 endopeptidase	EC:3.4.24.84
1130	837	-	Hypothetical protein	
1148	2698	+	Arginyl-tRNA synthetase	EC 6.1.1.19
2711	2896	+	Phage-shock protein PspC	
3012	3338	+	Transcriptional regulator, ArsR family	
3364	3768	+	Hypothetical protein (SRPBCC CalC domain)	
3872	4900	+	Fructose-1,6-bisphosphatase, GlpX type	EC 3.1.3.11
5784	5089	-	Acetyltransferase, GNAT family	
6043	5885	-	Hypothetical protein	
7294	6266	-	Fructose-bisphosphate aldolase class I	EC 4.1.2.13
7341	7883	+	Hypothetical protein	
8734	8024	-	Hypothetical protein	
8835	9026	+	Hypothetical protein	
9143	9496	+	Two-component response regulator	
9544	12378	+	Putative two-component system sensor kinase	
12489	12731	+	Hypothetical protein	
12758	13963	+	Two-component system response regulator	
13993	14871	+	Alkanesulfonates ABC transporter ATP-binding protein	
14858	15712	+	ABC transporter permease protein	
15732	16736	+	Taurine-binding periplasmic protein TauA	
16943	17179	+	Hypothetical protein	

17414	18160	+	Putative integral membrane protein
18293	20002	+	Membrane protein MMPL family (partial)
20053	20307	+	Metalloprotease MEP1-like protein
20457	20657	+	Hypothetical protein
20801	21307	+	Hypothetical protein
21339	22637	+	Putative integral membrane protein
22634	23953	+	Putative integral membrane protein
23996	24549	+	Magnesium and cobalt transport protein CorA (partial)
24955	26007	+	L-alanine-DL-glutamate epimerase
26024	27049	+	Protein often near L-alanine-DL-glutamate epimerase (cell wall recycling)
27081	28211	+	Peptidase M29, aminopeptidase II (complete)
28223	29070	+	Peptidase M29, aminopeptidase II (partial)
29088	29519	+	Magnesium and cobalt transport protein CorA
30185	29607	-	Hypothetical protein
30572	30258	-	Hypothetical protein
30330	30789	+	HIT family hydrolase
31774	31046	-	Diadenosine tetraphosphatase

*KEGG: Kyoto Encyclopedia of Genes and Genomes (EC). Red text shows genes targeted for sub-cloning. Grey shading shows the gene product used to construct the phylogenetic tree of the STC.

Table B9. List of genes present on the metagenomic insert of STC 91.

Start	Stop	Strand	Gene/ Function	KEGG*
186	3	-	Hypothetical protein	
1195	200	-	Hypothetical protein	
2239	1214	-	Hypothetical protein	
3284	2244	-	Hypothetical protein	
3451	4473	+	Lipid A biosynthesis lauroyl acyltransferase	EC 2.3.1.-
4891	4604	-	Translation factor SAU5	
5601	4858	-	Translation factor SAU5; TsaC protein (YrdC-Sua5 domains)	
5867	5598	-	Hypothetical protein	
5960	6787	+	Putative sugar ABC transport system, permease protein YjfF	
7098	7829	+	Hypothetical protein	
8851	8033	-	Conserved protein	
9032	9154	+	Hypothetical protein	
9439	12120	+	Hypothetical protein	
13565	12063	-	Magnesium chelatase	
14126	13680	-	Hypothetical protein	
14215	14415	+	RNA polymerase sigma-70 factor	
14412	15158	+	Hypothetical protein	
15638	16213	+	PDZ serine protease domain protein	
16782	17255	+	Alkyl hydroperoxide reductase subunit C-like protein	
17868	17506	-	Hypothetical protein	
18162	17887	-	Hypothetical protein	

18359	18796	+	Molybdenum cofactor biosynthesis protein MoaE
19090	20170	+	Hypothetical protein
19793	20171	+	Endonuclease III (partial)

*KEGG: Kyoto Encyclopedia of Genes and Genomes (EC). Grey shading shows the gene product used to construct the phylogenetic tree of the STC.

Table B10. List of genes present on the metagenomic insert of STC 92.

