

GREAT AUSTRALIAN BIGHT RESEARCH PROGRAM

RESEARCH REPORT SERIES

Molecular assessment of biodiversity in the Great Australian Bight

Final Report GABRP Project 3.2

Jason E Tanner, Luciano B Beheregaray, Lev Bodrossy, Sharon Hook,
Andrew Oxley, Minami Sasaki, Jodie van de Kamp and Alan
Williams

GABRP Research Report Series Number 26

October 2017



DISCLAIMER

The partners of the Great Australian Bight Research Program advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised that no reliance or actions should be made on the information provided in this report without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, the partners of the Great Australian Bight Research Program (including its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

The GABRP Research Report Series is an Administrative Report Series which has not been reviewed outside the Great Australian Bight Research Program and is not considered peer-reviewed literature. Material presented may later be published in formal peer-reviewed scientific literature.

COPYRIGHT

©2017

THIS PUBLICATION MAY BE CITED AS:

Tanner, J.E., Beheregaray, L.B., Bodrossy, L., Hook, S., Oxley, A., Sasaki, S., van de Kamp, J. and Williams, A. (2017). Molecular assessment of biodiversity in the Great Australian Bight. Final Report GABRP Project 3.2. Great Australian Bight Research Program, GABRP Research Report Series Number 26, 150pp.

CONTACT

Prof Jason Tanner
SARDI
e: jason.tanner@sa.gov.au

FOR FURTHER INFORMATION

www.misa.net.au/GAB

GREAT AUSTRALIAN BIGHT RESEARCH PROGRAM

The Great Australian Bight Research Program is a collaboration between BP, CSIRO, the South Australian Research and Development Institute (SARDI), the University of Adelaide, and Flinders University. The Program aims to provide a whole-of-system understanding of the environmental, economic and social values of the region; providing an information source for all to use.

CONTENTS

CONTENTS.....	iii
List of figures.....	vii
List of tables.....	xii
Acknowledgements.....	xiii
1. Executive summary.....	1
2. Introduction	2
2.1 Overview	2
2.1.1 Sub-project 1: Barcoding.....	3
2.1.2 Sub-project 2: Hydrocarbon degraders.....	3
2.1.3 Assessment of baseline assemblages	4
2.2 Objectives.....	5
2.3 References	5
3. Molecular barcoding of selected benthic taxa	8
3.1 Executive summary	8
3.2 Introduction	8
3.2.1 Aim	9
3.2.2 Outputs and extensions	9
3.3 Methods.....	9
3.3.1 Sampling.....	9
3.3.2 Laboratory procedures.....	10
3.3.3 Data analysis	11
3.4 Results and Discussion	12
3.4.1 Optimisation of laboratory procedures	12
3.4.2 Summary statistics	20
3.4.3 Proof of concept of identification strategy.....	21
3.5 Conclusions and recommendations for future studies.....	23
3.6 References	23
APPENDIX 3.1: Summary of DNA extraction protocols tested for the different phyla and tissues..	25
APPENDIX 3.2: Species Identifications for each sample	27
APPENDIX 3.3: Neighbour-Joining trees	35
4. Spatial distribution and diversity in hydrocarbon degrading microbes in the Great Australian Bight I: Functional Gene Distribution	39
4.1 Executive summary	39

4.2	Introduction	39
4.3	Materials and Methods.....	40
4.3.1	Sample collection	40
4.3.2	DNA extractions	41
4.3.3	Database Construction.....	42
4.3.4	Amplicon Generation and Sequencing	42
4.3.5	Bioinformatics	43
4.3.6	Data Analysis and Statistics.....	43
4.4	Results.....	44
4.4.1	Next-Gen Sequencing results	44
4.4.2	Diversity overview.....	44
4.5	Discussion.....	51
4.6	References	53
5.	Spatial distribution and diversity in hydrocarbon degrading microbes in the Great Australian Bight II: Analysis of bacterial and archaeal community structure	56
5.1	Executive summary	56
5.2	Introduction	56
5.3	Material and methods	57
5.3.1	Sample collection	57
5.3.2	DNA extractions	58
5.3.3	PCR and Sequencing.....	59
5.3.4	Bioinformatics	59
5.3.5	Phylogeny.....	59
5.3.6	Data Analysis and Statistics.....	60
5.4	Results	60
5.4.1	Sequencing output and quality assurance.....	60
5.4.2	Community composition.....	64
5.4.3	Archaea	74
5.4.4	Diversity Indices	77
5.4.5	Relationships between community composition and environmental variables	81
5.5	Discussion.....	91
5.6	References	93
6.	Development of a functional genomics assay to measure the abundance of hydrocarbon degrading bacteria in the Great Australian Bight	96

6.1	Executive summary	96
6.2	Introduction	96
6.3	Materials and methods	97
6.3.1	Study sites and sampling design	97
6.3.2	DNA extractions	98
6.3.3	PCR primer design	98
6.3.4	Detection and quantification of <i>alkB</i> , <i>c23o</i> and <i>pmoA</i> genes via qPCR.....	105
6.3.5	Statistical Analysis	105
6.4	Results	105
6.4.1	<i>alkB</i>	105
6.4.2	<i>pmoA</i>	107
6.4.3	<i>c23o</i>	108
6.5	Discussion.....	109
6.6	References	111
7.	A comparison of ecogenomic tools for assessing biodiversity in the Great Australian Bight	113
7.1	Executive summary	113
7.2	Introduction	114
7.2.1	Objectives.....	114
7.3	Methods	114
7.3.1	Sample collection and DNA extraction	114
7.3.2	Deep-sequencing of benthic sediment samples	115
7.3.3	Bioinformatics sequence data processing	116
7.3.4	Statistical analyses	117
7.4	Results and Discussion	118
7.4.1	18S rRNA gene based estimates of benthic diversity	118
7.4.2	COI estimates of benthic diversity	126
7.4.3	Comparisons between COI-5P and 18S rRNA markers for estimating benthic diversity	137
	Appendix 7.1. List of benthic sediment samples collected over a range of depths in 2013 and 2015 from the GAB and the corresponding gene target/s sequenced.....	142
7.5	References	145
8.	Discussion.....	147
9.	Appendix 1: data management.....	149
9.1	Raw datasets created.....	149

9.2	Data processing and derived datasets	149
9.3	Data curation and archive	149
9.4	Data access, use agreements and licensing	149
9.5	Publication of datasets.....	149
10.	Appendix 2: project publications	150
10.1	Papers	150
10.2	Presentations	150

LIST OF FIGURES

Figure 3.1. The number of sequences in the final data set for each phylum.	20
Figure 3.2. The proportion of sample identification for a given taxonomic level using p-distance thresholds (species: 0.00 – 1.65%, genus: 1.66 – 10.36%, family: 10.37 – 12.90%, order: 12.91 – 15.00% and class: 15.01 – 21.00%).	21
Figure 3.3. Neighbour-Joining tree for the genus <i>Ophiomusium</i> based on Kimura 2-parameter genetic distance with 1000 bootstraps. Samples with coloured circles are from the current study. Green circles indicate that the sample was identified at the species level (p-dis < 1.65%) using our proposed strategy, while red circles indicate that the sample was identified at the genus level (1.65% < p-dis > 10.36%) using threshold values. Sequences from samples without coloured circles were obtained from databases (NCBI and BOLD) and their accession numbers and sample IDs are shown after the species names.	22
Figure 4.1. Locations sampled by the <i>RV Southern Surveyor</i> for sediments. Green, red, blue and black depth contours are 200, 1000, 2000 & 3000 m respectively, green polygons indicate the BP exploration leases, and the blue polygon the benthic protection zone of the Great Australian Bight Marine Park. T1, T2 etc indicate transect numbers.	41
Figure 4.2. Rarefaction curve for gene sequence diversity for the <i>alkB</i> gene. Number of different sequences is shown on the y axis, and the number of sequencing reads are on the x-axis. Each station is shown in a different colour. Although additional new sequences are discovered with the addition of more sequencing depth, the curves have nearly plateaued, demonstrating that our efforts have captured the vast majority of environmental diversity.	44
Figure 4.3. A cladogram showing sequence relatedness amongst sediment samples collected from the GAB. Percent similarity is on the vertical bar. Black lines indicate statistically significant, solid relationships. Red lines indicate relationships with limited statistical support.	45
Figure 4.4. Principal co-ordinates plot showing variables that correlate with differences in diversity for the three selected functional genes, with replicates from each station plotted. For clarity, transects, not sample ID, are shown.	45
Figure 4.5. A cladogram showing <i>alkB</i> sequence relatedness amongst water samples collected from the GAB. Percent similarity is on the vertical bar. Black lines indicate statistically significant, solid relationships. Red lines indicate relationships with limited statistical support.	47
Figure 4.6. Principal co-ordinate analysis plot showing variables that correlate with differences in <i>alkB</i> sequence diversity.	47
Figure 4.7. Physical characteristics of the water column for the transects surveyed.	48
Figure 4.8. A cladogram, or hierarchical clustering tree based on phylogenetic relationships, based on nucleotide sequences of the <i>alkB</i> gene.	49
Figure 4.9. A cladogram based on nucleotide sequences of the <i>pmoA</i> gene.	50
Figure 4.10. A cladogram based on nucleotide sequences of the <i>c23o</i> gene.	51
Figure 5.1. Transects and sampling locations used in the 2013 and 2015 cruises. Sampling locations used for development of the functional assays are shown in red. BP's lease blocks are shown in orange, and the light blue shading indicates Commonwealth marine reserves.	58

Figure 5.2. Rarefaction curve of bacterial benthic samples, showing the number of new bacterial taxa identified (y-axis) relative to the number of sequences analysed (x-axis) for each station.	62
Figure 5.3. Rarefaction curve of bacterial pelagic samples, showing the number of new bacterial taxa identified (y-axis) relative to the number of sequences analysed (x-axis) for each station.	62
Figure 5.4. Rarefaction curve, showing the number of new archaeal taxa identified in benthic samples (y-axis) relative to the number of sequences analysed (x-axis) for each station.	63
Figure 5.5. Rarefaction curve, showing the number of new archaeal taxa identified in pelagic samples (y-axis) relative to the number of sequences analysed (x-axis) for each station.	63
Figure 5.6. The representation of different taxonomic groups of bacteria in benthic samples collected at each station. Depth and transect are on the x axis, and % contribution of each taxonomic group to the overall composition of the sample are on the y axis. Only taxonomic groups that make up more than 0.001% of the community are shown, chloroplast and mitochondrial rRNA have been removed. Each vertical line represents a different sample, but for clarity, only depths are indicated in the x-axis labels. The taxonomic resolution in panel A is to phylum, panel B shows class, panel C shows order.	67
Figure 5.7. The relative abundance of bacteria in orders (panel A) or families (panel B) known to contain hydrocarbon degrading organisms. The station (in order of transect) and depth are shown on the x axis, and the abundance on the y axis. For clarity, only depth is labelled.....	69
Figure 5.8. The representation of different taxonomic groups of bacteria in pelagic samples collected at each station. Depth and transect are on the x axis, and % contribution of each taxonomic group to the overall composition of the sample is on the y axis. Only taxonomic groups that make up more than 0.001% of the community are shown, chloroplast and mitochondrial rRNA have been removed. Each vertical line represents a different sample, but for clarity, only depths are labelled. The taxonomic resolution in panel A is to phylum, panel B shows class, panel C shows order.	73
Figure 5.9. Comparative abundances of different groups of <i>Archaea</i> collected from benthic samples at the different depths. Depth and transect are on the x axis, and % contribution of each taxonomic group to the overall composition of the sample are on the y axis. Only taxonomic groups that make up more than 0.001% of the community are shown. For clarity, station numbers are not shown. Panel A is taxonomic resolution to phylum, panel B shows class, panel C shows order.....	75
Figure 5.10. (Prev page) Comparative abundances of different groups of <i>Archaea</i> collected from pelagic samples at the different depths. Depth and transect are on the x axis, and % contribution of each taxonomic group to the overall composition of the sample are on the y axis. Only taxonomic groups that make up more than 0.001% of the community are shown. For clarity, station numbers are not shown. Panel A is taxonomic resolution to phylum, panel B shows class, panel C shows order.	77
Figure 5.11. Biodiversity metrics in the benthic bacterial samples. Panel A shows diversity as determined by either the Shannon (H) or Simpson (1-lambda) metrics; Panel B shows evenness for each station.....	78
Figure 5.12. Biodiversity metrics in the pelagic bacterial samples. Panel A shows diversity as determined by either the Shannon (H) or Simpson (1-lambda) metrics; Panel B shows evenness for each station.....	79

Figure 5.13. Biodiversity metrics in the benthic archaeal samples. Panel A shows diversity as determined by either the Shannon (H) or Simpson (1-lambda) metrics; Panel B shows evenness for each station.....	80
Figure 5.14. Biodiversity metrics in the pelagic archaeal samples. Panel A shows diversity as determined by either the Shannon (H) or Simpson (1-lambda) metrics; Panel B shows evenness for each station.....	81
Figure 5.15. Relationships between bacterial benthic samples as visualised via principle co-ordinates analysis.....	82
Figure 5.16. Relationships between families containing known hydrocarbon degrading taxa in benthic samples as visualised via principle co-ordinates analysis.....	83
Figure 5.17. Relationships between the bacterial composition of different pelagic samples as visualised via principle co-ordinates analysis.	83
Figure 5.18. Relationships between the archaeal composition of different benthic samples as visualised via principle co-ordinates analysis.	84
Figure 5.19. Relationships between the archaeal composition of different pelagic samples as visualised via principle co-ordinates analysis.	84
Figure 5.20. Relationships between community composition in benthic bacterial samples and environmental variables as determined via distance-based redundancy analysis (panel A) or principle co-ordinates analysis (panel B).	87
Figure 5.21. Relationships between community composition in pelagic bacterial samples and environmental variables as determined via distance based redundancy analysis (panel A) or principle co-ordinates analysis (panel B).	89
Figure 5.22. Relationships between community composition in benthic archaeal samples and environmental variables as determined via distance based redundancy analysis (panel A) or principle co-ordinates analysis (panel B).	90
Figure 5.23. Relationships between community composition in pelagic archaeal samples and environmental variables as determined via distance based redundancy analysis (panel A) or principle co-ordinates analysis (panel B).	91
Figure 6.1. Transects and sampling locations used in the 2013 and 2015 surveys. Sampling locations used for development of the functional assays are shown in red. BP's lease blocks are shown in orange, and the light blue shading indicates Commonwealth marine reserves.	98
Figure 6.2. (prev pages) Clades selected for designing <i>alkB</i> (panel A), <i>pmoA</i> (panel B), and <i>c23o</i> (panel C) qPCR primers. The relative abundance of each taxon is given in the numbers in blue next to each clade name. These abundances can be compared to their representation in the literature as a whole (in black) and to their abundance in the Gulf of Mexico (in brown). The numbers of amplicons from bacteria in each clade are in green. Clades outlined in light brown contain sequences that have only been found in the Great Australian Bight. Clades outlined in dark brown contain sequences that have been found in the GAB as well as sequences that have been described in other studies. Clades outlined in black only contain sequences from studies in the literature and do not have members that were identified in this study.....	103
Figure 6.3. Number of copies of <i>alkB</i> detected with the different primer sets in sediment samples collected from the GAB. Transect 2 is shown in black, Transect 3 in red, and Transect 4 in blue.	

Inequalities in the bottom left of each panel indicate statistically significant differences between depths within a transect, while those in the bottom right indicate differences between transects within a depth (ANOVA followed by a Tukey's HSD pairwise comparison, $p < 0.05$). 107

Figure 6.4. Number of copies of *pmoA* detected with the different primer sets in sediment samples collected from the GAB. Transect 2 is shown in black, Transect 3 in red, and Transect 4 in blue. Inequalities in the bottom left of each panel indicate statistically significant differences between depths within a transect, while those in the bottom right indicate differences between transects within a depth except for panel A, where differences are between the main effect of depth (ANOVA followed by a Tukey's HSD pairwise comparison, $p < 0.05$). 108

Figure 6.5. Number of copies of *c23o* detected with the different primer sets in sediment samples collected from the GAB. Transect 2 is shown in black, Transect 3 in red, and Transect 4 in blue. Where shown, inequalities in the bottom left of each panel indicate statistically significant differences between depths within a transect, while those in the bottom right indicate differences between transects within a depth (ANOVA followed by a Tukey's HSD pairwise comparison, $p < 0.05$). 109

Figure 7.1. Rarefaction curves depicting the number of observed 18S-V4 OTUs obtained from 94 benthic surface sediment samples. Charts plot the number of observed OTUs against the number of reads obtained from (a) pre-filtered and (b) filtered ($> 0.01\%$ abundance) data. 119

Figure 7.2. Numbers of 18S-V4 OTUs assigned at different taxonomic levels for major phyla following interrogation against a curated SILVA 18S database using the NCBI taxonomy. 120

Figure 7.3. Mean relative abundances of OTUs at the phylum level across depths within the GAB based on analysis of the 18S-V4 gene region. 121

Figure 7.4. Ordination (PCO) plot of the global 18S-V4 OTU profiles from 93 benthic surface sediment samples collected over a range of depths from the GAB. 122

Figure 7.5. Ordination (PCO) plots of the relative abundances of 18S-V4 OTUs for major phyla from 93 benthic sediment samples collected over a range of depths in the GAB. 123

Figure 7.6. K-dominance plots representing the cumulative dominance of 18s-V4 OTUs across depths in the GAB. 124

Figure 7.7. Shade plot representing the distribution of the 50 most important 18S-V4 OTUs. Stations have been grouped using Bray-Curtis similarities on the fourth root transformed data, by hierarchical agglomerative clustering with group average linking and SIMPROF (Type 1) tests. OTUs have been grouped using the Index of Association by hierarchical agglomerative clustering with group average linking and SIMPROF (Type 3) tests. White space denotes absence of an OTU at that Station; depth of the grey scale is a linear proportion of the fourth-root transformation of abundance data. 125

Figure 7.8. Relationships between 18S community composition and environmental parameters as determined by distance-based linear modelling and visualised with distance-based redundancy analysis. Only variables contributing to the final selected model are included. Temperature and depth were strongly collinear ($R > 0.9$), and thus temperature was retained while depth was removed from the analysis. 126

Figure 7.9. Rarefaction curves depicting the number of observed COI-5P OTUs obtained from the 65 benthic sediment samples obtained in 2013. Charts plot the number of observed OTUs against the number of reads obtained from (a) pre-filtered and (b) filtered ($> 0.01\%$ abundance) data. 127

Figure 7.10. Total numbers of COI-5P OTUs assigned to major phyla based on interrogation of sequences against the BOLD systems database.	128
Figure 7.11. Numbers of COI-5P OTUs assigned at different taxonomic levels for major phyla following interrogation of the GenBank COI collection. The % sequence identity cut-off values used at each taxonomic level are indicated in the key.	130
Figure 7.12. Mean relative abundances of OTUs at the phylum level across depths within the GAB based on analysis of the COI-5P gene region.....	132
Figure 7.13. Ordination (PCO) plot of the global COI-5P OTU profiles from 65 benthic sediment samples collected over a range of depths from the GAB.	133
Figure 7.14. Ordination (PCO) plots of the relative abundances of COI-5P OTUs for all taxa and major phyla from 65 benthic sediment samples collected over a range of depths in the GAB.....	134
Figure 7.15. K-dominance plots representing the cumulative dominance of COI-5P OTUs across depths in the GAB.	135
Figure 7.16. Bubble plots depicting the relative abundances of example macro- and microfaunal COI-5P OTUs across all 65 benthic sediment samples collected in 2013 from a range of depths within the GAB. Black dots represent individual samples and grey bubbles the relative abundance.....	136
Figure 7.17. Relationships between COI community composition and environmental parameters as determined by distance-based linear modelling and visualised with distance-based redundancy analysis. Only variables contributing to the final selected model are included. Temperature and depth were strongly collinear ($R > 0.9$), and thus temperature was retained while depth was removed from the analysis.	137
Figure 7.18. Comparison of the total numbers of 18S and COI-5P OTUs assigned to major phyla from 61 matched benthic sediment samples obtained in 2013 from the GAB. Common and unique taxa which could be classified to the phylum level are marked for each gene target with black and red dots respectively.	139
Figure 7.19. Ordination (PCO) plots comparing the (a) COI-5P and (b) 18S global OTU profiles obtained from the sequencing of 61 matched samples collected over a range of depths from the GAB in 2013.	140
Figure 7.20. Ordination (PCO) plots comparing the relative abundances of COI-5P and 18S OTUs for four major phyla across depths from 61 matched GAB benthic sediment samples collected in 2013.	141

LIST OF TABLES

Table 3.1. PCR cycles and primer combinations used to amplify the COI region for GAB samples.	10
Table 3.2. Samples used to calculate intraspecific diversity.....	13
Table 3.3. Results of DNA extractions and corresponding protocol for each phylum.....	17
Table 3.4. Summary of efforts on DNA extraction, PCR and sequencing.	18
Table 4.1. Illumina adapter and primer sequences used to generate amplicons for high throughput sequencing.	43
Table 4.2. Sediment grain size distribution with depth for each of the sampling transects. Values are percent composition by volume up to 1 mm, and percent composition by weight for fractions > 1 mm.	46
Table 5.1. The number of reads and OTUs derived from the sequencing results, divided into data sets and partitioned by environment, where 16S are bacterial sequences, and A16S are archaeal sequences.	61
Table 5.2. A comparison of the taxonomic identification of microbes identified in the Great Australian Bight to those enriched in samples originating from the response to the <i>DeepWater Horizon</i> oil spill. Closed symbols denote families, open symbols genera.....	68
Table 5.3. Relationships between community composition and environmental parameters as determined by distance-based linear modelling. Only variables contributing to the final selected model are included. Values for Pseudo-F and P represent values for marginal tests for each variable individually.	85
Table 5.4. Sediment grain size distribution (%) with depth for each of the sampling transects. Values are percent composition by volume up to 1 mm, and percent composition by weight for fractions > 1 mm.	88
Table 6.1. Primers used in qPCR reactions	104
Table 7.1. List of COI-5P OTUs which match the COI sequences obtained from individual GAB barcoded specimens. The total numbers of sequence reads are indicated and are given for each depth category.....	131

ACKNOWLEDGEMENTS

Funding for this study was provided by the Great Australian Bight Research Program - a collaboration between BP, CSIRO, the South Australian Research and Development Institute (SARDI), the University of Adelaide, and Flinders University. The Program aims to provide a whole-of-system understanding of the environmental, economic and social values of the region; providing an information source for all to use. The crew and scientists, especially Mark Green, Mark Lewis, Matt Sherlock, Jeff Cordell and Michael Watson on board the *RV Southern Surveyor* and John Highton, Mark Green, Phil de Boer and Matt Sherlock on board the *RV Investigator* are thanked for obtaining samples. Franzis Althaus is thanked for her assistance with figures.