Start	Stop	Strand	Gene/ Function	KEGG/ TCDB*
16	972	+	Transmembrane efflux protein	
1127	2041	+	Luciferase-like protein	
2697	2056	-	Menaquinone biosynthesis methyltransferase	
3215	2898	-	Integral membrane protein TerC	
3258	3560	+	Hypothetical protein	
4021	3686	-	Hypothetical protein	
4188	5054	+	Putative oxidoreductase	
5118	5657	+	Nitroreductase	
5806	6972	+	Glucose/sorbosone dehydrogenase	
7052	7624	+	Hypothetical protein (DisA_N domain)	
7689	7955	+	Zinc finger UBP-type protein	
7975	8922	+	Oxidoreductase	
8953	10590	+	Na^+/H^+ antiporter	TC 2.A.36.7.1
10766	13180	+	FIG01004535: Hypothetical protein	
13410	13213	-	Maleylacetate reductase	EC 1.3.1.32
13763	14365	+	Mobile element protein	
15349	14399	-	Putative metal-dependent membrane protease	
16472	15480	-	Putative metal-dependent membrane protease	
17535	16540	-	Maleylacetate reductase	EC 1.3.1.32
17854	18837	+	Glyoxalase family protein	
19012	19575	+	DoxX family protein	

19641	20645	+	Mg/Co/Ni transporter MgtE	
20765	20962	+	Mg/Co/Ni transporter MgtE	
20968	22173	+	Potassium channel protein	
22283	24151	+	Hypothetical protein	
25388	24105	-	ATPase, AAA family	
25665	25423	-	Hypothetical protein	
25817	26078	+	Predicted zinc-dependent protease	
26313	27455	+	Hypothetical protein	
28085	27795	-	Hypothetical protein	
29021	28227	-	Hypothetical protein (Endonuclease/DNA excision repair)	
29470	29270	-	Hypothetical protein	
29996	29553	-	Cell division inhibitor, CDP-paratose 2-epimerase	
30267	31268	+	Threonine dehydratase	EC 4.3.1.19
31813	31325	-	Hypothetical protein	

*KEGG: Kyoto Encyclopedia of Genes and Genomes (EC); TCDB: Transporter Classification Database (TC). Red text shows genes targeted for sub-cloning. Grey shading shows the gene product used to construct the phylogenetic tree of the STC.

Appendix C

>ABC_transporter_substrate-binding_domain

atgaggatgaaaggcttcacgaagctcgtcgcggcgaccctgggaatcgcgcttctgaccggcggcgccggctgcacgcgca
aggacgacgacgggaacctcatggtcttccacggctcctaccgcgacgacgtgaagacgatggatccggccaacgcctacga
cacgatcagcctcgatacggtcgcctcgatctacgagagcctctaccagtactcgtacctggccgagccctacaagctcgtgccg
cttctcgccgccgacatgcccaaggtctccgacgacaagctcacctacacgatcccgatccgcaaaggcatcaagttccagga
cgacccttgcttcaaggagaccaacggcaaaggccgcgagctcaaggcccaggacttcgtctacgcctggaaacgactggct
ctgccgtcgcttcagtcgcagggctggtggatcttcgacggcaaggtgagcggcatcaacgacttccacaagaagctcaccca
gacgcccaaaggcgacgtgccccgggtgttcaacgagcccgtcgaaggcatcaaggccgttgacgactacacgatccagatc
gagctcacgcgcccttacccgcagcttcttcacgtgctcgcgatggcgttcaccgcgcccgtcccgcacgaggccgtgaccgcg
tatggcgacgagaacggcaacctctcggatcacccggtcggaacggggccgttcctcctgaagacctgggaccgcggccacc
ggatcgtcctcgagcgcaacccgaactaccaccccgagttctacccgaccgaaggcagcgtgagcttcaggaagaaaaacct
gctcgtcgacgccgggaagccccttcccttcctcgaccgcgtcagcatgaaagtgatccgcgagcagcagccgcagtggctga
gcttcatgaaaggcgagctcgacctgcttcccttgccgcgcgacaacttcaaccaggcgatcacgaacacgacgagcctctcg
ccggagctcgcctcgaagggcatccggctcacgatcgagaacggcgtggcgttctactacctctcattcaacatgaaggaccgc
ctgctcggcggcaacaaagcgcttcgccaggcgctctcctcggcgatcgaccgcgagaagtggatcaacctcttcacgaacgg
cacgggccagaagatggtgaacgcccttccgcccggagtgccggatcggccgcagggctcgaagctcaagtacgactacga
cccgacgcgcgccaaagagcttctccgaaaggcgggctatcccgaaggcaagggcctgccggcgatcaacatcgacatgcg
cggctcggccagcatcgatcggcagatgggcgagttcttctcccagcagcttgccgcgatcggcgtgaagacgaacgtgatcta
caacacgttccccgccttcctcgagaagacgaagcaaggcaaccaccagatctcctacggtggctggacgatggattacccgg
acgccgagaacgtctatcagctgctttacggcccgaacaaagcgcccgggcccggcgaagccaacttcgatcaccccgagtt
caaccaggtgttcgcgaagatggccgtcatggaatcgagccccgcccgagccaagctcgtcgccgagctcgacgagattctcc
aggaagaagtcccctgggccctcggctactacgtgaccgagtaccggctctctcaagcctggcttctcaactaccgcgcgagcg
agatcatcctgaacaagtacaagtactaccgggtgaatcgcgacgtgaagaagcgctaccaggaagtaaagtaa

>RelA-SpoT-like_hypothetical_protein

atgtcctctgaggcgctgcgcaaaaaattcgagttccgttgggaggttctcgacatcatcgtcggtggccggtcgtccctggatctg
accctgggcttccacatcacgaaccgcgaggaggcggaccgcttcctccaaagctacggctacgacgtcgatgacccgatcg
agcaagccgaggcgctcggaaacttccacgaggcgctgaacttcatccgtcgctacttccttcagcccgagaaccccgacggg
ctcaggctcgaggtgccgcgcaagctgctcgagctcaccgacatgcgcgacttgttcatgatgtcgggcatgagctacccggga
cagagctccgacacccaggggcttcttctcaggaactgggcctgcagcatccttaaggtgatgcacacgatcgcgcacgtggat
caggacatccggagcgcgtacttctcggacatccagcttcagatcttcgatcgctactacaaagtcatccagcgcgacgccgac
ggcgtgctttacctcggcgagcgggaggacgatccataccgggtgaacctcgaggctttcgagacgaaggccaggaagtccc
gcgactcgacgctcttgaagctgcttcacaagcccggcaacgtggccgaggacatcttcgaccgcgtgggcctgcgattcgtca
cggccacgcgcctggacgcgatccgcgtggtgaagttcctccaggaaaagcggatcatcatgccgccgaacatcaagccgag

ccgatcgcggaacacgttgatcgacctcgagcacctgcgctcggagctcgaccagaggattccccaggtcgcggacgccgct
gatttcagctcggacgcctggctgaagagcctggagcaggagctgaagttcccgaaagagacggtgacgaaccccgagaac
ccgttcacgagcgagtactaccgcgcgatccagttcacttcgcgacagctcatcaagctcaggaacccgctcttcagcgacttga
aggagctgaaaaccctgtcccgctccggcacgatcgacggcgaggccgccaagatcctcgatcgcatcgacttgaagtacatc
cagaaggagactcgcttcttctatccctacgaagtgcaagtcgtcgatgtccggagcgccgaggagaacgagaaggggcgaa
gcgcgcacagcgagtacaagcgctcgcaagtccatgccgctctccggcgcgtgatgggagcgctcgcggatgccgttcgatag