We thank Thierry Laperousaz and Andrea Crowther from the South Australian Museum for facilitating delivery of samples to Flinders University for barcoding, and Emily Daws for assistance in the laboratory. We also acknowledge Peter Teske for contributing expertise with the choice of COI primers and advice on sequence strategies. LBB acknowledges salary support from an Australian Research Council Future Fellowship (FT130101068).

We would like to gratefully thank Dr S Catalano (Molecular Sciences, SARDI Aquatics) for laboratory support; Dr A Camarinha-Silva (Institute of Animal Science, Animal Nutrition Group, University of Hohenheim, Germany) for providing direction and helpful comment with the NGS assays; and Dr M Wos-Oxley (Flinders University) for assistance with sequence data analysis. The authors also wish to acknowledge the Australian Genome Research Facility - AGRF and their staff (L McGrath, P Gooding, S Tyagi and N Kasinadhuni) for providing exceptional support in the sequencing and bioinformatics processing of the customised NGS assays.

1. EXECUTIVE SUMMARY

As part of a broader project to assess spatial patterns in benthic biodiversity in the Great Australian Bight (GAB), we employed a number of molecular approaches to complement the traditional approaches that are reported elsewhere. Due to the potential for oil and gas exploration in the region, we were particularly interested in the presence of microbial taxa that might have the potential to degrade hydrocarbons. We also employed two metabarcoding approaches to assessing the eukaryotic assemblages present, and undertook barcoding of individual specimens in an attempt to improve the potential to provide species level identifications to the molecular barcodes obtained from environmental samples.

Samples of individual infauna and epifauna from two field surveys (2013 and 2015) in the GAB were DNA barcoded for species identification using the cytochrome c oxidase subunit 1 (COI) region of the mitochondrial genome. The samples successfully analysed corresponded to 139 Arthropoda, 106 Echinodermata, 27 Mollusca, 14 Annelida, 11 Sipuncula and 6 Cnidaria species. We then employed two molecular metabarcoding approaches, using the nuclear 18S rRNA gene and COI, to assess biodiversity in a range of sediment samples from the GAB. For both genes, approximately 1000 operational taxonomic units (OTUs - putative species) were identified, although the majority of these could not be assigned species level taxonomy (86.5% and 95.5% respectively). Only seven of 303 taxa from the GAB that were individually barcoded were identified in the 65 sediment samples that were successfully analysed for the COI gene. Both gene regions show a clustering of assemblage structure with depth, although there were some clear differences in the taxa detected.

The diversity and relative abundance of key genes in three pathways of hydrocarbon degradation: *alkB*, which is involved in alkane degradation, *c23o* which is involved in the degradation of aromatic compounds, and *pmoA*, which metabolises methane, was determined in sediment and water samples collected from the GAB. We found numerous copies of each gene in all sediment samples, and of *alkB* and *c23o* in the water samples, suggesting that bacteria in the GAB have the capacity to degrade oil. Furthermore, the sequences of these genes were unique, suggesting that while these bacteria have the capacity to degrade hydrocarbons, their rates of response to sources of hydrocarbons, and thus rates of oil degradation, may not be easily predicted based on studies conducted elsewhere. Using the same samples, we analysed the community composition of Bacteria and Archaea. The bacterial assemblage was more diverse than the Archaea, and contained a variety of taxa related to those that have increased in abundance in response to oil spills in other locations. The sediment and water samples had different microbial communities, and the sediment samples were clustered more tightly, indicating a higher degree of relatedness. The microbial communities in sediment samples collected from the continental edge (200 and 400 m) differed from those collected from the continental slope (depths greater than 1000 m). Following this, we designed quantitative PCR primers specific to the microbial taxa indigenous to the GAB. Although these bacteria were rare (between 10^{-5} to 0.1% of the total bacterial population), they were present at every station and every sample we analysed. These qPCR assays can be used as both a high throughput screening tool to measure the spatial and temporal duration of impact from routine discharges of petroleum and monitor environmental fate and persistence following an oil spill.

2. INTRODUCTION

2.1 Overview

The Benthic Biodiversity Theme provides the first knowledge of benthic biodiversity structure (composition, distribution and standing stock) in the deep waters of the Great Australian Bight (GAB). It will enable this information to be built into ecosystem models for the GAB, and provides the basis for understanding what is necessary to monitor the potential future impacts of oil and gas exploration and/or production, as well as other activities, on the structure of deep benthic biological communities, and the scope to do this using leading-edge and cost-effective molecular techniques that are consistent with an Australian national standard. This project evaluates leading-edge molecular techniques to identify informative and cost-effective biodiversity metrics relevant to future ecological monitoring, while traditional approaches are covered in Project 3.1 (Williams et al. 2017).

Benthic communities are a critical component of both shallow and deep marine ecosystems as they utilise surface productivity, alter the physical and chemical condition of the sediment and sediment-water interface, and transfer energy to higher trophic levels. The biota (both eukaryotic and prokaryotic) which interact with the sediment contain a range of life-forms, complex reciprocal relationships, ecological niches and life-cycles, and exploit the environment in a myriad of ways. Many of these species also have a larval dispersal stage, which forms part of the complex pelagic assemblage present in the overlying water column. This overlying pelagic assemblage also contains the organisms that are responsible for the primary productivity of most deep ocean waters, and ultimately provides the food sources that the benthic assemblage requires, as well as forming the basis for much of the food chain that supports a wide range of economically and socially important species. Both natural and anthropogenic stressors can significantly alter the composition and functional attributes of benthic and pelagic environments. Whilst macrobenthic invertebrates have traditionally been the focus of benthic community structure monitoring programs, and the analysis of photosynthetic pigments is often used for pelagic monitoring, no single approach can accurately assess and monitor these complex environments. Furthermore, the microbial assemblages present in these habitats are almost completely unknown, and the links between taxonomic composition and ecological functioning in this group are poorly understood, although they play an important role in many of the ecological processes in both benthic and pelagic habitats. Because of the poor understanding of the links between taxonomy and process, understanding and monitoring ecosystem structure and function require separate tools.

Developing the key elements of an ecological monitoring program will require new and innovative approaches. This is because traditional taxonomic methods alone are unlikely to provide the required resolution and quantification of the largely unknown benthic fauna in the project timeframe, and because the distribution of hydrocarbons from natural seeps, and their influence on biological communities, is unknown. Accordingly, this project has refined and applied the latest molecular methods for monitoring ecological composition, whilst developing approaches for understanding the relationships between bacteria and degradation of hydrocarbons, a potential key driver for the system's resilience to oil spills, as well as a key indicator of anthropogenic influence.

These innovations have been developed and delivered through three sub-projects: 1) a molecular-based program which uses 'barcoding' to identify key taxonomic groups of the region (Chapter 3); 2) the development of a microarray for examining the microbially-mediated degradation of hydrocarbons (Chapters 4-6); and 3) the development of a rapid and reliable ecogenomic technique which can provide a broad characterisation of all microbial and eukaryote taxa present (Chapter 7). Ecogenomics are methods that allow the examination of genetic material from complex

environmental samples containing many different organisms. Ecogenomic methods have been effectively used for a wide range of taxa, localities and environs (i.e. sediments and water) such as Sydney Harbour sediments (Chariton et al. 2010), marine anoxic water (Stoeck et al. 2010), ecological characteristics of viruses (Roossinck et al. 2010) and biotic assemblages of soil DNA (Epp et al. 2012). The use of traditional surveillance methods (i.e. physical collection, sorting and taxonomic identification) is time consuming, expensive and places workload pressure upon Australia's dwindling taxonomic expertise. The development of molecular tools for ecological monitoring potentially allows for more complete, rapid and cost effective assessment of biodiversity and ecological shifts to be conducted (Baird & Hajibabaei 2012). These methods, once developed and validated, will be applicable to a range of marine habitats that are currently poorly understood. Collectively, it is envisaged that these three approaches will provide ecological end-points to reliably and rapidly assess system structure and function.

2.1.1 Sub-project 1: Barcoding

Species identification underpins any research in marine biology, being a fundamentally important process for assessments of biodiversity, community structure and ecosystem functioning. However, marine environments impose numerous challenges for species identification including difficulties with sampling key habitats and problems associated with the patchy distribution of many marine species. DNA barcoding has revolutionised the task of identification by providing reliable, inexpensive, and rapid ways to identify species, discover new ones and discriminate between cryptic species (i.e. discrete species that are difficult or impossible to distinguish morphologically) (Hebert et al. 2003; Marshall 2005; Meusnier et al. 2008). In this method, identification is performed by using DNA sequences from a small fragment of the genome (the "DNA barcode"), with the aim of contributing to a wide range of studies in which traditional taxonomic identification is not practical.

A large number of researchers have embraced DNA barcoding as a practical tool to be used in conjunction with traditional morphology-based studies. For instance, The International Barcode of Life project (iBOL, www.ibol.org) is the largest biodiversity genomics project ever undertaken. It includes hundreds of scientists from 25 nations who collaborate to construct a DNA-based identification system for all multi-cellular life. The barcode of life database (BOLD) contains over 6 million sequences from more than 262,000 species, and is continually expanding. DNA barcoding is now well established for animals and mostly based on sequences from the mitochondrial cytochrome c oxidase 1 (COI) gene (see www.barcodeoflife.org). The COI gene normally shows sequence variation that is high enough between species to discriminate them from one another, and low enough within species so researchers can distinguish intra- from inter-specific genetic variation (Herbert et al. 2003; Marshall 2005; Meusnier et al. 2008).

Barcoding is particularly useful where putative morphological species are actually complexes of cryptic species, allowing the number of true taxa present to be distinguished. Barcoding will also allow a match to be made between the traditional morphological taxonomy and the results of the ecogenomic biodiversity assessments.

2.1.2 Sub-project 2: Hydrocarbon degraders

Much of the oil that is released into the environment, whether by natural processes or anthropogenic activity, is degraded by bacteria. The populations of hydrocarbon degrading bacteria within natural communities increase following exposure to hydrocarbons from 1% in pristine environments to 10% in oil contaminated environments (Atlas 1981, 1995). Measures of the rates of

increase have traditionally relied on culture based techniques, but in the past decade, molecular analysis has become more prevalent (e.g. Hazen et al. 2010; He et al. 2007).

Understanding the function of a microbial community may be best approached by measuring the abundance of genes with known roles in key metabolic functions (He et al. 2010). Even though similar functions are often carried out by phylogenetically distinct members of the community, key genes involved in these processes are often conserved (Widada et al. 2002). Previous studies have measured the abundance of genes involved in hydrocarbon degradation to better understand the microbial communities involved in oil degradation in diverse environments including Chinese oil fields (Liang et al. 2011) or following the Deepwater Horizon oil spill in the Gulf of Mexico (Lu et al. 2012; Rivers et al. 2013). The findings of these studies have provided complementary data to those conducted using a phylogenetic approach (Lu et al. 2012).

To assess the potential of the indigenous microorganisms in the GAB to degrade hydrocarbons, we use an approach similar to previous studies (e.g. Rhee et al. 2004, He et al. 2010, Lu et al. 2012), and designed quantitative PCR based assays for genes involved in the degradation of aromatic compounds following a high throughput sequencing campaign. Genes enabling the utilisation of different hydrocarbons, specifically alkane hydroxylase, catechol dioxygenase and methane mono-oxygenase have been targeted (Inagaki et al. 2004, Rhee et al. 2004, Head et al. 2006, Valliancourt et al. 2006). By measuring the abundance of these and other functional genes, we can provide an understanding of the relative levels of hydrocarbons currently present in the GAB, an understanding of the capacity of indigenous bacteria to degrade hydrocarbons, and a tool for rapid response monitoring for any accidental releases of oil.

2.1.3 Assessment of baseline assemblages

To develop a molecular monitoring tool for the assessment of GAB sedimentary ecosystem changes we draw upon previous studies to determine the benthic eukaryotic assemblages (see Chariton et al. 2010, Stoeck et al. 2010, Bucklin et al. 2011, Lodge et al. 2012, Leray et al. 2013, Leray and Knowlton 2015) and supplement these with similar analysis of the microbial (prokaryotic) assemblages (Kirchman et al. 2010). We determine the baseline eukaryotic assemblages of benthic samples by High Throughput Sequencing (HTS) targeting the 18S rRNA gene and the 5' region of the mitochondrial cytochrome c oxidase 1 gene, COI-5P (as a metabarcoding approach). This potentially allows the examination of all eukaryotic biota (including metazoa, algae, fungi and protists), enabling the inclusion of biota generally excluded from monitoring programs. This approach complements the DNA barcoding of select GAB specimens, thereby enabling fine resolution sampling of some taxa, whilst encompassing the spatio-temporal patterns of a large portion of the organisms present within the system.

HTS microbiomic surveys of the prokaryotic (Bacteria and Archaea) assemblages will provide complementary data to the eukaryotic HTS and the microbial hydrocarbon degradation components of the study. These analyses will describe the global structural (taxonomic) diversity of benthic sediment microbial prokaryote communities. These are key drivers of the underlying biogeochemical processes and the most responsive communities to major disturbance events, and can be used as biological indicators for exploring changes in the underlying structural/functional capacity of an ecosystem. Community-based indicators are generally more sensitive and descriptive than indicators using single or few species.

2.2 Objectives

1. Provide an overview of the biodiversity of sedimentary assemblages in the GAB (sub-project 3).
2. Develop an effective and rapid molecular monitoring tool to detect any functional changes in the hydrocarbon-degrading microbial assemblages of the GAB (sub-project 2).
3. Assess the presence of hydrocarbon degrading microbes in the GAB, match these to chemical analyses of hydrocarbons in the sediments, and infer the potential of the assemblage to cope with any future increase in hydrocarbon levels (sub-project 2).
4. Undertake biodiversity identification using molecular barcoding of selected taxonomic groups of specific interest for future monitoring (sub-project 1).

2.3 References

- Atlas R. 1981. Microbial degradation of petroleum hydrocarbons: and environmental perspective. *Microbiological reviews* 45: 180-209.
- Atlas R.M., 1995. Petroleum biodegradation and oil spill bioremediation. *Marine Pollution Bulletin*. 31: 178-182.
- Baird D.J. & Hajibabaei M. 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*. 21: 2039-2044.
- Bucklin A., Steinke D. & Blanco-Bercial L. 2011. DNA barcoding of marine metazoa. *Annual Review of Marine Science* 3: 471-508.
- Chariton A.A., Court L.N., Hartley D.M., Colloff M.J. and Hardy C.M. 2010. Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Frontiers in Ecology* 8: 233-238.
- Epp L.S., Boesenkool S. et al. 2012. New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular Ecology* 21: 1821-1833.
- Hazen T.C., Dubinsky E.A., DeSantis T.Z., Andersen G.L., Piceno Y.M., Singh N., Jansson J.K., Probst A., Borglin S.E., Fortney J.L., Stringfellow W.T., Bill M., Conrad M.E., Tom L.M., Chavarria K.L., Alusi T.R., Lamendella R., Joyner D.C., Spier C., Baelum J., Auer M., Zemla M.L., Chakraborty R., Sonnenthal E.L., D'Haeseleer P., Holman H.Y., Osman S., Lu Z., Van Nostrand J.D., Deng Y., Zhou J., Mason O.U. 2010. Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science* 330:204-208.
- He Z., Deng Y., Van Nostrand J.D., Tu Q., Xu M., et al. 2010. GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *The ISME Journal* 4: 1167-1179.
- Head I.M., Jones D.M., Roling W.F.M. 2006. Marine microorganisms make a meal of oil. *Nature Reviews Microbiology* 4:173-182.
- Hebert P.D., Cywinska A., Ball S.L., deWaard J.R. 2003. Biological identifications through DNA barcodes. *Proceedings Biological Sciences* 270:313-321.
- Inagaki F., Tsunogai U., Suzuki M., Kosaka A., Machiyama H., Takai K., Nunoura T., Nealson K.H., Horikoshi K. 2004. Characterization of C-1-metabolizing prokaryotic communities in methane seep

- habitats at the Kuroshima Knoll, southern Ryukyu arc, by analyzing *pmoA*, *mmoX*, *mxoF*, *mcrA*, and 16S rRNA genes. *Applied and Environmental Microbiology* 70:7445-7455.
- Kirchman D.L., Cottrell M.T., Lovejoy C. 2010. The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environmental Microbiology* 12: 1132-1143.
- Leray M., Knowlton N. 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences USA* 112: 2076-2081.
- Leray M., Yang J.Y., Meyer C.P., Mills S.C., Agudelo N., Ranwez V., Boehm J.T., Machida R.J. 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology* 10:34.
- Liang Y., Van Nostrand J.D., Deng Y., He Z et al., 2011. Functional gene diversity of soil microbial communities from five oil-contaminated fields in China. *The ISME Journal* 5: 403-413.
- Lodge D.M., Turner C.R., Jerde C.L., Barnes M.A., Chadderton L., Egan S.P., Feder J.L., Mahon A.R., Pfrender M.E. 2012. Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology* 21: 2555-2558.
- Lu Z., Deng Y., Van Nostrand J.D., He Z., Voordeckers J., Zhou A., Lee Y-J., Mason O.U., Dubinsky E.A., Chavarria K.L., Tom L.M., Fortney J.L., Lamendella R., Jansson J.K., D'Haeseleer P., Hazen T.C., Zhou J. 2012. Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *The ISME Journal* 6:451-460
- Marshall E. 2005. Taxonomy. Will DNA bar codes breathe life into classification? *Science*, 307:1037.
- Meusnier I., Singer G.A.C., Landry J.F., Hickey D.A., Hebert P.D.N., Hajibabaei M. 2008. A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, 9:214.
- Rhee S.-K., Liu X., Wu L., Chong S.C., Wan X., Zhou J. 2004. Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Applied and Environmental Microbiology* 70: 4303-4317.
- Rivers A.R., Sharma S., Tringe S.G., Martin J., Joye S.B., Moran M.A. 2013. Transcriptional response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill. *The ISME journal* 7:2315-2329.
- Roossinck M.J., Saha P., Wiley G.B., Quan J., White J.D., Lai H., Chavarria F., Shen G., Roe B.A. 2010. Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Molecular Ecology* 19 (Supp. 1): 81-88
- Stoeck T., Bass D., Nebel M., Christen R., Jones M.D.M., Breiner H., Richards T.A. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology* 19 (Supp. 1): 21-31
- Valliancourt F.H., Bolin J.T., Eltis L.D. 2006. The ins and outs of ring-cleaving dioxygenases. *Critical Reviews in Biochemistry and Molecular Biology*. 41:241-267.

Widada J., Nojiri H., Kasuga K., Yoshida T., Habe H., Omori T. 2002. Molecular detection and diversity of polycyclic aromatic hydrocarbon-degrading bacteria isolated from geographically diverse sites. *Applied Microbiology and Biotechnology* 58:202-209.

Williams A., Tanner J.E., Althaus F., Sorokin S.J., MacIntosh H., Green M., Brodie P., Loo M. 2017. Great Australian Bight Benthic Biodiversity Characterisation. Final Report GABRP Project 3.1. Great Australian Bight Research Program, GABRP Research Report Number 19, 361pp.

3. MOLECULAR BARCODING OF SELECTED BENTHIC TAXA

Minami Sasaki and Luciano B. Beheregaray

School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide, SA. 5001.

Reprinted with minor amendments from:

Sasaki M, Beheregaray LB (2016) Molecular assessment of benthic and pelagic biodiversity in the Great Australian Bight: Barcoding. Great Australian Bight Research Program, December 2016, 43pp.