>HD-hydrolase_domain
atggatcagcccgaaagccttctctcagcctatatccccgtggacatccaaacgttccgaagctccacgcgcgacatcagcttcg
acgtcttcttgaagctctccgaagacaacatcgcgcacgtgttctcgcggacgaccggcctcgattacaagcgcctggctcagta
cattcagaaaggcgtccgcgagctctacatccgccgcgaggacgagaaagcttaccaggagttcacgagccgcacggccga
gaagatcttcgccgatcccaagacttcgcaagagaagaagatcgccacgctgctcaacatgaccgagcagaacatggccga
gctcttcacgcagttcgcggtttcggacgacacggccgcctcgtcagccaaggtgatccgcggctacgtcacgctgatgaccga
gaatcccaagtccttggcgacgatcctgaagctcgtctcgcacggcgagtacttgtattaccactcgatcgccgtctcggtgttcag
cttgttcatcgcgcgcgcttcgggccagttcaaccagaagacgctcgagatcgtcgggctcgggggctttctccatgacatcggcg
cgactcacctgcctaaggagatcgcctgctcccatgaggatctgaccgaggagcagtgggcgcagatgcgcgagcatccgaa
gacgggccttcagatggtcgaaggcacgcctgcgatcccggacgaggtccgctacatcatctaccagcatcacgaggagccc
ggaggcggtggctacccgaacgggcttcgcggcccggtgatctattatccggccaagatcgtggctctggccgatgcgttcagc
gctttgatctcgaagcgcccctcgcgcccggcttatacggtcgatgaggcgatcgagctacttcgcagtcaggctgggaagtacg
atcgcgatctcgtgcggatcctcgcttcggtctttgctcgcgaggcgacgaagaaggctgcgtga

>Cation_export_system_protein
atgctcaacgccgtcatccgcttctccctgcagaaccggctcctggtggtggcgctggcggccctgctgctggtctacggcacctg
ggtcctctccaagctgcccgtggacgtcttccccgacctggaccgccccaccgtggccatccaggcggaggcccatggcctgtc
gcccgaggaggtggaggccctggtggtctaccccatcgagaccgccctcaacggggccaccggggtggagcgggtccgctc
ggcctcggcgccgggcctgggtctcgtcttcgccgagttcgcctggggaaccgatatctacaccgatcggcagatcgtggcgga
gaagctccagggcatccggagcgccctgcccgagggcgtgaacccgtccctggggcccatctcctccatcatgggcgagatc
atgctgatcagcgtgagcggcgactccctctcccccatggagctgcgcaccctggcggactggaccatcaagccgcgcctcctc
agcatccccggcgtgtcccaggtcatcgccatcggcggcgacgccaagcagtaccaggtggtggtcgaccccgatcggctgc
gccagcacggggtgaccctgcgcgaggtggaggaggccgtgaaggagaccaatgccggatcctccggcgggttcctggtgc
gcgattacaccgagagcgtggtccgcaccctgggccgggtggagggcatggaggatctccgcgccaccgtggtgtcccgccg
cgacggcgggagcatcaccctggaccaggtggcggacctccgcttcgggtcgcccccccaagcgcggctccgccggcgccga
cggcgggccggccgtgatcctcagcgtcaagaagcagcccggcgccgccaccatcgagctcacggagcgcatcgacgccg
ccctggcggacctgcgcaagggcctgccctcccacgtgaaggtcaacgccaacctgttccgccagtcggtgttcatcgccgccg
ccatcaagagcgtcgaggacgccctgcgggacggcagcatcctggtcatcgtggtgctcttcctcttcctgctcaacttccgcacc
accttcatcaccctcaccgccatccccctctccatcctggtcacggccctggtcttccgggccttcgacatctccatcaacaccatg
accttgggcggcctcgccgtggccatcggcgagctcgtggacgacgccatcgtggacgtggagaacgtcttccggcggctcaa

ggagaaccgcctcaaggagcggcccgatcccttcctggaagtcatcttcaaggcctcctccgagatccgcaactccatcgtcttc
gccaccgtggtggtggtgctcgtcttcgtgcccctcttcgccctgggcggcatcgagggccgcatcttcgccccctgggcatctcc
tacatcatctccatcctggcctccctcctcgtctccctcaccgtgaccccggccctctgcgccttcctcctccccggcgccaagttcct
gggccgcgagggcgacggcttcctggtccgcgccctgaagcgctggaaccgcaaggtcttcctggagccggcgctggcgcgc
ccgaagcccttcctcgccggggccgtcctcctgctcgccgccgccctggcctccttcccctttcctgggacgggaattcctgccgcc
cttcaacgagggcaccgccaccctcaacatcttctccccgcccggcacctccctggcggagagcgaccgcatcgggggcatc
gccgagaggctggtcctccaggtccccgaggtcctgaccacgggccgccgcaccggccgcgccgaggaggacgaccatgc
cgagggcgtccactacaccgagatcgacgtggacctcaagccttcggggcggagccgcgacgagatcctggggcatatccgc
gcgcagctggaccagatcccgggcgtggccacgagcatcggccagcccatctcccaccgcatcgaccatatcctctcgggcat
ccgcgcccaggtggcggtcaagatcttcggggacgacctggccctcctccgggacaaggccgccgaggtggaggccctgctg
cgggggatcccgggcgtggtggacctgtcgacggagaagcaggtgctggtgccccaggtggagatcaagctggaccgcgag
gccttgaggctccatggcctgcgcgccggcgaggtgagcgagctggtggagacggccttgagcgggcgcatcgcgggcgag
gtgctggaggatcagaagcgcttcgccctggtggtgaggctggagggggccgaccggggcatcgagggcatcggcgacctg
ctcatcgacatccccggcggggagcgcatccccctctccatggtggccgaggtgcggccggccctgggccccaaccagatcat
gcgcgagaacgtgcggcgccgcatcgtggtgcaatgcaataccgaagggcgcgacctgggctcggtggtgcgcgacatcca
ggccgccgtcgccggcgaggtgtccctgccctccggctggttcatcgagtacggcgggcagttcgagagccagcagcgggcct
cccgcatgatcgccctgctcggcatcctctccctgatcggcatggccctggccctctacctccacttccgctcggcgcccctggtcg
gcctcatcctcctcaacatcccccttcgccctggtgggatcggtcctggcggtgtggcttaccacccgcaccttctcggtggggtccct
ggtcggcttcgtcaccctgtgcggcatctcggcgcgcaacggggtcatgatgctgtcgcattacctccacctgctccgcgaggaa
ggggaggaatggtccttcgccatggtgatccgcggggcctcggagcgcctggtgccggtgctgatgaccgccttgaccgccgcc
ctggccctggtcccgctcatgctcaaccccggcgagccgggcaaggagatcctctatcccgtggccgtggtcatcttcggcggcc
tcgtctcctccaccctgatgaacctggccctgatgccggccttgttctggaactacggccggaaggccgtcgcgcgcattctgaac
ccccccaccaggagcctgccatga

>Peptidase_M29

gtgaacgacccgcggatcgaccggctcgggcggctgatcgtcgaatactcgctcgagctgggggagggcgactcacttcgcgt
cgacggctcggagcttgcggcggcgcacatgctcgcactgctgcgcgcggcactcggccgcggcgcgaacgcgtcgacgaa
cgtggcgctcggcggcctcgccgagctcctggtggcgcgcggcagcgacgagcagatcggcttcatctcgcagatggcgcgc
aacgagatcgagaacgtcgacgcggtcgtgacgctctggtcggaggcgaacacgcgcgcattcacgcgcgccgacccgcg
acggcacgcgctccatctcgcgacgcggcgacagctgagcgagcggctgcgggagcggatcgccgcgggcgaggcgcgct
ggtgcggcacgctcgtgccgacccacgcgttcgcgcaggacgcggagatgtcgctcgccgagttcgaggacttcgtctaccgg
gcgctgcacgtcgacggcgaggaggacgctgtcgagtactggcgacggcgcgaggcggagctcgacgcgatcgccgcggg
tcttgccgacgtgcgcgagctgcggatcgtcgggcccgacaccgacatcaccgtcggcgtcgacgggcgccggtggaaggtc
gccggcggccggcgcaacctccccgacggcgaggtcttcaccagcccggtcgagacggatacgcgcggcgagatccgcttct
ccttcccggcgatcttccaaggtgtcgaggtggaggacgtgcggctgcgcttcgaggacgggcgcgtcgtcgcgcacgaggcg
ggccgcggccgcgagcacctcgcggcgatgctcgacctcgacgacggtgcgcgtgtcctcggcgaggtcgcgttcggcctca
actacgagatcgaccgcttcacccgcaacatcctctacgacgagaagatcggcggcacgatgcacctcgcgctcggttccggg

ttcacggacatcggcggcgagaacgtttccgcgctgcactgggacatgatctgcgacctgcgcgcggacggcgaggtgtacgc
cgacggcgaggtcgtgtggcgcgcaggccggttcctgaaggagccggcgatggcggatgtgggctag