3.1 Executive summary

Species identification underpins any research in marine biology, being a fundamentally important process for assessments of biodiversity, community structure and ecosystem functioning. The aim of this sub-project was to use DNA barcoding to develop selective genetic datasets to complement and advance morphological taxonomy within informative benthic groups. Benthic samples from two field surveys (2013 and 2015) in the Great Australian Bight (GAB) were transferred to Flinders University's Molecular Ecology Laboratory, where all analyses and results reported here were carried out. The DNA extracted from these samples was used to amplify, via the polymerase chain reaction (PCR), the cytochrome c oxidase subunit 1 (COI) region of the mitochondrial genome that has been used as a DNA barcode for species identification. We optimised and developed several protocols for DNA extraction and amplification of the COI gene across a range of marine phyla from the GAB region. We also implemented analytical approaches to circumvent the lack of information about morphological identification, which allowed the proposal of thresholds of COI sequence similarity to identify samples at the species, genus and/or family levels. The identifications corresponded to 139 Arthropoda, 106 Echinodermata, 27 Mollusca, 14 Annelida, 11 Sipuncula and 6 Cnidaria. Identifications are provided for each individual sample as a series of barcode IDs and sequence data. We also found that species identification was substantially improved by using both the National Center of Biotechnology Information (NCBI) and Barcode of Life Database (BOLD) sequence databases. Finally, according to the species distribution records from the Atlas of Living Australia and BOLD, we found three species not previously documented in Australia: *Sperosoma biserialatum* (Echinodermata), *Granulifusus nopinicus* (Mollusca) and *Ebalia nux* (Arthropoda). We propose that the COI region should be considered as an appropriate and cost-effective tool for biodiversity assessment in future ecological monitoring in the GAB region, either as individual sequencing approaches, or using next-generation sequencing platforms. Information from morphological identification and species geographic ranges should be incorporated into future studies to improve the power of DNA barcoding for species identification in the GAB.

3.2 Introduction

The overall aim of the Benthic Biodiversity Theme is to provide the first knowledge of benthic biodiversity structure (composition, distribution and standing stock) in the deep waters of the Great Australian Bight (GAB). The sub-project 1 'Barcoding' is a molecular-based program that uses "DNA barcodes" to identify key taxonomic groups in the GAB region. It provides a contribution to the evaluation of leading-edge molecular techniques to identify informative and cost-effective biodiversity metrics relevant to future ecological monitoring.

Species identification underpins any research in marine biology, being a fundamentally important

process for assessments of biodiversity, community structure and ecosystem functioning. However, marine environments impose numerous challenges for species identification including difficulties sampling key habitats and problems arising from the patchy distribution of many marine species. DNA barcoding has revolutionised the task of identification by providing reliable, inexpensive, and rapid ways to identify known species, discover new ones and discriminate between cryptic species (i.e. discrete species that are difficult or impossible to distinguish morphologically) (Hebert et al. 2003; Marshall 2005; Meusnier et al. 2008). In this method, identification is performed by using DNA sequences from a small fragment of the genome (the “DNA barcode”), with the aim of contributing to a wide range of studies in which traditional taxonomic identification is impractical.

A large number of researchers have embraced DNA barcoding as a practical tool to be used in conjunction with traditional morphology-based studies. For instance, [The International Barcode of Life project](http://www.ibol.org) (iBOL, www.ibol.org) is the largest biodiversity genomics project ever undertaken. It includes hundreds of scientists from 25 nations who collaborate to construct a DNA-based identification system for all multi-cellular life. DNA barcoding is now well established for animals, and is based on sequences from the mitochondrial cytochrome c oxidase 1 (COI) gene (see www.barcodeoflife.org). The COI gene normally shows sequence variation that is high enough between species to discriminate one from another, and low enough within species to distinguish intra- from inter-specific genetic variation (Herbert et al. 2003; Marshall 2005; Meusnier et al. 2008).

3.2.1 Aim

The aim of this sub-project is to use DNA barcoding to develop selective genetic datasets to complement and advance morphological taxonomy within informative benthic groups.

3.2.2 Outputs and extensions

The main output of this sub-project will be a series of barcodes for selected taxa of interest, which can then be used to form the basis of the molecular monitoring sub-project. In addition, we expect the major findings of the Barcoding sub-project to be published in leading journals in marine science.

3.3 Methods

3.3.1 Sampling

All tissue samples required for this project were provided by Project 3.1 (Williams et al. 2017) as part of two GAB surveys referred herein as ‘Survey 2013’ and ‘Survey 2015’. Samples from ‘Survey 2013’ were divided into groups of ‘infauna’ and ‘epifauna’ organisms, whereas all samples from ‘Survey 2015’ were ‘epifauna’ organisms. All samples came from either sediment cores or beam trawls undertaken at 6 depth horizons (200, 400, 1000, 1500, 2000 & 3000 m) along 5 north-south transects in the central and eastern GAB. The majority of infaunal specimens collected in 2013 were made available for barcoding, with the exception of a small number of samples considered too important to destroy (they were new species, or exceptionally good examples of species poorly represented in museum collections). For epifauna, either tissue samples or for smaller individuals, whole specimens, were collected primarily for abundant species, with some unusual specimens also targeted. Where possible, for each individual trawl, 5 specimens of each taxon were sampled. Logistics prevented comprehensive sampling of all taxa sampled, and the limited knowledge of what was likely to be present in the trawls prior to sampling commencing meant that target taxa had to be selected based on what turned up in the trawls. Samples were frozen aboard and then preserved in ethanol (>95%) prior to being transferred to the Molecular Ecology Lab at Flinders University (MELFU) for DNA barcoding.

3.3.2 Laboratory procedures

Most of the infauna samples were very small in size (~1-2 mm), whereas the majority of the epifauna samples were substantially larger (>5 mm). Samples were previously identified at the phylum level, and mostly included Arthropoda, Echinodermata, Annelida and Mollusca. Other phyla included Sipuncula, Porifera, Chordata (Ascidiacea), Cnidaria and Brachiopoda. Due to the high diversity of tissue types and sizes, we had to carry out optimisation of DNA extractions for each phylum using a variety of methods and conditions. We tested several protocols including DNeasy blood and tissue kit (Qiagen), CTAB, salting out protocol (Sunnucks & Hales 1996), Chelex 100 (Bio-Rad) and Vertebrate Lysis Buffer protocol (VLB) (Ivanova et al. 2006). The details about the methods and modifications of each DNA extraction protocol are presented in Appendix 3.1. DNA integrity was assessed by gel electrophoresis, and quality and quantity were measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific). The total DNA was then used to amplify, via the polymerase chain reaction (PCR), the 605 base pairs of the COI region of the mitochondrial genome that has been used as DNA barcode for most species identification projects (Hebert et al. 2003; Ward et al. 2005; Bucklin et al. 2011). Several primer combinations and PCR conditions were trialled (Table 3.1). The 10 µL PCRs consisted of 2 µL of 5x PCR buffer, 0.25 mM of dNTPs, 1.5 mM of MgCl₂, 0.3 µM of each primer, 0.2 mg/mL of BSA (NEB), 1 U of Mango Taq (Bioline), 1 µL of template DNA and ddH₂O. Resulting PCR products were visualised on 1% agarose gel using 1x TAE buffer (40 mM Tris (pH 7.6), 20 mM acetic acid and 1mM EDTA), with 1 µL of each reaction. Successful amplifications were diluted in ddH₂O to approximately 100 ng/µL if necessary and unidirectionally sequenced in an ABI3730XL (Applied Biosystems) automated sequencer at Macrogen, Korea. Primers used for sequencing correspond to those used for PCR and include LCO1490, LCOechi1aF1, LoboF1 and jgLCO.

Table 3.1. PCR cycles and primer combinations used to amplify the COI region for GAB samples.

LCO1490 x HCO2198			LCOechi1aF1 x HCO2198			jgLCO x jgHCO		
94C	1 min		94C	1 min	x1	94C	1min	x1
40-47C	1 min	x35-40	94C	30 sec		94C	30sec	
72C	1 min		45C	90 sec	x5	45C	90sec	x5
72C	10 min	x1	72C	1 min		72C	1min	
LoboF1 x LoboR1			94C	30 sec		94C	30sec	
94C	2 min	x1	54C	90 sec	x45	54C	90sec	x45
94C	30 sec		72C	1 min		72C	1min	
47C	90 sec	x45	72C	5 min	x1	72C	5min	x1
72C	1 min							
72C	5 min	x1						

3.3.3 Data analysis

Resultant sequences were visually inspected in GENEIOUS 10.0.5 (Kearse et al. 2012) and the data quality using Phred score was also checked in CODONCODE ALIGNER v.6.0.2 (CodonCode Aligner, Centerville, MA, USA). Any bases with Phred score of less than 20 were considered as missing data (i.e. N). The presence of a stop codon was examined with MESQUITE v.3.10 (Maddison & Maddison 2016) and any sequences with intermediate stop codons along the fragment were removed from the dataset. Sequences from the same phylum were aligned in MEGA 7.0.21 (Kumar et al. 2016) to check for the presence of indels and to remove the primer region. Clean and aligned sequences were then blasted using the NCBI BLAST (Altschul et al. 1990) based on default settings with expected threshold of 10 and match/mismatch score of 0.5 for possible contaminations. These sequences were also submitted to the animal identification system (IDS) implemented in BOLD (Ratnasingham & Hebert 2007) to obtain the identification of organisms that are the nearest match to our samples. Because of the lack of morphological ID for the barcoded samples, the highest similarity ID from either BLAST search or IDS from BOLD, regardless of percentage, was used to name our samples. Sequences with >500 bp length and with less than 1% of missing data were kept to compile a final data set.

To improve sample identification we used a direct sequence comparison approach known as the best close match (BCM), with support from phylogenetic trees. The BCM was proposed by Meier et al. (2006), who compared the success rate of species identification based on phylogenetic trees and direct sequence comparison. These authors found that direct sequence comparison had better identification and a relatively low misidentification. BCM involves calculating the frequency of intraspecific uncorrected distance (p-distance) to find the threshold value below 95% where all intraspecific distances can be found. If sequences do not have any barcode match even below the threshold value, they will remain unidentified. The remaining sequences are then compared to those of their closest match. Identification is considered successful if the taxon name was identical while identification fails when the taxon differs. If there are more than two species that match below the threshold, identification is considered ambiguous. This method can be used when 1) you have replicates from species and genus, and 2) a reference sequence that is closely related to your samples.

Since we did not have enough replicates at species and genus levels, it was not feasible to use our dataset to calculate intraspecific distance. Instead, we used a different method and calculated intraspecific distance in TAXON DNA/SPECIES IDENTIFIER 1.8 (Meier et al. 2006). We obtained COI sequences from NCBI GenBank and BOLD representing both Echinodermata (genera *Amphiophiura*, *Benthopecten*, *Mellita* and *Meridiastra*) and Arthropoda (genera *Acanthephyra*, *Ebalia*, *Eumunida*, *Liocarcinus* and *Periclimenes*). These phyla were chosen since most of our samples were represented by Echinodermata and Arthropoda. Sequences accessed from GenBank and BOLD corresponded to the same genera as those in our dataset with the aim of obtaining genetic distances from similar organisms. The number of species per genus ranged from 4 to 6 and the number of replicates per species ranged from 2 to 10 (Table 3.2). Because intraspecific distance varies among taxa, we averaged distances below 95% from 4 Echinodermata genera and 5 Arthropoda genera so that thresholds could be applied to our dataset. This was also done for samples with only few available conspecific sequences as they cannot be used in Taxon DNA/Species Identifier. The lowest average interspecific distance and average congeneric distance calculated from Taxon DNA/Species Identifier were used as thresholds to identify genus and family, respectively. We then used 15% and 21% for

classifying order and class, respectively, based on visual examinations of NCBI GenBank and BOLD results.

Uncorrected pairwise genetic distances (p-distance) were calculated between our samples and the closest match sequences in MEGA 7.0.21 and the threshold was applied to the pairwise p-distance to identify the taxonomic level. In order to confirm the results of identification using BCM, we also constructed Neighbour-Joining trees (NJ) based on Kimura-2-parameter with 1000 bootstraps. This was only done for species with available congeneric sequences. Finally, the known distributions of species identified in this study was checked against the Atlas of Living Australia (Atlas 2015) and BOLD (Ratnasingham & Hebert 2007).

3.4 Results and Discussion

3.4.1 Optimisation of laboratory procedures

Although different protocols yielded varying levels and qualities of DNA extractions for different phyla, the salting out protocol yielded the highest concentration of cleaner DNA (260/280 and 230/280 NanoDrop measurements) for all major phyla such as Arthropoda, Annelida, Echinodermata and Mollusca (Table 3.3). These results were more apparent for Echinodermata, except for the class Holothuroidea of which mucous slime appeared during the DNA precipitation step hindering the success of extraction and thus affecting downstream application. We found no substantial differences in DNA extraction based on the way tissues were preserved (frozen or ethanol). The optimisation of DNA extraction protocols was achieved for all phyla, except for Porifera (DNA yield was much better for denser morphological species than coarser ones) and Foraminifera.

Table 3.2. Samples used to calculate intraspecific diversity.

Bold **G** indicates that sequences were sourced from GenBank and bold **B** indicates that sequences were sourced from BOLD.

Echinodermata			
Genus	Species	Sample ID	Comments
<i>Amphiophiura</i>	<i>Amphiophiura laudata</i>	EU869873.1 G	
		EU869874.1 G	
		EU869875.1 G	
		EU869876.1 G	
		KU894936.1 G	
	<i>Amphiophiura spatulifera</i>	KU894926.1 G	
		KU894927.1 G	
	<i>Amphiophiura superba</i>	HM400324.1 G	
		HM400325.1 G	
		HM400326.1 G	
		HM400327.1 G	
	<i>Amphiophiura urbana</i>	EU869877.1 G	
		EU869878.1 G	
		EU869879.1 G	
		EU869881.1 G	
		EU869880.1 G	
<i>Benthopecten</i>	<i>Benthopecten acanthonotus</i>	HM542917.1 G	
		HM542918.1 G	
		HM542919.1 G	
		HM542921.1 G	
		HM542920.1 G	
	<i>Benthopecten claviger</i>	HM542922.1 G	
		HM542923.1 G	
		HM542924.1 G	
	<i>Benthopecten munidae</i>	NZEC431-09 B	
		NZEC433-09 B	
		NZEC440-09 B	
	<i>Benthopecten pikei</i>	NZEC439-09 B	
		NZEC441-09 B	
		NZEC442-09 B	
<i>Mellita</i>	<i>Mellita grantii</i>	KF204793.1 G	
		KF204794.1 G	
		KF204795.1 G	
		KF204796.1 G	
		KF204797.1 G	
	<i>Mellita longifissa</i>	KF204798.1 G	
		KF204799.1 G	
		KF204800.1 G	
		KF204807.1 G	
		KF204830.1 G	

<i>Mellita</i>	<i>Mellita notabilis</i>	KF204749.1 G	They are grouped together for analysis as their p-dis ranged from 0 - 0.72%. <i>M. kanakoffi</i> was suggested to be junior synonym of <i>M. notabilis</i> (Coppard et al. 2013).
		KF204756.1 G	
		KF204757.1 G	
		KF204758.1 G	
		KF204759.1 G	
	<i>Mellita kanakoffi</i>	KF204773.1 G	
		KF204774.1 G	
		KF204775.1 G	
		KF204776.1 G	
		KF204777.1 G	
	<i>Mellita quinquiesperforata</i>	KF204724.1 G	They are grouped together for analysis as their p-dis ranged from 0 - 0.72%. <i>M. isometra</i> was suggested to be synonymise to <i>M. tenuis</i> (Coppard et al. 2013).
		KF204725.1 G	
		KF204726.1 G	
		KF204727.1 G	
		KF204728.1 G	
	<i>Mellita tenuis</i>	KF204729.1 G	
		KF204857.1 G	
		KF204859.1 G	
		KF204842.1 G	
		KF204858.1 G	
<i>Meridiastra</i>	<i>Mellita isometra</i>	KF204860.1 G	
		KF204848.1 G	
		KF204849.1 G	
		KF204850.1 G	
		KF204851.1 G	
	<i>Meridiastra calcar</i>	KF204852.1 G	
		EU869944.1 G	
		EU869945.1 G	
		EU869946.1 G	
		EU869947.1 G	
	<i>Meridiastra gunnii</i>	EU869948.1 G	
		AY458432.1 G	
		AY458434.1 G	
		AY458436.1 G	
		AY458437.1 G	
	<i>Meridiastra medius</i>	AY458438.1 G	
		AY458477.1 G	
		AY458473.1 G	
		AY458474.1 G	
		AY458476.1 G	
	<i>Meridiastra occidens</i>	AY458501.1 G	
		AY458505.1 G	
		AY458502.1 G	
		AY458503.1 G	
		AY458504.1 G	
	<i>Meridiastra oriens</i>	AY458450.1 G	
		AY458451.1 G	
		AY458452.1 G	

		AY458453.1 G	
		AY458455.1 G	
		Arthropoda	
Genus	Species	Sample ID	Comment
<i>Acanthephyra</i>	<i>Acanthephyra eximia</i>	KP759360.1 G	
		KP759359.1 G	
		KP759358.1 G	
		KP759357.1 G	
		KP759356.1 G	
	<i>Acanthephyra curtirostris</i>	KP076164.1 G	
		KP076163.1 G	
		KP076162.1 G	
		KP076161.1 G	
		GU183785.1 G	
	<i>Acanthephyra pelagica</i>	KP759361.1 G	
		KP076182.1 G	
		KP076179.1 G	
		KF930998.1 G	
		JQ305961.1 G	
	<i>Acanthephyra purpurea</i>	KP076173.1 G	
		KP076171.1 G	
		KP076170.1 G	
		GU183787.1 G	
	<i>Acanthephyra quadrispinosa</i>	KP759362.1 G	
		KP759363.1 G	
		KP076178.1 G	
<i>Ebalia</i>	<i>Ebalia cranchii</i>	KT209387.1 G	
		KT209435.1 G	
		KT209457.1 G	
		KT209478.1 G	
		KT209497.1 G	
		KT209523.1 G	
	<i>Ebalia nux</i>	JQ348856.1 G	
		JQ348859.1 G	
		JQ348860.1 G	
	<i>Ebalia tuberosa</i>	KT208846.1 G	
		KT209089.1 G	
		KT209151.1 G	
		KT209162.1 G	
		KT209479.1 G	
	<i>Ebalia tumefacta</i>	KT208981.1 G	
		KT208779.1 G	
		KT209285.1 G	
		KT209393.1 G	
		KT209463.1 G	
<i>Eumunida</i>	<i>Eumunida annulosa</i>	EU243354.1 G	
		EU243403.1 G	
		EU243408.1 G	

		EU243461.1 G
	<i>Eumunida capillata</i>	EU243341.1 G
		EU243342.1 G
		EU243343.1 G
		EU243344.1 G
	<i>Eumunida keijii</i>	EU243337.1 G
		EU243338.1 G
		EU243340.1 G
		EU243514.1 G
	<i>Eumunida minor</i>	EU243502.1 G
		EU243548.1 G
		EU243550.1 G
		EU243551.1 G
	<i>Eumunida parva</i>	EU243345.1 G
		EU243346.1 G
		EU243518.1 G
		EU243521.1 G
		EU243522.1 G
	<i>Eumunida spinosa</i>	EU243500.1 G
		EU243501.1 G
		EU243513.1 G
		EU243533.1 G
<i>Liocarcinus</i>	<i>Liocarcinus pusillus</i>	GQ268534.1 G
		KT209364.1 G
		KT208625.1 G
	<i>Liocarcinus depurator</i>	GQ268531.1 G
		KT209206.1 G
		KT209341.1 G
		KT209461.1 G
		KT209530.1 G
	<i>Liocarcinus holsatus</i>	GQ268538.1 G
		KT208798.1 G
		KT208821.1 G
		KT209279.1 G
		KT209350.1 G
	<i>Liocarcinus marmoreus</i>	GQ268535.1 G
		KT208606.1 G
		KT209205.1 G
		KT209214.1 G
		KT209329.1 G
		KT209516.1 G
	<i>Liocarcinus navigator</i>	cGQ268537.1 G
		KT208849.1 G
		KT208868.1 G
		KT209223.1 G
		KT209424.1 G
		KT209515.1 G
<i>Periclimenes</i>	<i>Periclimenes imperator</i>	GQ415634.1 G

	GQ415635.1 G
	GQ415636.1 G
	GQ415637.1 G
<i>Periclimenes rathbunae</i>	KX090123.1 G
	KX090120.1 G
	KX090121.1 G
	KX090122.1 G
	KX090124.1 G
	KX090125.1 G
<i>Periclimenes soror</i>	GQ415629.1 G
	GQ415630.1 G
	GQ415631.1 G
	GQ415632.1 G
	GQ415633.1 G
<i>Periclimenes thermohydrophilus</i>	AB298102.2 G
	AB298103.2 G
	AB298104.2 G
	AB298105.2 G
	AB298106.2 G
<i>Periclimenes yucatanicus</i>	KU065010.1 G
	KX858822.1 G

Table 3.3. Results of DNA extractions and corresponding protocol for each phylum.

O: satisfactory result, X: poor result, Δ: inconsistent result (protocol works for some samples but not all), NA: protocol was not used for the phylum, D¹: indicates denser morphotype Porifera and C²: indicates coarser morphotype Porifera.

DNA protocol	Amphipoda	Echinodermata	Polychaeta	Mollusca	Porifera
DNeasy	O	Δ	O	O	D ¹ : Δ / C ² :X
CTAB	NA	Δ	NA	NA	X
Salting out	O	O	O	O	NA
Chelex	X	X	X	NA	X
VLB	NA	X	NA	NA	X

The combination of primers LCO1490 and HCO2198 (Folmer et al. 1994) was the most effective across the taxonomic diversity of our sample. For Echinodermata, a different forward primer, LCOech1aF1, together with HCO2198 was used for most samples, whereas a combination of LoboF1 and LoboR1 (Lobo et al. 2013) was used for several Arthropoda and Cnidaria, and a combination of jgLCO and jgHCO (Geller et al. 2013) was used for Porifera. We also tested combinations of polyLCO and polyHCO (Carr et al. 2011), MinibarF1 and MinibarR1 (Meusnier et al. 2008) and COLceF and

COLceR (Hoareau & Boissin 2010) during optimisation but results were inconsistent and they were not used to produce a final dataset (Table 3.4).

The PCR amplifications were also improved as different primer sets were tested beside the better performing Folmer primers. For example, we observed signs of amplification for Echinodermata with Folmer primers but resultant sequences were mostly contaminated with teleost (i.e. fish) DNA. Examples of source of teleost contamination included *Cetonus globiceps* and *Coelorinchus acanthiger*. These species are found along the GAB and were sampled in the same round of bottom trawling as the samples used for this project. We propose that echinoderms absorbed the mucoprotein coating of these teleosts before they were individually processed, which accounts for the high concentration of foreign DNA in the tissue of our target samples. A range of dissecting methods and tissue sources were attempted from the echinoderm samples with inconsistent results. In the best cases we needed 3-4 rounds of processing (i.e. dissecting, DNA extraction, PCR and sequencing) for each echinoderm sample, as opposed to one round as initially planned, in order to obtain a non-teleost ID for the echinoderm. We also tested whether DNA quantity or a combination of primers was more important for amplification of these targeted samples. The LCOech1aF1 x HCO2198 combination successfully amplified Echinodermata sequences while the Folmer primers often continued to amplify fish DNA when using the same DNA template.

None of the sequences obtained from Brachiopoda and Chordata (Ascidiacea) were from target species, suggesting that none of the primers were suitable for these phyla. Another recurrent issue we experienced relates to a low rate of amplification success for the 2013 infauna samples. For instance, the PCR protocol optimised for the 2015 epifauna Arthropoda failed to amplify many of the infauna Arthropoda (Table 3.4), an issue attributed to the very small size of these samples.

Table 3.4. Summary of efforts on DNA extraction, PCR and sequencing.

¹ - Total number of samples received from two surveys, ² - Total number of unique samples that were sequenced regardless of quality, ³ - Total number of DNA extractions carried out for a given phylum, ⁴ - Total number of PCRs carried out for a given phylum, ⁵ - The number of PCRs carried out for each primer combination, ⁶ - Total number of sequences obtained (including replicates) and ⁷ - Total number of sequences in the final dataset for a given phylum (after quality control).

‘Survey 2013’ infauna samples							
Phylum	Total # ¹	Seq # ²	All DNA # ³	All PCR # ⁴	# of PCR for each primer combination ⁵ 140 (Folmer), 4 (Minibar), 7 (jgLCO), 14 (Lobo), 67 (polyLCO)	All seq # ⁶	Final seq # ⁷
Annelida	109	6	82	232		15	0
Arthropoda	59	4	52	255	190 (Folmer), 59 (Lobo), 3 (jgLCO), 3 (Minibar)	20	0
Echinodermata	13	0	11	38	12 (Folmer), 24 (LCOechi1aF1), 2 (Lobo)	0	0
Miscellaneous	17	0	10	15	15 (Folmer)	0	0
TOTAL	198	10	155	540		35	0
‘Survey 2013’ epifauna samples							
Phylum	Total # ¹	Seq # ²	All DNA # ³	All PCR # ⁴	# of PCR for each primer combination ⁵	All seq # ⁶	Final seq # ⁷

Annelida	5	2	8	17	17 (Folmer)	3	1
Arthropoda	12	12	19	24	24 (Lobo)	15	10
Cnidaria	1	1	2	1	1 (jgLCO)	1	0
Echinodermata	34	16	85	187	32 (COLceF), 12 (LCOechi1aF1), 141 (Folmer), 2 (Lobo)	39	12
Mollusca	6	5	7	12	12 (Folmer)	4	4
Porifera	19	4	42	163	94 (jgLCO), 63 (Folmer), 6 (Lobo)	7	0
Sipuncula	2	1	3	4	4 (Folmer)	3	1
TOTAL	79	41	166	408		72	28

'Survey 2015'

	All		All			All seq	Final
	Total	Seq	DNA	PCR			
Phylum	# ¹	# ²	# ³	# ⁴	# of PCR for each primer combination ⁵	# ⁶	seq # ⁷
Annelida	19	13	28	33	29 (Folmer), 4 (Lobo)	15	13
Arthropoda	153	145	156	196	168 (Folmer), 28 (Lobo)	155	129
Brachiopoda	14	12	28	28	28 (Folmer)	15	0
Chordata	10	8	20	44	34 (Folmer), 10 (Lobo)	14	0
Cnidaria	12	11	13	43	17 (Folmer), 3 (jgLCO), 33 (Lobo)	12	6
Echinodermata	275	100	299	593	85 (Folmer), 2 (COLceF), 506 (LCOechi1aF1)	154	93
Mollusca	45	32	43	65	65 (Folmer)	33	23
Sipuncula	16	12	16	35	35 (Folmer)	25	11
Dendrogramma	3	2	3	3	3 (Folmer)	2	0
TOTAL	360	335	606	1040		425	275

SUMMARY

	All		All			All seq	Final
	Total	Seq	DNA	PCR			
Phylum	# ¹	# ²	# ³	# ⁴	# of PCR for each primer combination ⁵	# ⁶	seq # ⁷
Annelida	133	21	118	282		33	14
Arthropoda	224	161	227	475		190	139
Brachiopoda	14	12	28	28		15	0
Chordata	10	8	20	44		14	0
Cnidaria	13	11	13	43		12	6
Dendrogramma	3	2	3	3		2	0
Echinodermata	322	116	395	818		193	105
Miscellaneous	17	0	10	15		0	0
Mollusca	51	37	50	77		37	27
Porifera	19	4	42	163		7	0
Sipuncula	18	13	19	39		28	12
TOTAL	824	385	925	1987		531	303

3.4.2 Summary statistics

From the total number of samples received (n=824), we carried out 925 DNA extractions, 1,987 rounds of PCRs and obtained 531 sequences. These numbers include replication due to issues with teleost contamination and minute size of tissue samples (described above). After quality control, we obtained a total of 303 high quality sequences, which correspond to 139 Arthropoda, 106 Echinodermata, 27 Mollusca, 14 Annelida, 11 Sipuncula and 6 Cnidaria (Figure 3.1). The aligned sequences from each phylum were trimmed to 565 bp (all Arthropoda), 563 bp (all Cnidaria), 535 bp (three classes of Echinodermata), 556 bp (all Mollusca), 587 bp (all Annelida) and 533 bp (all Sipuncula). The aligned sequences from the Echinodermata class Holothuroidea were trimmed separately from other Echinodermata because many of the available Holothuroidea sequences were approximately 200 bp different in length compared to the standard barcode region. Sequence lengths for Holothuroidea samples ranged from 434 to 605 bp. Data from the following phyla were not available for identification due to low annealing during PCR (Brachiopoda and Chordata), due to the presence of a stop codon approximately in the middle of the fragment (Porifera) or due to an excess of missing data.

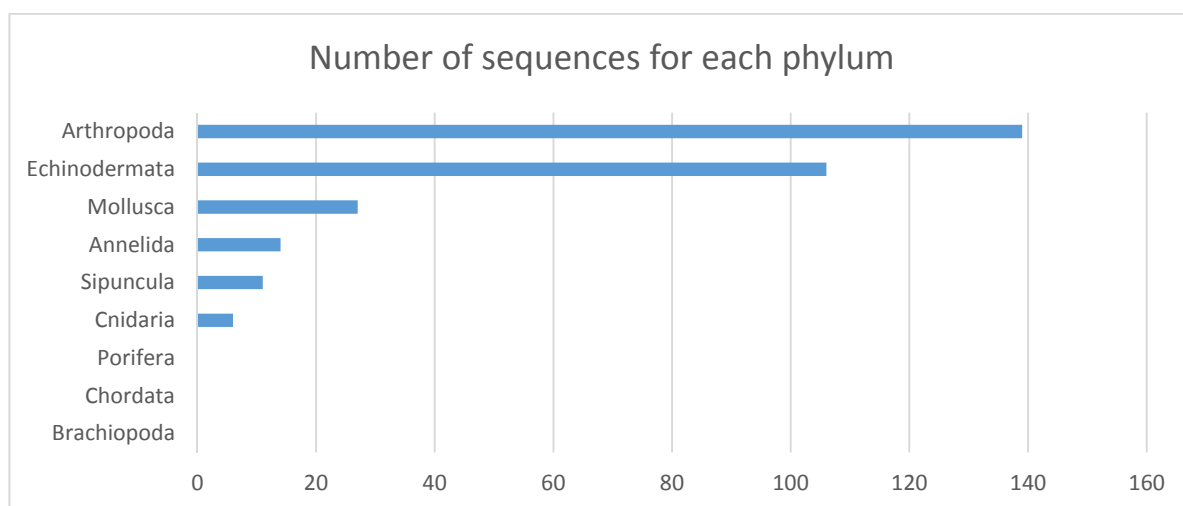


Figure 3.1. The number of sequences in the final data set for each phylum.

Searches conducted in Blast returned species ID with similarity percentages ranging from 79% to 100% (see Appendix 3.2 for IDs for each sample). Uncorrelated pairwise distances between our sample sequences and GenBank sequences ranged from 0% to 20.54%. Our estimates of average sequence distance based on five Arthropoda and four Echinodermata genera indicated that intraspecific distances ranged from 0% to 1.65%, interspecific distance were up to 10.36% and intrageneric genetic distances were up to 12.9%. These thresholds, based on average sequence distances, were then used to identify species, genus, and family, respectively. They successfully identified a number of species, genera and families for Echinodermata, Mollusca, Arthropoda and Cnidaria. All Annelida and Sipuncula samples were only identified at the class level (Figure 3.2).

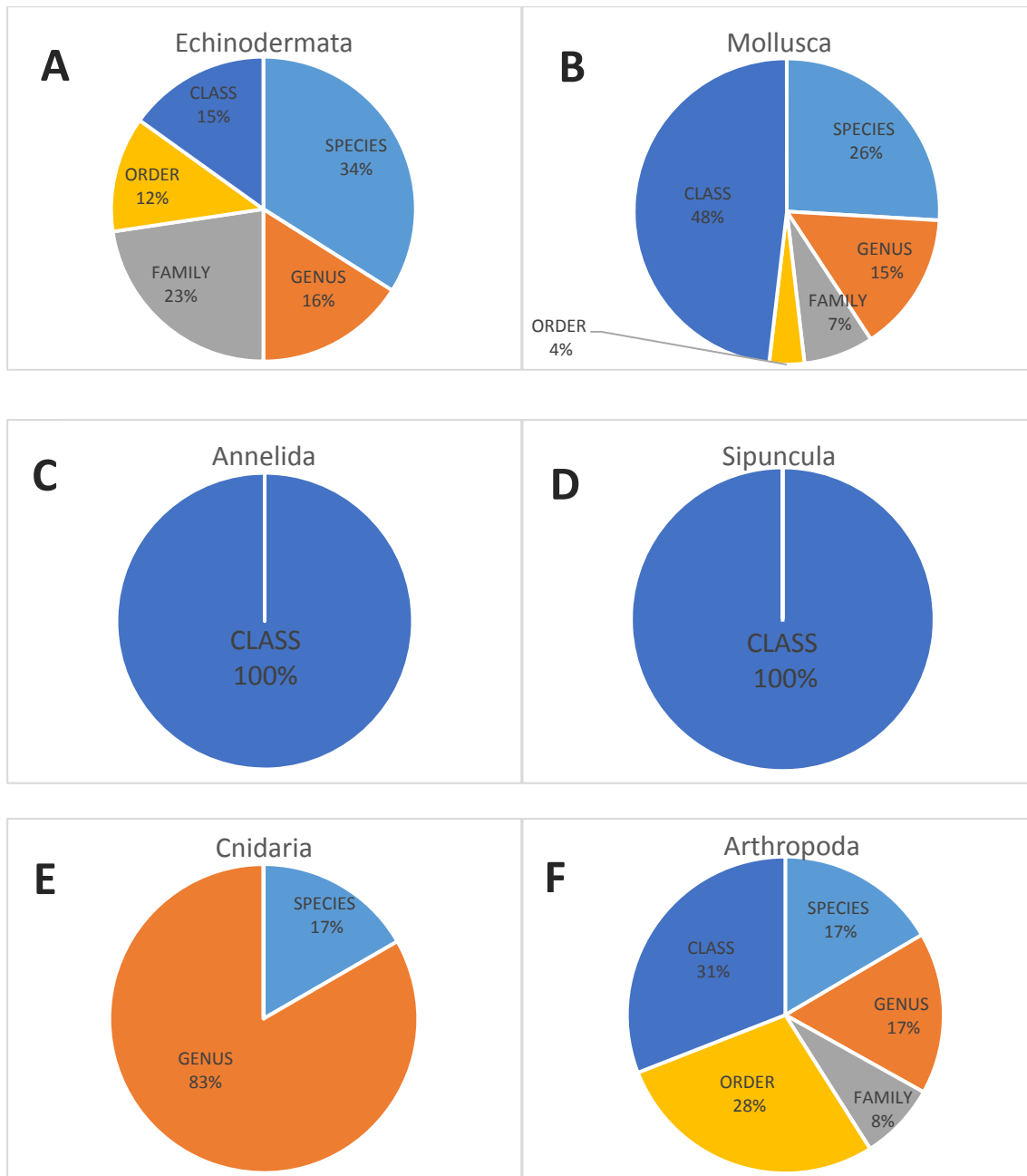


Figure 3.2. The proportion of sample identification for a given taxonomic level using p-distance thresholds (species: 0.00 – 1.65%, genus: 1.66 – 10.36%, family: 10.37 – 12.90%, order: 12.91 – 15.00% and class: 15.01 – 21.00%).

3.4.3 Proof of concept of identification strategy

The proof of concept for our identification strategy and resulting thresholds is exemplified below using a Neighbour-Joining (NJ) tree that we constructed for the Echinodermata genus *Ophiomusium* (Figure 3.3). Results from additional NJ trees for the other genera are shown in Appendix 3.3. The tree shows that most of our samples with species level identification clustered within the same clade with samples from the database (NBCI and BOLD), whereas the samples with genus level identification were clustered with two different species (*O. scalare* and *O. asperum*). Interestingly, pairwise p-distance between *O. asperum* and *O. scalare* ranged from 0.92% to 1.48%, which

probably means that these samples were obtained from the same species under this BCM threshold. Using additional loci and coalescent-based analytical approaches, such as species delimitation methods (e.g. Yang 2015), is recommended to test for species boundaries and provide stronger delimitation support for such situations. Although our *O. anisacanthum* sample (ABTC132345, red circle) was identified only to the genus level using BCM threshold ($p\text{-dis} = 2.71\%$), it did form a single and discrete clade that only contained other samples identified as *O. anisacanthum* (Figure 3.3).

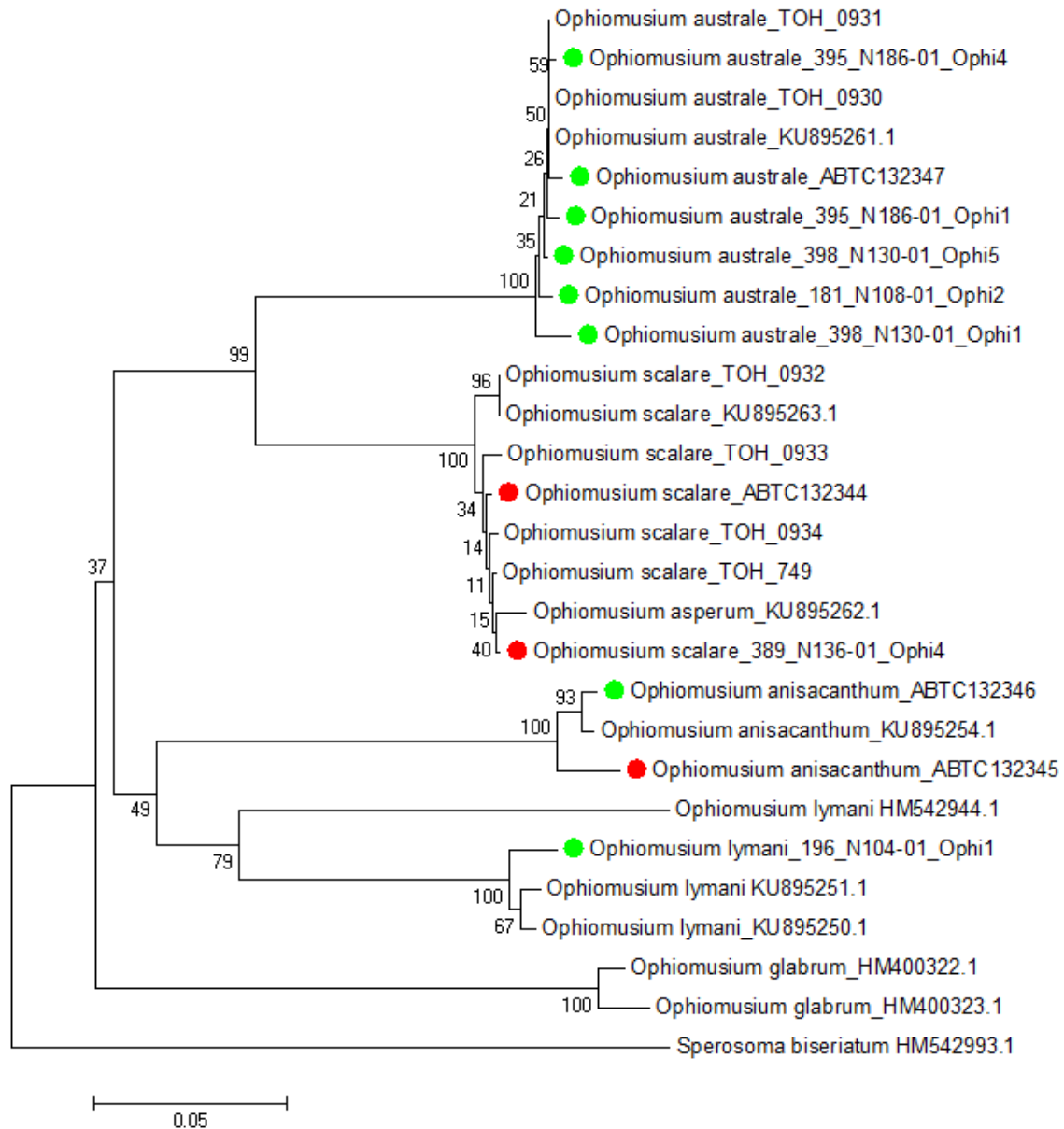


Figure 3.3. Neighbour-Joining tree for the genus *Ophiomusium* based on Kimura 2-parameter genetic distance with 1000 bootstraps. Samples with coloured circles are from the current study. Green circles indicate that the sample was identified at the species level ($p\text{-dis} < 1.65\%$) using our proposed strategy, while red circles indicate that the sample was identified at the genus level ($1.65\% < p\text{-dis} < 10.36\%$) using threshold values. Sequences from samples without coloured circles were obtained from databases (NCBI and BOLD) and their accession numbers and sample IDs are shown after the species names.

3.5 Conclusions and recommendations for future studies

Our results have shown that the approach of DNA barcoding using sequences from the mitochondrial DNA COI gene successfully identified key marine taxonomic groups in the GAB. We optimised and developed a number of protocols for the extraction of DNA and the amplification and sequencing of the barcoding COI gene across a range of marine phyla from the GAB region. These developments enabled the extraction of 925 samples, 1,987 rounds of PCRs and the generation of 531 raw and 303 high quality sequences. We also implemented analytical approaches to circumvent the lack of information about morphological identification, which enabled us to propose thresholds of COI sequence similarity to identify samples at the species, genus and/or family levels. Identifications based on the data generated and analysed as part of this report correspond to 139 Arthropoda, 106 Echinodermata, 27 Mollusca, 14 Annelida, 11 Sipuncula and 6 Cnidaria. Identifications are provided for each individual sample as a series of barcode IDs and sequence data. The sequences obtained are available via the eResearch SA repository database.

Overall, more taxonomic units (i.e. species, genus and family levels) were obtained for Echinodermata and Arthropoda due to the larger number of sequences obtained in the final data set for these phyla. We also found that species identification was substantially improved by using both the NCBI and BOLD databases. The large number of OTUs that could not be assigned to species reflects the poorly known nature of the deepwater fauna in the GAB, and hence their lack of representation in the NCBI and BOLD databases. As the final results from the morphological taxonomy become available, it will be possible to match these to the barcodes generated, substantially increasing the number of barcodes that can be matched to species level identifications. Finally, according to the species distribution records from the Atlas of Living Australia and BOLD, we found three species that are new to Australia. These are *Sperosoma biserialatum* (Echinodermata), *Granulifusus nopinicus* (Mollusca) and *Ebalia nux* (Arthropoda).