>Na+H+_Antiporter
gtggagcacctggagattatcgtcctctggctcctgatggccgtggcagggctcgcactgctcgcccgcgtcttgcgcgtccccta
cccgatcttcctcgtcatcgggggcctggctctgggcttcatgcccggcatacccgaggtcgagctcaagccggatctggtcttgct
catcttcctccccccctctcctgtactcggcggcgttcttctcctctttgcgagatctgcgggcaaacatcaggccgatagggctcctct
ccgtggggctcgtcgtgctcacgatggtggtcgtcgccggggtagcgcactgggcggtggggctcccgtgggcggcggcgttcg
ttctgggggcgatcgtctcgccgaccgacccggtggcggcgacggccgtcgcagagcggctcggggtgccgcgccggatcgt
gaccgtcctcgaaggcgagagcctgatcaacgacgggacggcgctcgttctgtaccaggcagccgtgcgcgtggccgtcggc
accgtggccttcggggtcttctcgttcgtggaggcggggctgcgcttcgtgctcggcatcctggtcggggtggcgatcgggctcgc
ggtcggctgggtcatcgccttcgtgcggcgcaggatcgaggatcctctggtggagatcacgatctccatattcaccggctacgcc
gcctacctgccagccgaggaactcaacgcctcagagatctttatgggctcgggagtgccggcttcgggcgtgctcgccgcggtc
accgccgggctctacctgaaccggcgcgcccccgagatggtctccccgcgcacccgcctgcaggcgctcgcggtgtgggagg
tgatcatctttatgctcaactcgctcctctttatcctcatcgggttgcaactcccgaacgtactcggagagttctcggaggaaccggcg
gggaccctcgtgctctacgccgtgctcgtgagcctggtggtggtcgtgacgcgcgtcttatgggtctttccggccacgtacatcccc
cgctgggcgagccgccgcctgcgcgagcgcgacccctctccccctggcaggccgtcaccgtcatcgcctacgcgggcatga
ggggcgccgtctctctggcggccgcgctcgcgctacccctgacgacgcagagcggcgacgcgttccccgggcgcgacctgat
aatcttcctgaccttctgcgtcatcctcgtcaccctcgtcctgcagggccttagcctgccggtcctgatccggtggctcggcctagag
gacgacgggatcagcgcacgcgaggaggtggcggcccgcctccgggctgccgaggcggccctggccaggatcgaggagc
tcgccggagaggattgggtgcgcgaggacaccgccgaacgcatgcgcggcctctacggttatcggcggcggcgcttcgccgct
cgagccgccgaagccgacggcgatgaggagtacgaagactacgaggggcgctccgtcgccttccagcgcctccggcgcga
gttgctcgaggcggagcggtcggccctgctacgcttgaggagcgagggaaagataggcgaggaggccatgcgccgcgtgga
gcgggacctggacctcgaagatcagcggttggagatctag