The COI region should be considered as an appropriate and cost-effective tool for biodiversity assessment in future ecological monitoring in the GAB region. That could be achieved using individual sequencing approaches (such as the one implemented here based on direct Sanger sequencing), or by using high throughput sequencing platforms for parallel acquisition of DNA barcode sequences from hundreds of specimens simultaneously (e.g. Shokralla et al. 2014). Undoubtedly, the power of DNA barcoding for species identification can be improved substantially with the availability of extensive DNA sequence data from the regional biota being studied, and with information from morphological data (including identification) and from records about the ranges of the targeted taxonomic units. The final morphological taxonomy for the samples targeted was not available in time to be used in this GAB sub-project. Finally, resolution of some technical issues identified in this project would enable more efficient application of DNA barcoding to the benthic fauna of the GAB. Improvements include development of specific primers for Echinodermata and sampling methods to target larger infauna samples during fieldwork.

3.6 References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403-410.
- Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA Barcoding of Marine Metazoa. *Annual Review of Marine Sciences*, **3**, 471-508.

- Carr CM, Hardy SM, Brown TM, Macdonald TA, Hebert PDN (2011) A Tri-Oceanic perspective: DNA barcoding reveals geographic structure and cryptic diversity in Canadian polychaetes. *Plos ONE*, **6**, e2232.
- Coppard SE, Zigler KS, Lessios HA (2013) Phylogeography of the sand dollar genus *Mellita*: Cryptic speciation along the coasts of the Americas. *Molecular Phylogenetics and Evolution*, **69**, 1033-1042.
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294-299.
- Geller J, Meyer C, Parker M, Hawk H (2013) Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*. **13**, 851-861.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *The Royal Society*, **270**, 313-321.
- Hoareau TB, Boissin E (2010) Design of phylum-specific hybrid primers for DNA barcoding: addressing the need for efficient COI amplification in the Echinodermata. *Molecular Ecology Resources*, **10**, 960-967.
- Ivanova NV, Dewaard JR, Hebert PD (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, **6**, 998-1002.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647-1649, doi: 10.1093/bioinformatics/bts199.
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, doi: 10.1093/molbev/msw054.
- Lobo J, Costa PM, Teixeira MA, Ferreira MS, Costa MH, Costa FO (2013) Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. *BMC Ecology*, **13**, doi: 10.1186/1472-6785-13-34.
- Maddison WP, Maddison DR (2016) Mesquite: a modular system for evolutionary analysis. Version 3.10 <http://mesquiteproject.org>.
- Marshall E (2005) Taxonomy. Will DNA barcodes breathe life into classification? *Science*, 307:1037.
- Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA Barcoding and Taxonomy in Diptera: A Tale of High Intraspecific Variability and Low Identification Success. *Systematic Biology*, **55**, 715-728, doi: 10.1080/10635150600969864.
- Meusnier I, Singer GA, Landry J-F, Hickey DA, Hebert PD, Hajibabaei M (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, **9**, 214.
- Shokralla S, Gibson JF, Nikbakht H, Janzen DH, Hallwachs W, Hajibabaei M (2014) Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Molecular Ecology Resources*, **14**, 892-901.
- Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular biology and evolution*. *Molecular Biology Evolution*, **13**, 510-524.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) Barcoding Australia's fish species. *Philosophical Transactions of the Royal Society*, **360**. 1847-1857.

Williams A, Tanner JE, Althaus F, Sorokin SJ, MacIntosh H, Green M, Brodie P, Loo M (2017) Great Australian Bight Benthic Biodiversity Characterisation. Final Report GABRP Project 3.1. Great Australian Bight Research Program, GABRP Research Report Series Number 19, 361pp.

Yang Z (2015) The BPP program for species tree estimation and species delimitation. *Current Zoology* **61**, 854–865.

APPENDIX 3.1: Summary of DNA extraction protocols tested for the different phyla and tissues

Protocol	DNA Extraction Procedure
DNeasy Blood & Tissue kit (Qiagen)	<ol style="list-style-type: none"> 1. Cut up to 25 mg tissue (up to 10 mg spleen) into small pieces, and place in a 1.5 mL microcentrifuge tube. For rodent tails, place one (rat) or two (mouse) 0.4–0.6 cm lengths of tail into a 1.5 mL microcentrifuge tube. Add 180 µL Buffer ATL. Earmark the animal appropriately. 2. Add 20 µL proteinase K. Mix thoroughly by vortexing, and incubate at 56°C until the tissue is completely lysed. Vortex occasionally during incubation to disperse the sample, or place in a thermomixer, shaking water bath, or on a rocking platform. 3. Vortex for 15 s. Add 200 µL Buffer AL to the sample, and mix thoroughly by vortexing. Then add 200 µL ethanol (96–100%), and mix again thoroughly by vortexing. 4. Pipet the mixture from step 3 (including any precipitate) into the DNeasy Mini spin column placed in a 2 mL collection tube (provided). Centrifuge at >6,000 x g (8,000 rpm) for 1 min. Discard flow-through and collection tube. 5. Place the DNeasy Mini spin column in a new 2 mL collection tube (provided), add 500 µL Buffer AW1, and centrifuge for 1 min at >6,000 x g (8,000 rpm). Discard flow-through and collection tube. 6. Place the DNeasy Mini spin column in a new 2 mL collection tube (provided), add 500 µL Buffer AW2, and centrifuge for 3 min at 20,000 x g (14,000 rpm) to dry the DNeasy membrane. Discard flow-through and collection tube. 7. Place the DNeasy Mini spin column in a clean 1.5 mL or 2 mL microcentrifuge tube (not provided), and pipet 30 µL ddH₂O directly onto the DNeasy membrane. Incubate at room temperature for 1 min, and then centrifuge for 1 min at >6,000x g (8,000 rpm) to elute.
CTAB	<ol style="list-style-type: none"> 1. For mollusks and echinoderms mix 5 mL of 2×CTAB and 0.5 mL of Proteinase K, 20 mg/mL in a sterile container. Add 50 µL of Lysis Mix to each Eppendorf tubes containing small pieces of tissue (1-3 mm³). Cover with fresh strip caps. Incubate at 56°C for a minimum of 6 hours or overnight to allow digestion. Centrifuge at 1,500 g for 1 min to remove any condensate from the cap strips. 2. Add 100 µL of Plant Binding Buffer (PBB) to each sample using multichannel pipette. Incubate for 5 min at RT. 3. Mix lysate 5-10 times by pipetting, transfer the lysate (about 150 µL) from the Eppendorf tube into the silica membrane spin column placed on collection tube using pipette. Close the tubes. Centrifuge at 5,000 g for 5 min to bind DNA to the silica membrane. 4. First wash step: Add 180 µL of Protein Wash Buffer (PWB) to each spin column. Close the lids and centrifuge at 5,000 for 2 min.

	<ol style="list-style-type: none"> 5. Second wash step: Add 750 µL of Wash Buffer (WB) to each spin column. Close the lids and centrifuge at 5,000 for 5 min. 6. Open the lids and place spin columns on the rack. Incubate at 56°C for 30 min to evaporate residual ethanol. 7. Position new Eppendorf tubes and place the spin columns (without collection tubes) on top to collect DNA. Dispense 20 µL of ddH₂O (prewarmed to 56°C) directly onto the membrane in each well of GF plate and incubate at room temperature for 1 min. 8. Centrifuge spin columns in the Eppendorf tubes at 5,000 g for 5 min to collect the DNA eluate. Remove the spin column and discard it. <p>(Ivanova et al. 2016)</p>
Salting out	<ol style="list-style-type: none"> 1. 1-3 mg of tissue were crushed in a 1.5mL microfuge tube and incubated at 55°C in 600 µL TNES (50 mM Tris, pH 7.5, 400 mM NaCl, 20 mM EDTA, 0.5% SDS) with 100 µg/ml Proteinase K. After overnight incubation, proteins were precipitated with 170 µL of 5M NaCl and hard shaking for 15 sec. 2. Proteins were pelleted in a microfuge at 14,000 rpm for 5 min. then DNA was precipitated from the decanted supernatant with 1 volume 100% ethanol. DNA was pelleted, washed in 70% ethanol twice, air-dried, and dissolved in 10-15 µL of sterile water. <p>(Sunnucks & Hales 1996)</p>
Chelex (Bio-rad)	<ol style="list-style-type: none"> 1. Add small sample (1-2 mm) in the Eppendorf tube with 80 µL of 10% Chelex 100 (Bio-rad) and incubate sample at 99°C for 20 min.
VLB	<ol style="list-style-type: none"> 1. Mix 45 µL of VLB (100 mM NaCl, 50 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0 and 0.5% SDS) and 5 µL of proteinase K (10 mg/ mL) in the Eppendorf tube and add 1-2 mm³ of tissue. Close the lid and incubate overnight at 56°C and then centrifuge at 1,000 g for 1 min. 2. Add 100 µL of Binding Mix (1 part of 96% ethanol and 1 part of Binding Buffer, BB, (6 M GuSCN, 20 mM EDTA pH 8.0, 10 mM Tris-HCL pH 6.4 and 4% Triton X-100)) to each tube, mix and centrifuge at 1,000 g for 20 s. 3. Open the lid and transfer 125 µL of each lysate into a silica membrane spin column with collection tube. Close the lid of spin column and centrifuge at 5,000 g for 5 min to bind DNA to the silica membrane. 4. For the first wash step, add 180 µL of Protein Wash Buffer (73% of 96% ethanol and 27% of BB) to each spin column and centrifuge at 5,000 g for 2 min. 5. For the second wash step, add 750 µL of Wash Buffer (60% ethanol, 50 mM NaCl, 10 mM Tris-HCL pH 7.4 and 0.5 mM EDTA pH 8.0) to each spin column and centrifuge at 5,000 g for 5 min. 6. Open the lid, place spin column in new Eppendorf tube and incubate at 56°C for 30 min to evaporate residual ethanol. Add 30 µL of ddH₂O to each spin column, close the lid and incubate at room temperature for 1 min. 7. Centrifuge spin column in the Eppendorf tube at 5,000 g for 5 min to collect eh DNA elute. Remove the spin column and discard it. <p>(Ivanova et al. 2006)</p>

APPENDIX 3.2: Species Identifications for each sample

ECHINODERMATA: Species (<1.65%), Genus (<10.36%), Family (<12.9%), Order (<15%), Class (<21%)				
ID	Most similar species ID	p-dis	Accession	ID level
181_N111-01_Ophi1	<i>Amphiophiura distincta</i>	0%	KU894937.1	Species
181_N111-01_Ophi5	<i>Amphiophiura distincta</i>	0%	KU894937.1	Species
389_N146-01_Ophi1	<i>Amphiophiura urbana</i>	0.54%	EU869880.1	Species
389_N146-01_Ophi2	<i>Amphiophiura urbana</i>	0.91%	EU869880.1	Species
202_N114-01_Aste	<i>Benthopecten pikei</i>	0.37%	33398	Species
202_N114-01_Aste2	<i>Benthopecten pikei</i>	0%	33398	Species
281_N106-01_Aste3	<i>Benthopecten pikei</i>	0.37%	33398	Species
ABTC132393_Holo	<i>Holothuria austrinabassa</i>	0%	EU220818.1	Species
ABTC132346	<i>Ophiomusium anisacanthum</i>	0.72%	KU895254.1	Species
181_N108-01_Ophi2	<i>Ophiomusium australe</i>	0.54%	KU895261.1	Species
395_N186-01_Ophi1	<i>Ophiomusium australe</i>	0.36%	KU895261.1	Species
395_N186-01_Ophi4	<i>Ophiomusium australe</i>	0.18%	KU895261.1	Species
398_N130-01_Ophi1	<i>Ophiomusium australe</i>	1.26%	KU895261.1	Species
398_N130-01_Ophi5	<i>Ophiomusium australe</i>	0.18%	KU895261.1	Species
ABTC132347	<i>Ophiomusium australe</i>	0.36%	KU895261.1	Species
196_N104-01_Ophi1	<i>Ophiomusium lymani</i>	1.63%	KU895250.1	Species
330_N120-01_Ophi	<i>Ophiomyxa crinita</i>	0.56%	KU895172.1	Species
389_N126-01_Ophi1	<i>Ophiomyxa crinita</i>	0.56%	KU895172.1	Species
389_N126-01_Ophi2	<i>Ophiomyxa crinita</i>	0.00%	KU895172.1	Species
389_N126-01_Ophi3	<i>Ophiomyxa crinita</i>	0.56%	KU895172.1	Species
389_N145-01_Ophi1	<i>Ophiothrix aristulata</i>	0.36%	KF663151.1	Species
389_N145-01_Ophi2	<i>Ophiothrix aristulata</i>	0.18%	KF663151.1	Species
389_N145-01_Ophi3	<i>Ophiothrix aristulata</i>	0.90%	KF663103.1	Species
389_N145-01_Ophi4	<i>Ophiothrix aristulata</i>	0.54%	KF663138.1	Species
389_N145-01_Ophi5	<i>Ophiothrix aristulata</i>	0.18%	KF663151.1	Species
159_N101-01_echi	<i>Phormosoma bursarium</i>	0%	33522	Species
382_N132-01_echi2	<i>Phormosoma bursarium</i>	0%	33643	Species
382_N132-01_echi4	<i>Phormosoma bursarium</i>	0%	33643	Species
382_N132-01_echi5	<i>Phormosoma bursarium</i>	0%	33643	Species
435_N109-01_Aste4	<i>Plutonaster complexus</i>	0.40%	33403	Species
435_N109-01_Aste1	<i>Plutonaster complexus</i>	0.40%	33403	Species
202_N106-01_echi	<i>Sperosoma biserialatum</i>	1.28%	HM542993.1	Species
202_N106-01_echi2	<i>Sperosoma biserialatum</i>	1.10%	HM542993.1	Species
202_N106-01_echi4	<i>Sperosoma biserialatum</i>	1.28%	HM542993.1	Species
202_N106-01_echi5	<i>Sperosoma biserialatum</i>	1.28%	HM542993.1	Species
281_N102-01	<i>Sperosoma biserialatum</i>	1.10%	HM542993.1	Species
395_N124-01_echi5	<i>Clypeaster japonicus</i>	10.20%	JQ341144.1	Genus
ABTC132360	<i>Ophiurnus vallincola</i>	4.15%	KU895179.1	Genus
276_N112-01_Ophi2	<i>Ophiocten hastatum</i>	3.97%	KJ620610.1	Genus
449_N112-01_Ophi2	<i>Ophiocten hastatum</i>	3.97%	KJ620610.1	Genus
ABTC132345	<i>Ophiomusium anisacanthum</i>	2.71%	KU895254.1	Genus
389_N136-01_Ophi4	<i>Ophiomusium asperum/Ophiomusium scalare</i>	0.9%/1.26%	KU895262.1/KU895263.1	Genus

ABTC132344	<i>Ophiomusium scalare/Ophiomusium asperum</i>	1.09%	KU895263.1/KU895262.1	Genus
151_N114-01_Ophi4	<i>Ophiosphalma fimbriatum</i>	3.88%	KU895360.1	Genus
151_N114-01_Ophi5	<i>Ophiosphalma fimbriatum</i>	3.88%	KU895360.1	Genus
216_N115-01_Ophi1	<i>Ophiosphalma fimbriatum</i>	4.43%	KU895360.1	Genus
202_N103-01_Aste	<i>Plutonaster complexus</i>	5.79%	33403	Genus
435_N109-01_Aste3	<i>Plutonaster complexus</i>	5.40%	33403	Genus
ABTC132366_echi	<i>Plutonaster complexus</i>	5.79%	33403	Genus
435_N109-01_Aste5	<i>Plutonaster knoxi/ Plutonaster complexus</i>	0.40%	15388/33403	Genus
207_N122-01_echi	<i>Salenocidaris profundus</i>	3.99%	KF642994.1	Genus
207_N122-01_echi3	<i>Salenocidaris profundus</i>	3.26%	KF642994.1	Genus
207_N122-01_echi5	<i>Salenocidaris profundus</i>	3.62%	KF642994.1	Genus
395_N124-01_echi	<i>Clypeaster japonicus</i> (Family: Clypeasteridae)	10.70%	JQ341144.1	Family
395_N124-01_echi2	<i>Clypeaster japonicus</i> (Family: Clypeasteridae)	12.30%	JQ341144.1	Family
395_N124-01_echi3	<i>Clypeaster japonicus</i> (Family: Clypeasteridae)	11.80%	JQ341144.1	Family
398_N133-01_echi	<i>Clypeaster japonicus</i> (Family: Clypeasteridae)	10.60%	JQ341144.1	Family
141_N117-01_Aste2	<i>Hymenaster pellucidus</i> (Family: Pterasteridae)	11.50%	KU495861.1	Family
207_N124-01_Aste1	<i>Hymenaster pellucidus</i> (Family: Pterasteridae)	11.50%	KU495861.1	Family
207_N124-01_Aste2	<i>Hymenaster pellucidus</i> (Family: Pterasteridae)	11.70%	KU495861.1	Family
207_N124-01_Aste4	<i>Hymenaster pellucidus</i> (Family: Pterasteridae)	11.50%	KU495861.1	Family
207_N124-01_Aste5	<i>Hymenaster pellucidus</i> (Family: Pterasteridae)	11.50%	KU495861.1	Family
216_N106-01_Aste2	<i>Hymenaster pellucidus</i> (Family: Pterasteridae)	11.50%	KU495861.1	Family
216_N106-01_Aste4	<i>Hymenaster pellucidus</i> (Family: Pterasteridae)	11.50%	KU495861.1	Family
216_N106-01_Aste5	<i>Hymenaster pellucidus</i> (Family: Pterasteridae)	11.50%	KU495861.1	Family
141_N131-01_Holo2	<i>Peniagone vignoni</i> (Family: Elpididae)	11.78%	HM196385.1	Family
ABTC132350_echi	<i>Phormosoma bursarium</i> (Family: Echinothuriidae)	11.78%	46086	Family
207_N123-01_Aste4	<i>Psilaster charcoti</i> (Family: Astropectinidae)	12.48%	38450.1	Family
216_N128.01_Aste	<i>Psilaster charcoti</i> (Family: Astropectinidae)	12.85%	38450.1	Family
276_N108-01_Aste2	<i>Psilaster charcoti</i> (Family: Astropectinidae)	12.85%	38450.1	Family
276_N108-01_Aste3	<i>Psilaster charcoti</i> (Family: Astropectinidae)	12.85%	38450.1	Family
276_N108-01_Aste4	<i>Psilaster charcoti</i> (Family: Astropectinidae)	12.85%	38450.1	Family
281_N101-01_Aste2	<i>Psilaster charcoti</i> (Family: Astropectinidae)	12.48%	38450.1	Family
281_N101-01_Aste3	<i>Psilaster charcoti</i> (Family: Astropectinidae)	12.48%	38450.1	Family
281_N101-01_Aste5	<i>Psilaster charcoti</i> (Family: Astropectinidae)	12.66%	38450.1	Family
281_N108-01_Aste2	<i>Psilaster charcoti</i> (Family: Astropectinidae)	12.48%	38450.1	Family
435_N109-01_Aste2	<i>Psilaster charcoti</i> (Family: Astropectinidae)	12.66%	38450.1	Family
ABTC132353_echi	<i>Ctenodiscus crispatus</i> (Order: Paxillosida)	14.21%	HLC-23907	Order
207_N123-01_Aste1	<i>Psilaster charcoti</i> (Order: Paxillosida)	13.22%	38450.1	Order
207_N123-01_Aste5	<i>Psilaster charcoti</i> (Order: Paxillosida)	13.22%	38450.1	Order

276_N108-01_Aste5	<i>Psilaster charcoti</i> (Order: Paxillosida)	13.22%	38450.1	Order
281_N101-01_Aste4	<i>Psilaster charcoti</i> (Order: Paxillosida)	13.41%	38450.1	Order
330_N132-01_Aste	<i>Thrissacanthias penicillatus</i> (Order: Paxillosida)	13.65%	RBCM EC00045	Order
389_N120-01_Aste	<i>Thrissacanthias penicillatus</i> (Order: Paxillosida)	13.90%	RBCM EC00045	Order
181_N105-01_Aste2	<i>Uniophora granifera</i> (Order: Forcipulatida)	13.54%	Ech031	Order
395_N154-01_Aste2	<i>Uniophora granifera</i> (Order: Forcipulatida)	13.54%	Ech031	Order
395_N154-01_Aste3	<i>Uniophora granifera</i> (Order: Forcipulatida)	13.72%	Ech031	Order
398_N138-01_Aste2	<i>Uniophora granifera</i> (Order: Forcipulatida)	13.72%	Ech031	Order
398_N138-01_Aste3	<i>Uniophora granifera</i> (Order: Forcipulatida)	13.54%	Ech031	Order
ABTC132342	<i>Uniophora granifera</i> (Order: Forcipulatida)	13.36%	Ech031	Order
202_N105-01_Aste1	<i>Hymenaster pellucidus</i> (Class: Asteroidea)	15.21%	KU495861.1	Class
202_N105-01_Aste2	<i>Hymenaster pellucidus</i> (Class: Asteroidea)	15.21%	KU495861.1	Class
202_N105-01_Aste3	<i>Hymenaster pellucidus</i> (Class: Asteroidea)	15.21%	KU495861.1	Class
202_N105-01_Aste5	<i>Hymenaster pellucidus</i> (Class: Asteroidea)	15.01%	KU495861.1	Class
276_N104-01_Aste	<i>Hymenaster pellucidus</i> (Class: Asteroidea)	19.49%	KU495861.1	Class
281_N104-01_Aste1	<i>Hymenaster pellucidus</i> (Class: Asteroidea)	15.21%	KU495861.1	Class
281_N104-01_Aste2	<i>Hymenaster pellucidus</i> (Class: Asteroidea)	15.21%	KU495861.1	Class
281_N104-01_Aste3	<i>Hymenaster pellucidus</i> (Class: Asteroidea)	15.01%	KU495861.1	Class
435_N111-01_Aste1	<i>Hymenaster pellucidus</i> (Class: Asteroidea)	15.21%	KU495861.1	Class
ABTC132355_ophi_aste	<i>Marthasterias glacialis</i> (Class: Asteroidea)	16.79%	HM107750.1	Class
155_N110-01_Holo	<i>Protelpidia murrayi</i> (Class: Holothuroidea)	17.24%	KF713386.1	Class
207_N127-01_Holo	<i>Protelpidia murrayi</i> (Class: Holothuroidea)	16.88%	KF713386.1	Class
276_N106-01_Holo	<i>Protelpidia murrayi</i> (Class: Holothuroidea)	16.88%	KF713386.1	Class
281_N103-01_Holo	<i>Protelpidia murrayi</i> (Class: Holothuroidea)	17.24%	KF713386.1	Class
ABCT132332_Holo	<i>Protelpidia murrayi</i> (Class: Holothuroidea)	17.06%	KF713386.1	Class
ABTC132365	<i>Psilaster charcoti</i> (Class: Asteroidea)	19.31%	GU227097.1	Class

MOLLUSCA: Species (<1.65%), Genus (<10.36%), Family (<12.9%), Order (<15%), Class (<21%)

ID	Closest Species ID	p-dis	Accession	ID level
186_N106-01_mol	<i>Granulifusus niponicus</i>	0.54%	KT753935.1	Species
186_N106-01_Mol3	<i>Granulifusus niponicus</i>	0.54%	KT753935.1	Species
186_N106-01_Mol4	<i>Granulifusus niponicus</i>	0.18%	KT753935.1	Species
186_N106-01_Mol5	<i>Granulifusus niponicus</i>	0.54%	KT753935.1	Species
389_N108-01_mol	<i>Granulifusus niponicus</i>	0.36%	KT753935.1	Species
389_N108-01_Mol2	<i>Granulifusus niponicus</i>	0.54%	KT753935.1	Species
389_N108-01_Mol3	<i>Granulifusus niponicus</i>	0.54%	KT753935.1	Species
141_N124-01_mol	<i>Aforia serranoi</i>	3.96%	KT448836.1	Genus
167_N102-01_mol	<i>Muusoctopus oregonensis</i>	3.98%	HM572180.1	Genus
196_N101-01_mol	<i>Muusoctopus oregonensis</i>	3.98%	HM572180.1	Genus
ABTC132319_mol	<i>Stoloteuthis japonica</i>	9.17%	AB591072.1	Genus
ABTC132348_mol	<i>Amoria hunteri</i> (Family: Volutidae)	12.48%	JN182226.1	Family
389_N107-01_mol	<i>Cranopsis cucullata</i> (Family: Fissurellidae)	10.67%	GQ160755.1	Family
ABTC132340	<i>Dentalium majorinum</i> (Order: Dentaliida)	14.29%	AY260823.1	Order
281_N125-01_Mol2	<i>Echinolittorina paytensis</i> (Class: Gastropoda)	18.10%	AJ623027.1	Class
155_N105-01_mol	<i>Lorabela pelseneeri</i> (Class: Gastropoda)	17.88%	KT448824.1	Class
276_N122-01_Mol3	<i>Lorabela pelseneeri</i> (Class: Gastropoda)	17.88%	KT448824.1	Class

276_N122-01_Mol4	<i>Lorabela pelseneeri</i> (Class: Gastropoda)	17.88%	KT448824.1	Class
281_N125-01_Mol3	<i>Lorabela pelseneeri</i> (Class: Gastropoda)	18.07%	KT448824.1	Class
281_N125-01_Mol4	<i>Lorabela pelseneeri</i> (Class: Gastropoda)	17.88%	KT448824.1	Class
281_N125-01_Mol5	<i>Lorabela pelseneeri</i> (Class: Gastropoda)	18.07%	KT448824.1	Class
449_N118-01_mol	<i>Lorabela pelseneeri</i> (Class: Gastropoda)	18.07%	KT448824.1	Class
449_N118-01_Mol2	<i>Lorabela pelseneeri</i> (Class: Gastropoda)	17.70%	KT448824.1	Class
449_N118-01_Mol3	<i>Lorabela pelseneeri</i> (Class: Gastropoda)	17.88%	KT448824.1	Class
449_N118-01_Mol4	<i>Lorabela pelseneeri</i> (Class: Gastropoda)	17.70%	KT448824.1	Class
449_N118-01_Mol5	<i>Lorabela pelseneeri</i> (Class: Gastropoda)	17.70%	KT448824.1	Class
ABTC132318_mol	<i>Sepia hirunda</i> (Class: Cephalopoda)	15.29%	AY530213.1	Class

ANNELIDA: Species (<1.65%), Genus (<10.36%), Family (<12.9%), Order (<15%), Class (<21%)

ID	Closest Species ID	p-dis	Accession	ID level
186_N108-01_poly	<i>Laetmonice filicornis</i> (Class: Polychaeta)	17.72%	JN852919	Class
186_N108-01_poly2	<i>Laetmonice filicornis</i> (Class: Polychaeta)	16.87%	JN852919	Class
186_N108-01_poly3	<i>Laetmonice filicornis</i> (Class: Polychaeta)	17.21%	JN852919	Class
186_N108-01_poly5	<i>Laetmonice filicornis</i> (Class: Polychaeta)	17.21%	JN852919	Class
191_N104-01_poly	<i>Laetmonice filicornis</i> (Class: Polychaeta)	17.55%	JN852919	Class
395_N146-01_poly	<i>Laetmonice filicornis</i> (Class: Polychaeta)	17.72%	JN852919	Class
395_N146-01_poly2	<i>Laetmonice filicornis</i> (Class: Polychaeta)	17.21%	JN852919	Class
395_N146-01_poly3	<i>Laetmonice filicornis</i> (Class: Polychaeta)	17.21%	JN852919	Class
395_N146-01_poly4	<i>Laetmonice filicornis</i> (Class: Polychaeta)	17.21%	JN852919	Class
395_N148-01_poly1	<i>Laetmonice filicornis</i> (Class: Polychaeta)	17.89%	JN852919	Class
395_N148-01_poly2	<i>Laetmonice filicornis</i> (Class: Polychaeta)	17.89%	JN852919	Class
395_N148-01_poly3	<i>Laetmonice filicornis</i> (Class: Polychaeta)	17.89%	JN852919	Class
395_N148-01_poly4	<i>Laetmonice filicornis</i> (Class: Polychaeta)	17.89%	JN852919	Class
ABTC132333_poly	<i>Laetmonice filicornis</i> (Class: Polychaeta)	17.21%	JN852919	Class

SIPUNCULA: Species (<1.65%), Genus (<10.36%), Family (<12.9%), Order (<15%), Class (<21%)

ID	Closest Species ID	p-dis	Accession	ID level
382_N150-01_sip	<i>Nephasoma diaphanes</i> (Class: Sipunculida)	18.64%	JN182658.1	Class
382_N150-01_Sip3	<i>Nephasoma diaphanes</i> (Class: Sipunculida)	18.64%	JN182658.1	Class
382_N150-01_Sip4	<i>Nephasoma diaphanes</i> (Class: Sipunculida)	18.64%	AY533769.1	Class
382_N150-01_Sip5	<i>Nephasoma diaphanes</i> (Class: Sipunculida)	18.64%	JN182658.1	Class
ABTC132337_sip	<i>Nephasoma diaphanes</i> (Class: Sipunculida)	19.02%	DQ300128.1	Class
151_N111-01_sip	<i>Sipunculus nudus</i> (Class: Sipunculida)	20.35%	GU014047.1	Class
151_N111-01_Sip4	<i>Sipunculus nudus</i> (Class: Sipunculida)	20.35%	GU014047.1	Class
435_N114-01_sip	<i>Sipunculus nudus</i> (Class: Sipunculida)	20.35%	AB542575.1	Class
435_N114-01_Sip2	<i>Sipunculus nudus</i> (Class: Sipunculida)	20.54%	AB542575.1	Class
435_N114-01_Sip3	<i>Sipunculus nudus</i> (Class: Sipunculida)	20.15%	AB542575.1	Class
435_N114-01_Sip4	<i>Sipunculus nudus</i> (Class: Sipunculida)	20.15%	KC404840.1	Class

CNIDARIA: Species (<1.65%), Genus (<10.36%), Family (<12.9%), Order (<15%), Class (<21%)

ID	Closest Species ID	p-dis	Accession	ID level
330_N133.01	<i>Corallimorphus profundus</i>	0.18%	KP938440.1	Species
151_N147.01_1	<i>Halopteris californica/Halopteris willemoesi</i>	0.89%	KF874202.1/KF874192.1	Genus
151_N147.01_3	<i>Halopteris californica/Halopteris willemoesi</i>	0.89%	KF874202.1/KF874192.1	Genus
151_N147.01_4	<i>Halopteris californica/Halopteris willemoesi</i>	0.89%	KF874202.1/KF874192.1	Genus
151_N123.01	<i>Stomphia didemon</i>	4.09%	KHBC-S13-0033-02	Genus
155_N123.01	<i>Stomphia didemon</i>	4.09%	KHBC-S13-0033-02	Genus

ARTHROPODA: Species (<1.65%), Genus (<10.36%), Family (<12.9%), Order (<15%), Class (<21%)

ID	Closest Species ID	p-dis	Accession	ID level
ABTC132323_arth	<i>Acantheephyra pelagica</i>	0%	KP076179.1	Species
435_N110-01_arth	<i>Acantheephyra quadrispinosa</i>	0.18%	KP759362.1	Species
207_N128-01_Arth2	<i>Colossendeis gigas</i>	1.06%	FJ716624.1	Species
181_N138-01_arth	<i>Ebalia nux</i>	1.24%	JQ348859.1	Species
181_N138-01_arth2	<i>Ebalia nux</i>	1.42%	JQ348859.1	Species
181_N138-01_arth3	<i>Ebalia nux</i>	1.42%	JQ348859.1	Species
181_N138-01_arth4	<i>Ebalia nux</i>	1.42%	JQ348859.1	Species
181_N138-01_arth5	<i>Ebalia nux</i>	1.42%	JQ348859.1	Species
398_N157-01_arth	<i>Ebalia nux</i>	1.42%	JQ348859.1	Species
398_N157-01_Arth2	<i>Ebalia nux</i>	1.42%	JQ348859.1	Species
398_N157-01_Arth3	<i>Ebalia nux</i>	1.59%	JQ348859.1	Species
398_N157-01_Arth4	<i>Ebalia nux</i>	1.24%	JQ348859.1	Species
398_N157-01_Arth5	<i>Ebalia nux</i>	1.24%	JQ348859.1	Species
395_N125-01_arth	<i>Ibacus alticrenatus</i>	0.18%	JN701659.1	Species
174_N147-01_arth	<i>Ibacus alticrenatus</i>	0%	JN701659.1	Species
398_N117-01_arth	<i>Ibacus alticrenatus</i>	0%	JN701659.1	Species
398_N117-01_Arth2	<i>Ibacus alticrenatus</i>	0%	JN701659.1	Species
398_N117-01_Arth3	<i>Ibacus alticrenatus</i>	0%	JN701659.1	Species
398_N117-01_Arth4	<i>Ibacus alticrenatus</i>	0%	JN701659.1	Species
398_N117-01_Arth5	<i>Ibacus alticrenatus</i>	0.18%	JN701659.1	Species
ABTC132330_arth	<i>Ibacus alticrenatus</i>	0%	JN701659.1	Species
196_N102-01_arth	<i>Neolithodes brodiei</i>	0.89%	EU493263.1	Species
ABTC132325_arth	<i>Pentacheles laevis</i>	0.71%	LC022127.1	Species
207_N128-01_Arth5	<i>Colossendeis gigas</i>	7.61%	FJ716626.1	Genus
207_N129-01_arth	<i>Colossendeis japonica</i>	3.09%	FJ716625.1	Genus
207_N129-01_Arth2	<i>Colossendeis japonica</i>	2.91%	FJ716625.1	Genus
207_N129-01_Arth3	<i>Colossendeis japonica</i>	2.91%	FJ716625.1	Genus
276_N118-01_Arth2	<i>Colossendeis japonica</i>	3.09%	FJ716625.1	Genus
155_N102-01_arth	<i>Colossendeis macerrima</i>	7.12%	JN018213.1	Genus
207_N128-01_arth	<i>Colossendeis macerrima</i>	7.30%	JN018213.1	Genus
207_N128-01_Arth3	<i>Colossendeis macerrima</i>	7.30%	JN018213.1	Genus
207_N129-01_Arth4	<i>Colossendeis macerrima</i>	3.56%	JN018213.1	Genus
276_N118-01_arth	<i>Colossendeis macerrima</i>	3.56%	JN018213.1	Genus
276_N120-01_arth	<i>Colossendeis macerrima</i>	7.30%	JN018213.1	Genus
276_N120-01_Arth2	<i>Colossendeis macerrima</i>	7.30%	JN018213.1	Genus
281_N122-01_arth	<i>Colossendeis macerrima</i>	7.12%	JN018213.1	Genus
207_N114-01_arth	<i>Glyphocrangon indonesiensis</i>	9.91%	JX486076.1	Genus
207_N114-01_Arth2	<i>Glyphocrangon indonesiensis</i>	9.91%	JX486076.1	Genus
207_N114-01_Arth3	<i>Glyphocrangon indonesiensis</i>	9.91%	JX486076.1	Genus
276_N121-01_arth	<i>Glyphocrangon indonesiensis</i>	9.91%	JX486076.1	Genus
330_N101-01_arth	<i>Metanephrops velutinus/Metanephrops andamanicus</i>	0.534% /1.25%	EU186132.1/EU186133.1	Genus
252_N116-01_arth_1	<i>Phronima sedentaria</i>	7.91%	GU145043.1	Genus
252_N116-01_Arth2_1	<i>Phronima sedentaria</i>	7.91%	GU145043.1	Genus
252_N116-01_Arth3_1	<i>Phronima sedentaria</i>	7.91%	GU145043.1	Genus
252_N116-01_Arth4_1	<i>Phronima sedentaria</i>	8.45%	GU145043.1	Genus
252_N116-01_Arth5_1	<i>Phronima sedentaria</i>	7.91%	GU145043.1	Genus

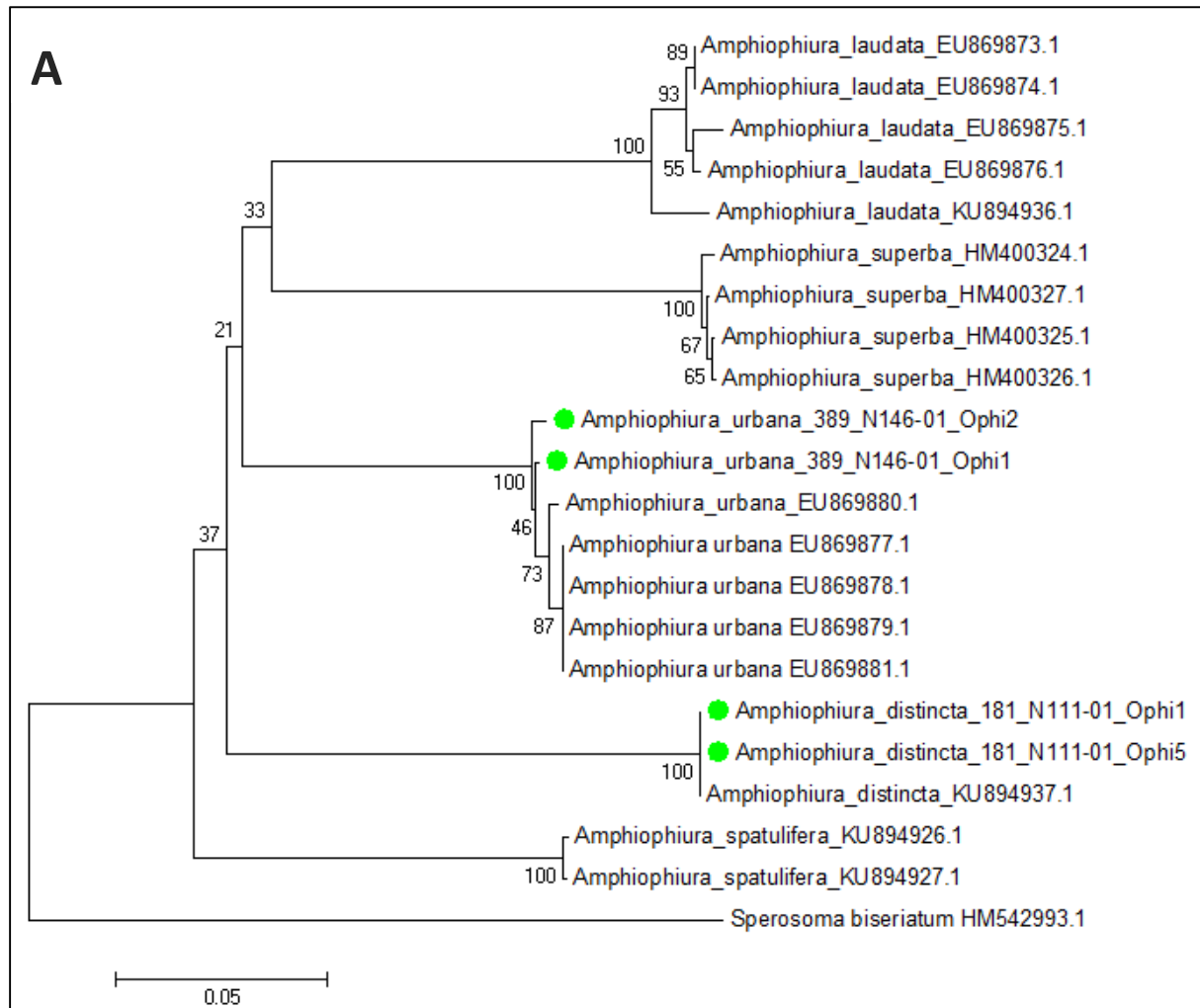
395_N107-01_Arth3	<i>Chlorotocus crassicornis</i> (Family: Pandalidae)	11.86%	JQ305891.1	Family
395_N107-01_Arth4	<i>Chlorotocus crassicornis</i> (Family: Pandalidae)	11.68%	JQ305891.1	Family
395_N107-01_Arth5	<i>Chlorotocus crassicornis</i> (Family: Pandalidae)	12.21%	JQ305891.1	Family
398_N116-01_Arth5	<i>Chlorotocus crassicornis</i> (Family: Pandalidae)	11.86%	JQ305891.1	Family
395_N107-01_Arth2	<i>Chlorotocus crassicornis</i> (Family: Pandalidae)	11.86%	JQ305891.1	Family
398_N116-01_Arth2	<i>Chlorotocus crassicornis</i> (Family: Pandalidae)	11.50%	JQ305891.1	Family
398_N126-01_Arth2	<i>Liocarcinus maculatus</i> (Family: Polybiidae)	11.92%	FJ174949.1	Family
398_N126-01_Arth3	<i>Lybia tessellata</i> (Family: Xanthidae)	12.28%	HM751017.1	Family
398_N126-01_Arth	<i>Perbrinckia punctata</i> (Family: Gecarcinucidae)	12.57%	GQ289662.1	Family
ABTC132326_arth	<i>Plesionika martia</i> (Family: Pandalidae)	10.80%	JQ306280.1	Family
330_N181-01_arth	<i>Propagurus gaudichaudii</i> (Family: Paguridae)	10.97%	ZSMA20111490	Family
398_N155-01_Arth5	<i>Dromia personata</i> (Order: Decapoda)	13.81%	JQ306068.1	Order
141_N134-01_arth	<i>Eumunida annulosa</i> (Order: Decapoda)	14.65%	EU243493.1	Order
216_N113-01_arth	<i>Eumunida annulosa</i> (Order: Decapoda)	14.83%	EU243493.1	Order
382_N138-01_arth4	<i>Eumunida capillata</i> (Order: Decapoda)	13.10%	EU243343.1	Order
382_N138-01_arth5	<i>Eumunida capillata</i> (Order: Decapoda)	13.10%	EU243343.1	Order
382_N138-01_arth3	<i>Eumunida capillata</i> (Order: Decapoda)	13.45%	EU243343.1	Order
449_N146-01_arth	<i>Eumunida multilineata</i> (Order: Decapoda)	14.51%	EU243546.1	Order
155_N134-01_arth	<i>Eumunida multilineata</i> (Order: Decapoda)	14.51%	EU243546.1	Order
382_N138-01_arth2	<i>Eumunida multilineata</i> (Order: Decapoda)	14.69%	EU243546.1	Order
435_N113-01_arth	<i>Eumunida multilineata</i> (Order: Decapoda)	14.69%	EU243546.1	Order
435_N113-01_Arth2	<i>Eumunida multilineata</i> (Order: Decapoda)	14.87%	EU243546.1	Order
435_N113-01_Arth3	<i>Eumunida multilineata</i> (Order: Decapoda)	14.87%	EU243546.1	Order
449_N145-01_arth	<i>Eumunida multilineata</i> (Order: Decapoda)	14.87%	EU243546.1	Order
398_N154-01_Arth2	<i>Galearctus timidus</i> (Order: Decapoda)	14.85%	JN701673.1	Order
398_N154-01_Arth3	<i>Galearctus timidus</i> (Order: Decapoda)	14.85%	JN701673.1	Order
389_N114-01_Arth3	<i>Glyphocrangon armata</i> (Order: Decapoda)	13.10%	HQ241546.1	Order
389_N114-01_Arth4	<i>Glyphocrangon armata</i> (Order: Decapoda)	13.10%	HQ241546.1	Order
389_N114-01_Arth5	<i>Glyphocrangon armata</i> (Order: Decapoda)	12.92%	HQ241546.1	Order
389_N114-01_arth	<i>Glyphocrangon armata</i> (Order: Decapoda)	12.92%	HQ241546.1	Order
330_N118-01_arth	<i>Glyphocrangon armata</i> (Order: Decapoda)	12.92%	HQ241546.1	Order
395_N104-01_arth1	<i>Haptosquilla hamifera</i> (Order: Stomatopoda)	14.87%	KM074037.1	Order
398_N114-01_arth1	<i>Haptosquilla hamifera</i> (Order: Stomatopoda)	14.87%	KM074037.1	Order
398_N114-01_Arth3	<i>Haptosquilla hamifera</i> (Order: Stomatopoda)	14.69%	KM074037.1	Order
398_N114-01_Arth4	<i>Haptosquilla hamifera</i> (Order: Stomatopoda)	14.87%	KM074037.1	Order
398_N156-01_arth	<i>Inachus dorsettensis</i> (Order: Decapoda)	13.88%	JQ306209.1	Order
398_N156-01_Arth4	<i>Inachus dorsettensis</i> (Order: Decapoda)	13.88%	JQ306209.1	Order
398_N156-01_Arth2	<i>Inachus dorsettensis</i> (Order: Decapoda)	14.06%	JQ306209.1	Order
398_N156-01_Arth3	<i>Inachus dorsettensis</i> (Order: Decapoda)	14.06%	JQ306209.1	Order
398_N156-01_Arth5	<i>Inachus dorsettensis</i> (Order: Decapoda)	13.70%	JQ306209.1	Order