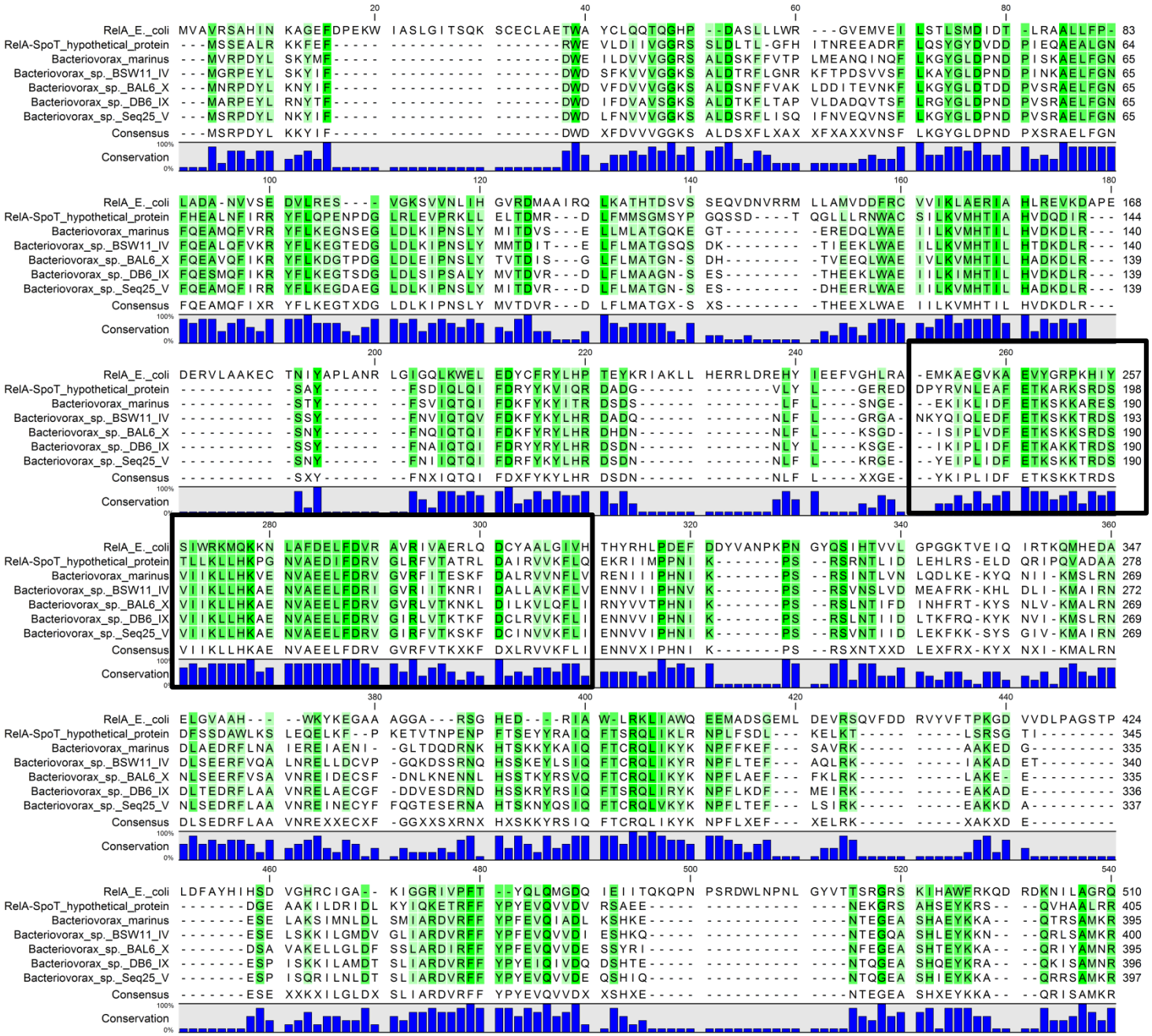


Figure C1. Multiple sequence alignment using Clustal Omega of amino acid sequences of *E. coli* RelA protein (only 600 residues shown), the RelA-SpoT hypothetical protein identified in this study and the top five BLASTp matches, all *Bacteriovorax* species. Shading indicates 71% (5/7) conservation, or greater, at a particular position. The N-terminal half of the RelA-SpoT domain is indicated by the black boxes, which highlights this conserved region. The C-termini of the proteins are not shown.

Figure C2. Multiple sequence alignment using Clustal Omega of the amino acid sequences of the HD-hydrolase protein identified here and the two closest BLASTp matches (*Bdellovibrio bacteriovorus* and *Hydrogenedentes* spp). The conserved HD domain (as shown in Figure 24) is indicated by the black boxes. Shading indicates 66% (2/3) conservation, or greater, at a particular position.

Table C1. Statistical significance of the average survival rate of *E. coli* BL21(DE3) transformants on 3.5% NaCl LB agar, as determined using Student's *t*-test (unpaired).

Culture	pET21a	RelA-SpoT:HD-hydrolase	Peptidase M29	Na^+/H^+ Antiporter
pET21a	1			
RelA-SpoT:HD-hydrolase	0.24	1		
Peptidase M29	0.0051*	0.068	1	
Na^+/H^+ Antiporter	0.07	0.078	$4.35e^{-06}$*	1

**P*-values <0.05, which indicates cultures with significantly different survival rates.