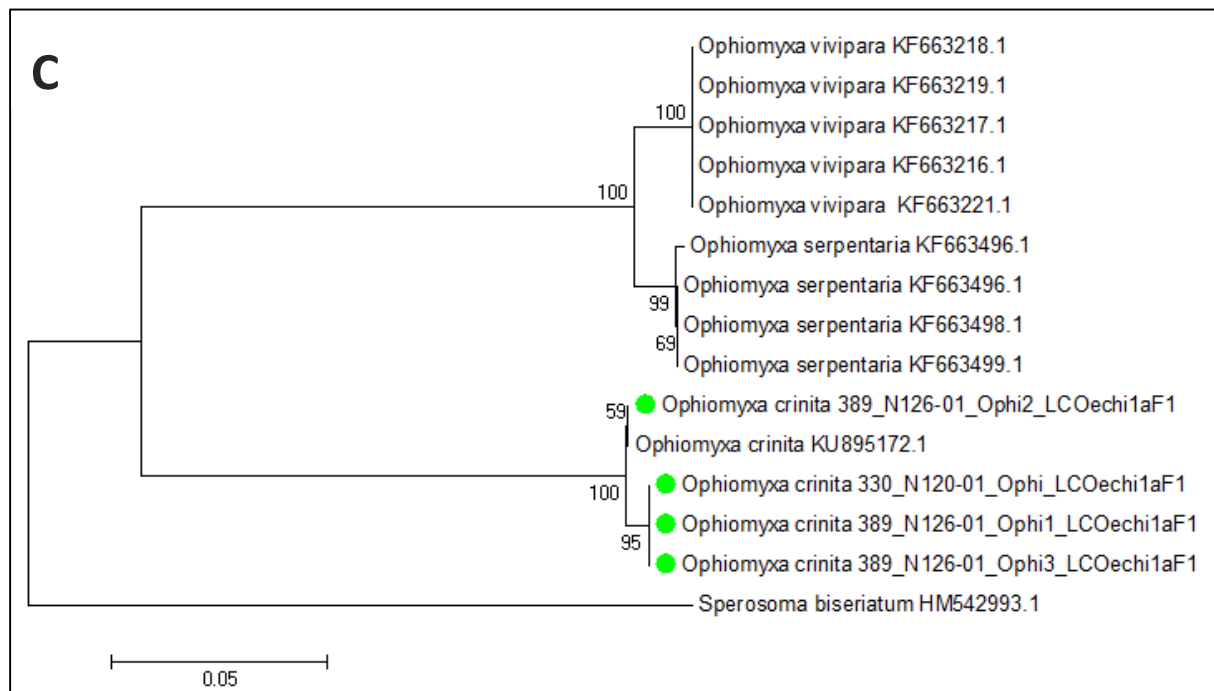
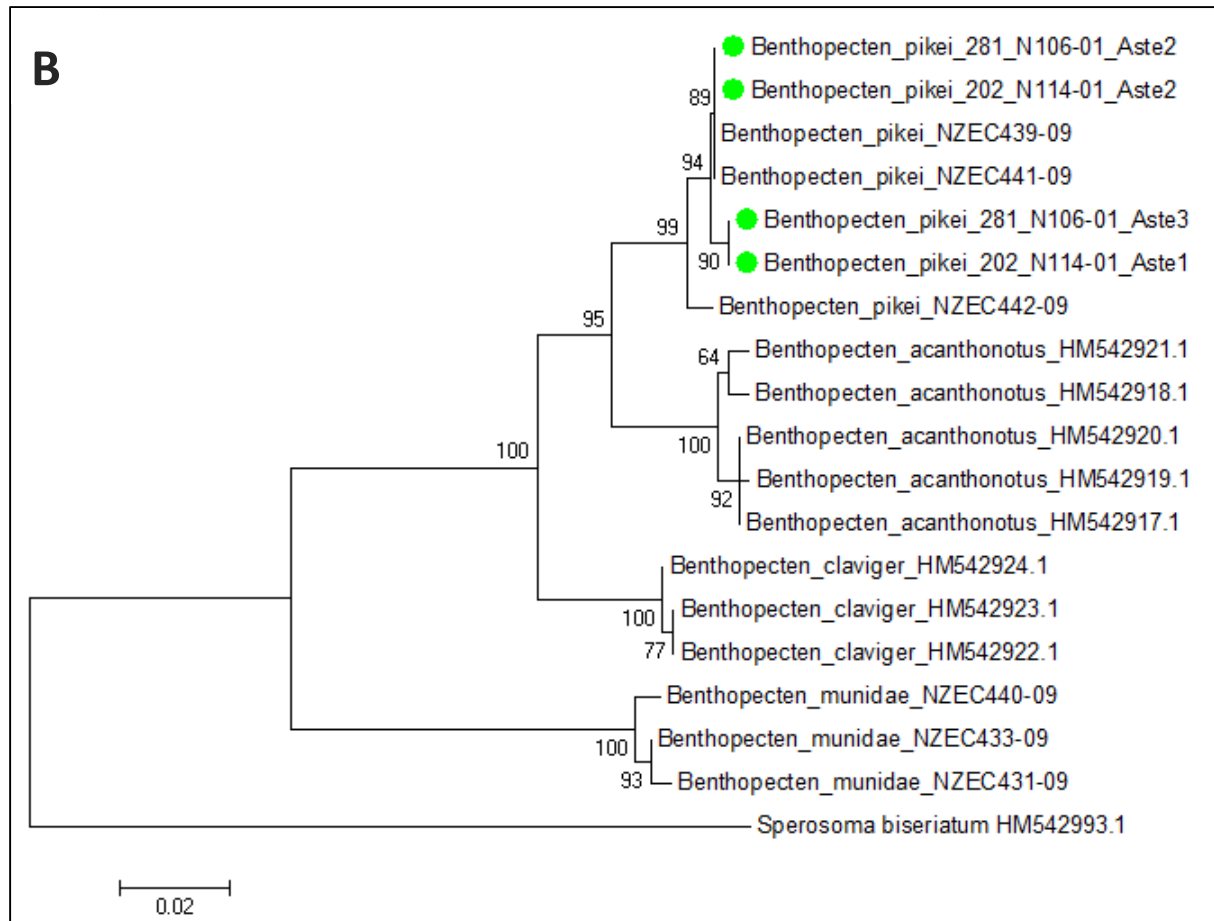
398_N155-01_Arth3	<i>Lauridromia dehaani</i> (Order: Decapoda)	13.98%	EU636986.1	Order
398_N155-01_arth	<i>Lauridromia dehaani</i> (Order: Decapoda)	13.63%	EU636986.1	Order
151_N121-01_Arth1	<i>Lissosabinea unispinosa</i> (Order: Decapoda)	14.16%	KP759421.1	Order
435_N105-01_arth	<i>Nematocarcinus lanceopes</i> (Order: Decapoda)	12.92%	EF407640.2	Order
435_N105-01_Arth3	<i>Nematocarcinus lanceopes</i> (Order: Decapoda)	12.92%	EF407640.2	Order
186_N103-01_arth	<i>Neonrosella vitiata</i> (Order: Sessilia)	14.54%	KM974424.1	Order
186_N103-01_Arth3	<i>Neonrosella vitiata</i> (Order: Sessilia)	14.36%	KM974424.1	Order
186_N103-01_Arth4	<i>Neonrosella vitiata</i> (Order: Sessilia)	14.54%	KM974424.1	Order
186_N103-01_Arth5	<i>Neonrosella vitiata</i> (Order: Sessilia)	14.01%	KM974424.1	Order
330_N123-01_arth	<i>Upogebia africana</i> (Order: Decapoda)	14.03%	DQ351379.1	Order
ABTC132373_arth	<i>Aegla alacalufi</i> (Class: Malacostraca)	15.75%	FJ471763.1	Class
292_N114-01_arth	<i>Aristeus antennatus</i> (Class: Malacostraca)	15.40%	JQ305889.1	Class
292_N115-01_arth	<i>Aristeus antennatus</i> (Class: Malacostraca)	15.40%	JQ305889.1	Class
292_N116-01_Arth	<i>Aristeus antennatus</i> (Class: Malacostraca)	15.40%	JQ305889.1	Class
292_N117-01_Arth	<i>Aristeus antennatus</i> (Class: Malacostraca)	15.75%	JQ305889.1	Class
292_N118-01_arth	<i>Aristeus antennatus</i> (Class: Malacostraca)	15.40%	JQ305889.1	Class
435_N105-01_Arth4	<i>Aristeus antennatus</i> (Class: Malacostraca)	15.40%	JQ305889.1	Class
ABTC132322_arth	<i>Aristeus antennatus</i> (Class: Malacostraca)	15.58%	JQ305889.1	Class
207_N106-01_arth	<i>Austromegabalanus psittacus</i> (Class: Maxillopoda)	18.58%	KJ756056.1	Class
207_N106-01_Arth2	<i>Austromegabalanus psittacus</i> (Class: Maxillopoda)	18.23%	KJ756056.1	Class
395_N140-01_arth	<i>Ceylonthelphusa cavatrix</i> (Class: Malacostraca)	16.64%	GQ982586.1	Class
395_N140-01_Arth2	<i>Ceylonthelphusa cavatrix</i> (Class: Malacostraca)	16.46%	GQ982586.1	Class
395_N140-01_Arth3	<i>Ceylonthelphusa cavatrix</i> (Class: Malacostraca)	16.46%	GQ982586.1	Class
398_N115-01_arth	<i>Chionoecetes tanneri</i> (Class: Malacostraca)	19.57%	JQ407478.1	Class
398_N115-01_Arth2	<i>Chionoecetes tanneri</i> (Class: Malacostraca)	19.57%	JQ407478.1	Class
398_N115-01_Arth4	<i>Chionoecetes tanneri</i> (Class: Malacostraca)	19.57%	JQ407478.1	Class
174_N142-01_arth	<i>Chionoecetes tanneri</i> (Class: Malacostraca)	19.57%	JQ407478.1	Class
330_N121-01_arth	<i>Chionoecetes tanneri</i> (Class: Malacostraca)	19.57%	JQ407478.1	Class
398_N155-01_Arth2	<i>Clinothelphusa kakoota</i> (Class: Malacostraca)	15.75%	GQ289664.1	Class
398_N155-01_Arth4	<i>Dromia personata</i> (Class: Malacostraca)	15.22%	JQ306068.1	Class
151_N121-01_Arth3	<i>Eumunida annulosa</i> (Class: Malacostraca)	15.01%	EU243493.1	Class
141_N134-01_arth3	<i>Eumunida annulosa</i> (Class: Malacostraca)	15.19%	EU243493.1	Class
216_N110-01_arth	<i>Eumunida annulosa</i> (Class: Malacostraca)	15.19%	EU243493.1	Class
216_N112-01_arth	<i>Eumunida annulosa</i> (Class: Malacostraca)	15.19%	EU243493.1	Class
274_N117-01_arth	<i>Eumunida annulosa</i> (Class: Malacostraca)	15.19%	EU243493.1	Class
274_N118-01_arth	<i>Eumunida annulosa</i> (Class: Malacostraca)	15.01%	EU243493.1	Class
274_N120-01_arth	<i>Eumunida annulosa</i> (Class: Malacostraca)	15.01%	EU243493.1	Class
382_N138-01_arth	<i>Eumunida minor</i> (Class: Malacostraca)	15.40%	EU243550.1	Class
ABTC132374_arth	<i>Eumunida sternomaculata</i> (Class: Malacostraca)	15.22%	EU243484.1	Class
395_N104-01_Arth3	<i>Haptosquilla hamifera</i> (Class: Malacostraca)	15.04%	KM074037.1	Class
395_N104-01_Arth4	<i>Haptosquilla hamifera</i> (Class: Malacostraca)	15.22%	KM074037.1	Class

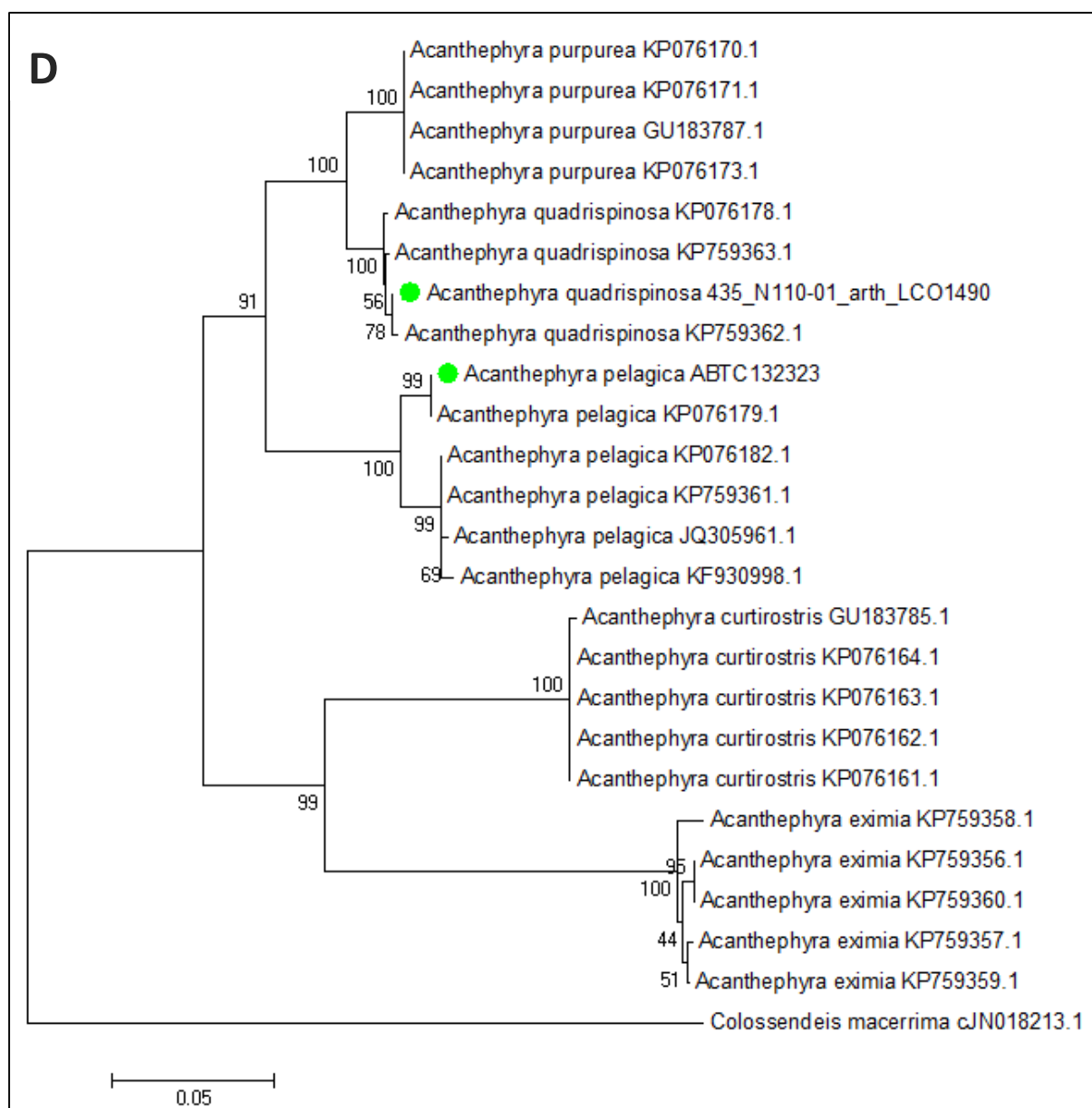
398_N114-01_Arth2	<i>Haptosquilla hamifera</i> (Class: Malacostraca)	15.04%	KM074037.1	Class
398_N114-01_Arth5	<i>Haptosquilla hamifera</i> (Class: Malacostraca)	15.04%	KM074037.1	Class
186_N103-01_Arth2	<i>Neonrosella vitiata</i> (Class: Maxillopoda)	15.07%	KM974424.1	Class
382_N120-01_arth	<i>Nephropsis atlantica</i> (Class: Malacostraca)	17.62%	JQ305974.1	Class
449_N142-01_arth	<i>Pagurus prideaux/Pagurus venturensis</i> (Class: Malacostraca)	16.07%	JQ306249.1/GU442192	Class
449_N144-01_arth	<i>Pagurus prideaux/Pagurus venturensis</i> (Class: Malacostraca)	16.07%	JQ306249.1/GU442192	Class
435_N105-01_Arth2	<i>Periclimenes wirtzi</i> (Class: Malacostraca)	16.46%	KU065009.1	Class
435_N105-01_Arth5	<i>Periclimenes wirtzi</i> (Class: Malacostraca)	16.28%	KU065009.1	Class
ABTC132329_arth	<i>Pugettia gracilis</i> (Class: Malacostraca)	18.65%	KX039768.1	Class
395_N140-01_Arth4	<i>Pugettia gracilis</i> (Class: Malacostraca)	15.75%	KX039768.1	Class
ABTC132324	<i>Synalpheus fritzmuelleri</i> (Class: Malacostraca)	17.70%	KJ595081.1	Class
ABTC132327_arth	<i>Synalpheus fritzmuelleri</i> (Class: Malacostraca)	17.52%	KJ595081.1	Class

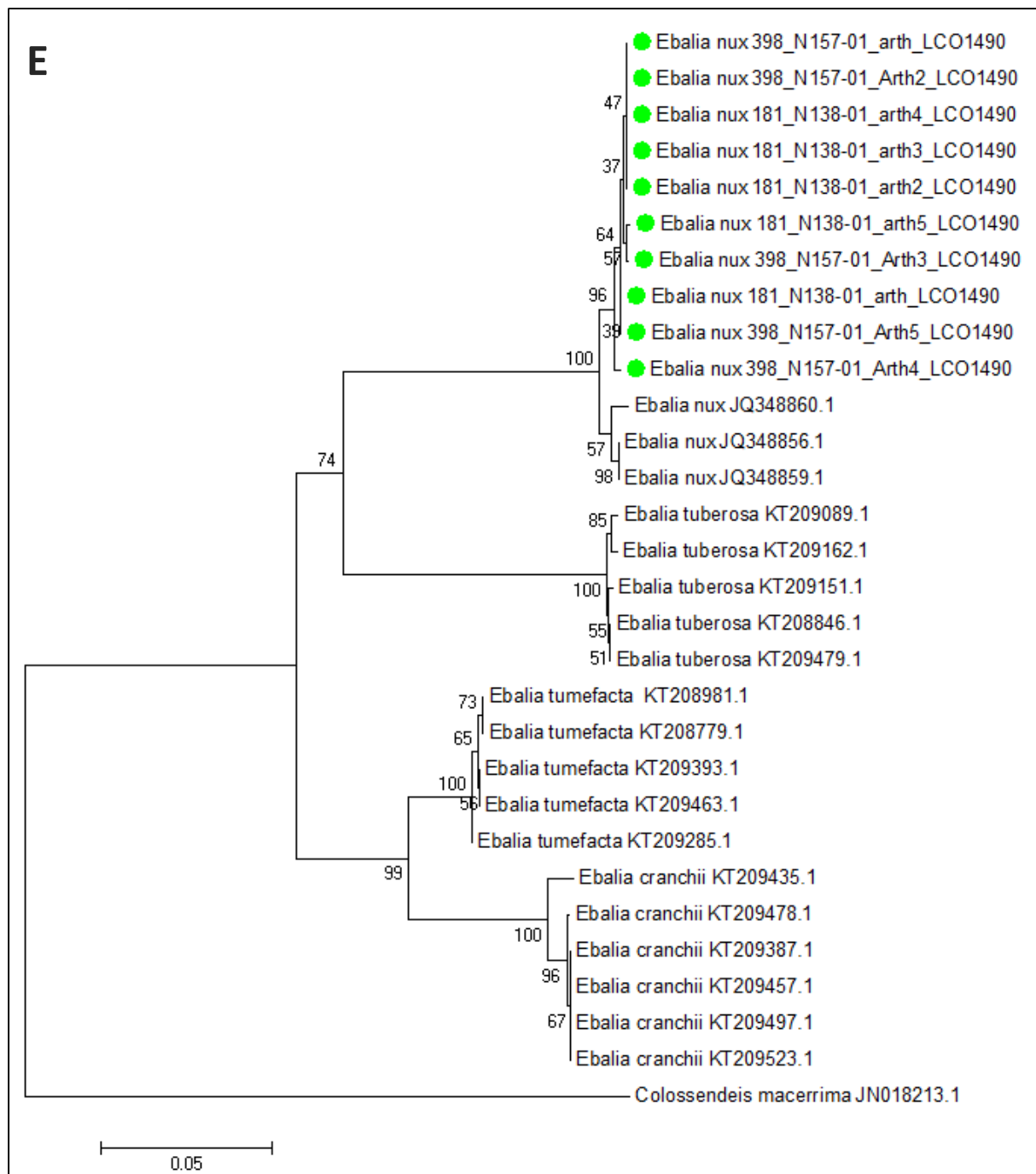
APPENDIX 3.3: Neighbour-Joining trees

Neighbour-Joining trees based on Kimura-2-paramter with 1000 bootstraps for genera used to calculate intra- and interspecific diversity. A: *Amphiophiura*, B: *Benthopecten*, C: *Ophiomyxa*, D: *Acanthephyra*, and E: *Ebalia*. Samples with coloured circles are from the current study. Green indicates that the sample was identified at species level using p-dis threshold.









4. SPATIAL DISTRIBUTION AND DIVERSITY IN HYDROCARBON DEGRADING MICROBES IN THE GREAT AUSTRALIAN BIGHT I: FUNCTIONAL GENE DISTRIBUTION

Sharon Hook¹, Jodie van de Kamp², Alan Williams², Jason E. Tanner³, Levente Bodrossy²

1. CSIRO Oceans and Atmosphere Lucas Heights, NSW 2234

2. CSIRO Oceans and Atmosphere Hobart, TAS 7001

3. SARDI Aquatic Sciences West Beach, SA 5024

Reprinted with minor amendments from:

Hook S, van de Kamp J, Williams A, Tanner JE, Bodrossy L (2016). Spatial distribution and diversity in hydrocarbon degrading microbes in the Great Australian Bight I: Functional Gene Distribution. Great Australian Bight Research Program. GABRP Research Report Number 8, Great Australian Bight Research Program, September 2016.

4.1 Executive summary

Bacteria are known to efficiently metabolise most components of oil, in particular alkanes and some aromatic compounds. They also efficiently metabolise much of the natural gas that is released during wellhead blowouts. As a consequence, being able to predict the metabolic capacity for oil degradation of bacteria indigenous to the Great Australian Bight (GAB) will be an important component of the oil spill risk assessment for this area. To address this in part, we studied key genes in three pathways of hydrocarbon degradation: *alkB*, which is involved in alkane degradation, *c23o* which is involved in the degradation of aromatic compounds, and *pmoA*, which metabolises methane. The diversity and relative abundance of these genes was determined in sediment and water samples collected from the GAB. We found numerous copies of each gene in all sediment samples, and of *alkB* and *c23o* in the water samples, suggesting that bacteria in the GAB have the capacity to degrade oil. Furthermore, the sequences of these genes were unique, suggesting that while these bacteria have the capacity to degrade hydrocarbons, their rates of response to sources of hydrocarbons, and thus rates of oil degradation, may not be easily predicted based on studies conducted elsewhere.

4.2 Introduction

Crude oil is a natural product derived from the degradation of organic material over geological time scales (Atlas & Hazen 2011). As a consequence, numerous groups of organisms have developed the ability to metabolise oil (Head et al. 2006, Atlas & Hazen 2011), including bacteria, algae and fungi. Following the recent *Deepwater Horizon* oil spill in the US Gulf of Mexico, approximately one half of the petroleum hydrocarbons released were oxidised by marine microbes (reviewed by Joye 2015). There are several factors that may have contributed to the rapid rates of degradation, including the chemical composition of the spilled oil (which was light and easily degraded) (Atlas & Hazen 2011), the high nutrient levels in the area, the high natural inputs of oil from seeps (Kappell et al. 2014, Yergeau et al. 2015), and the fact that the oil was dispersed, increasing the surface area that is available for oil degradation (Atlas & Hazen 2011, Prince 2015, Kleindienst et al. 2015a, b).

Since crude oil is a complex mixture of different chemicals, the degradation of crude oil is undertaken by a consortium of organisms using different pathways (reviewed in Head et al. 2006). The ability to degrade crude oil is found in diverse, taxonomically unrelated bacteria (Smith et al. 2013). Following the release of crude oil into the environment, the alkanes (linear arrangements of carbon) are degraded most rapidly (Abbasian et al. 2015). Alkane monooxygenase (also referred to as alkane hydroxylase, with the functional marker gene *alkB*) oxidises C₅ to C₁₆ length alkanes (Smith et al. 2013) and is the rate limiting step in this process. Polycyclic aromatic hydrocarbons (PAHs) are also readily degraded, although not as rapidly. The catechol-dioxygenases (including the catechol-2,3-dioxygenase, with the functional marker gene *c23o*; also referred to as *xylE*) are the enzymes responsible for cleaving the aromatic ring and a conserved step in the metabolism of PAHs (Meyer et al. 1999). This enzyme is thought to control the rate limiting step in the degradation of aromatic compounds (Okuta et al. 1998). Methane can also be released from production wells in large volumes. The enzyme methane mono-oxygenase (with the functional marker *pmoA*) catalyses the first step in the microbial conversion of methane to carbon dioxide under oxic conditions (Inagaki et al. 2004). Following the release of oil into the environment, these processes are thought to occur simultaneously, but to have patterns of relative abundance that proceed in an ecological succession, as depicted below. Indeed, following the *Deepwater Horizon* oil spill, this successional pattern was observed in the waters and sediments of the Gulf of Mexico (Dubinsky et al. 2013, reviewed in Yergeau et al. 2015).

Since the Great Australian Bight (GAB) may have substantial oil and gas reserves that could be developed in the future, we want to better understand the propensity for biodegradation following a potential release of oil into the environment. Consequently, we want to be able to predict the capacity of indigenous organisms in the GAB to degrade hydrocarbons – both as discharged as part of routine operations, and in the case of an oil spill. To that end, we characterised the diversity in three genes involved in different pathways in hydrocarbon degradation: *alkB*, involved in the oxidation of alkanes; *c23o*, involved in the “ring breaking” step of the metabolism of aromatic compounds, and *pmoA*, involved in methane metabolism. As our initial forays suggested that the diversity of sequences in the GAB was not well represented by sequences deposited in the National Center for Biotechnology Information (NCBI) databases, we used massively parallel high throughput sequencing (Illumina) to capture the diversity of sequences in sediment and water from the GAB. In a companion study, we examine the taxonomic composition of the microbial assemblages found in the sediments of the GAB (Chapter 5).

4.3 Materials and Methods

4.3.1 Sample collection

Sediment samples were collected from the surface of cores collected at the stations on transects T2 and T4 shown in Figure 4.1, taken during the *Southern Surveyor* cruise, 3-22 April 2013. A 6-core multicorer from KC (Denmark) was incorporated into an instrumented coring platform that could be controlled from the vessel and allowed reliable collection of sediment samples at depths between 200 and 2000 m (Sherlock et al. 2014). Triplicate sediment cores from each deployment were subsampled for microbial analysis using 30mm diameter minicores. The top 2 cm of each minicore was extruded, placed into a DNA free tube, and frozen until they were transferred to CSIRO laboratories for DNA extraction. Water samples were collected at maximum depth at each Station using a Niskin bottle on the coring platform (located ~ 1m off bottom). Microbial cells were collected by filtration of 2 L seawater through a 0.22 µm pore Sterivex™ GP filter (Millipore®, Massachusetts. Cat. # SVGPL10RC), using a 6 channel peristaltic pump. Pump tubing was rinsed with ~200ml seawater from the

appropriate depth prior to cell collection. Pumping continued for 1 min after the sample had cleared the filter to dry. Both ends of the filter were capped, placed in individual snap-lock bags and stored at -80°.

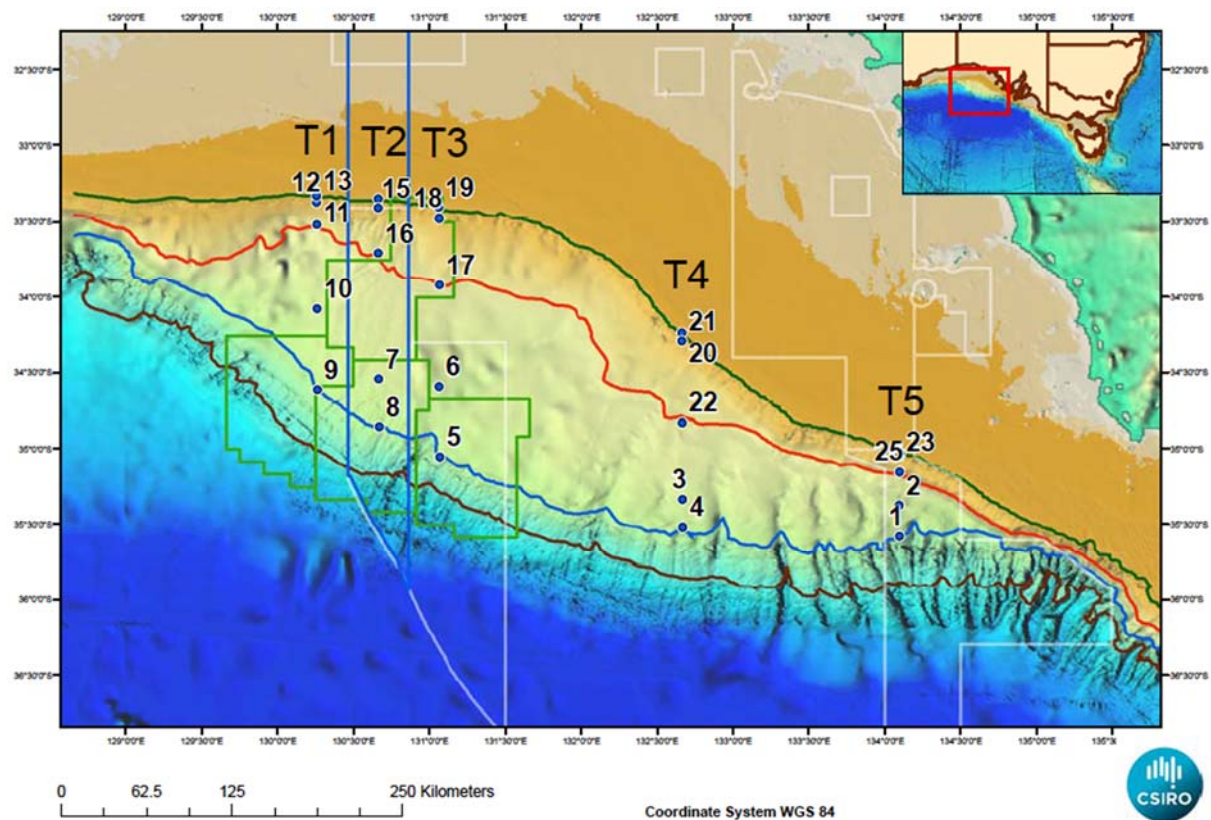


Figure 4.1. Locations sampled by the *RV Southern Surveyor* for sediments. Green, red, blue and black depth contours are 200, 1000, 2000 & 3000 m respectively, green polygons indicate the BP exploration leases, and the blue polygon the benthic protection zone of the Great Australian Bight Marine Park. T1, T2 etc indicate transect numbers.

4.3.2 DNA extractions

Sediment

Ten grams of sediment were used for each DNA extraction. DNA was extracted using the PowerMax® Soil DNA Isolation Kit (Mo Bio Laboratories Inc, USA), modified as follows: 10 minutes incubation at 70 °C after adding the lysis solution (C1), and extending the incubation times to 30 minutes. Once DNA was eluted, the sample was concentrated to dry pellets in a “speed vac” vacuum concentrator, and then washed in 100% ethanol to remove excess salts.

Water samples

DNA was extracted from sterivex filters following a modified version of the PowerWater® Sterivex™ DNA Isolation Kit (Mo Bio Laboratories Inc, USA) (Appleyard et al. 2011). Filters were removed from -80 °C and brought to room temperature before adding 1875 µL lysis buffer (200 mM sodium phosphate buffer, pH 7.0 containing: 1% CTAB, 2% PVP K30, 0.3 M NaCl and lysozyme at a final concentration of 5 mg mL⁻¹) and 125 µL MT Buffer (MP Biomedicals, LLC, USA) via the inlet valve. Sterivex filters were recapped and attached to a horizontal vortexer (Vortex-Genie 2, Mo Bio Laboratories Inc, USA) and vortexed at speed setting 6 for 1 h. A 3 mL syringe was attached to the inlet of the sterivex and using back pressure from the syringe the contents of the sterivex was removed and

split into 2 x 2.0 mL microfuge tubes (approximately 1 mL per tube). 900 µL of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) was added to each tube and mixed by inverting several times. After centrifuging at 13 000 rpm x 10 min, the supernatant was removed and combined in a new 2.0 mL microfuge. Proteinase K was added (20 µL of 20mg/mL stock) followed by incubation at 60 °C for 2 h. 500 µL of Chloroform:Isoamyl Alcohol (CI) (24:1) was added and mixed by inverting several times. After centrifuging at 13 000 rpm x 10 min, the supernatant was removed to a new 2.0 mL microfuge tube and the CI extraction step was repeated. Following centrifugation the supernatant was removed to a 5.0 mL tube and 3 mL of ST4 Buffer (prewarmed to 65 °C) was added. The barrel of a 20 mL syringe was attached to a filter column that was then attached to a vacuum manifold (Vac-Man® Laboratory Vacuum Manifold, Promega Corp, USA). The contents of the 5.0 mL tube was poured into the syringe barrel and pulled through the filter column using the vacuum. Once the entire volume had been pulled through the column the 20 mL syringe barrel was removed. With the vacuum still flowing the filter was washed with 500 µL ST5 followed by 500 µL ST6 and dried by continuing vacuum flow for a further 2 min after the ST6 was entirely pulled through. The vacuum was turned off and the filter transferred to a new 2.0 mL collection tube and allowed to air dry on the bench for 10 min. To elute DNA, the filter was incubated with 80 µL of 0.1 x TE at 37 °C for 45 min, followed by a final spin at 13 000 rpm x 2 min. The quality and quantity of all DNA was checked using a NanoDrop™ 8000 Spectrophotometer (Thermo Scientific™). DNA was aliquoted into multiple plates, vacuum dried and stored at -20 °C.

4.3.3 Database Construction

Sequences for *alkB*, *c23o* and *pmoA* were obtained from the NCBI Genbank databases. These sequences were aligned and phylogenetic trees were created using the ARB phylogenetic software package. Multiple probes were designed using principles outlined in previous studies (Stralis-Pavese et al. 2011, Abell et al. 2012, 2014). Upon validation, it became apparent that the diversity in the sequences in the GAB were not captured in the NCBI databases, even though these databases included several previous environmental surveys (Wasmund et al. 2009, Smith et al. 2013). As a consequence, we decided to capture the sequence diversity in water and sediment samples from the GAB via massively parallel sequencing.

4.3.4 Amplicon Generation and Sequencing

Amplicons for *alkB*, *pmoA* and *c23o* were sequenced using the Illumina MiSeq platform (Illumina, Inc., USA) for all samples, including replicates, from transects 2 and 4 (Figure 4.1). Illumina overhang adapter sequences were added to gene-specific primer pairs to generate amplicons for high throughput sequencing (Table 4.1). PCR reactions consisted of 2.5 µl 10x ImmoBuffer, 0.5 µl 10 mM dNTP, 1.25 µl 50 mM MgCl₂, 1 µl 10 µM forward primer, 1 µl 10 µM reverse primer, 0.25 µl BSA, 0.2 µl 5U/µl Immolase Polymerase and 10 ng DNA template in a total volume of 25 µl. Cycling parameters were: denaturation at 95 °C x 10 min; 11 cycles of 94 °C x 1 min, 65 °C x 1 min and 72 °C x 30 sec; followed by 24 cycles of 94 °C x 1 min, 60 °C x 1 min and 72 °C x 30 sec; and a final extension at 72 °C x 4 min. Amplicon products were purified using Agencourt AMPure XP (Beckman Coulter, Inc., USA) as per the manufacturer's instructions. Second stage PCR incorporating Nextera XT barcodes, purification, library generation and sequencing using the Illumina MiSeq platform (with 300 bp paired reads) were performed according to manufacturer's directions.

Table 4.1. Illumina adapter and primer sequences used to generate amplicons for high throughput sequencing.

	Forward sequence 5'-3'		Reverse sequence 5'-3'	Reference
Illumina forward overhang adapter	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[gene specific sequence]	Illumina reverse overhang adapter	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[gene specific sequence]	
alkB-1F	AAYACNGCNCAYGARCTNGGNCAYAA	alkB-1R	GCRTGRTGRTCNCGARTGNCGYTG	Kloos et al. 2006
pmoA189 GC	GGNGACTGGGACTTCTGG	pmoA682	GAASGCNGAGAAGAASGC	Holmes et al. 1995
c230F	AAGAGGCATGGGGGCGCACCGGTTTCATCA	c230R	CCAGCAAACACCTCGTTGCGGTTGCC	Sei et al. 1999

4.3.5 Bioinformatics

Amplicon sequences were analysed following the bioinformatics workflow established for the Bioplatforms Australia (BPA) funded Biome of Australian Soils (BASE) project (Bissett et al. 2016) (workflow details: <https://ccgapps.com.au/bpa-metadata/base/information>) with the following exceptions: after initial quality control, trimming and merging of read pairs, FASTA format sequences were extracted from FASTQ files and assembled into one file per amplicon. RDP FrameBot (Wang et al 2013) was used to detect frameshift errors in functional gene nucleotide sequences by comparing the DNA sequences to known protein sequences for each gene. FASTA files were submitted separately to the RDP FrameBot pipeline (fungene.cme.msu.edu/FunGenePipeline/framebot/form.spr) resulting in FASTA files containing only frameshift-corrected DNA sequences. These FASTA files were used to generate *de novo* OTUs at 97% similarity using the open reference OTU picking pipeline USEARCH 64 bit v8.0.1517 (Edgar 2010) with the UPARSE algorithm (Edgar 2013) and representative sequences for each OTU. Representative sequences were imported and aligned into the existing ARB databases for each gene as described above.

4.3.6 Data Analysis and Statistics

Square matrices containing presence and abundance data for all OTUs across all samples were generated using the python command `uc2otutab.py` provided on the drive5 Bioinformatics and software services site (<http://www.drive5.com/>) to assist in processing OTU data. Rarefaction curves were generated using the `rarefaction.single` command in MOTHUR (Schloss et al 2009). Preliminary community composition analysis was performed using the Primer-E Multivariate Statistics for Ecologists software package (Version 7; Clarke and Gorley 2015). For community composition analysis, the relative abundance data were fourth root transformed and a Bray-Curtis dissimilarity matrix constructed. Hierarchical agglomerative cluster analysis with the group average method was used to examine groupings in the data, which were then further visualised using principal co-ordinates analysis. Pearson correlations with relevant environmental data collected concordantly were overlaid on the later.

4.4 Results

4.4.1 Next-Gen Sequencing results

After QC and removal of chimeras, Illumina sequencing for *alkB* identified 1 417 985 sequences that could be grouped into 626 Operational Taxonomic Units (OTUs – sequences that are 97% similar, the functional equivalent of a species) from both water and sediment samples. Sequencing for *pmoA* produced 1 035 016 sequences from both water and sediment samples that could be grouped into 224 OTUs, and sequencing for *c23o* yielded 2 541 848 sequences from both water and sediment samples that were clustered into 940 OTUs. The rarefaction curves, shown in Figure 4.2 for *alkB*, demonstrate that we have captured the majority of diversity in the sampled environment with our sequencing efforts.

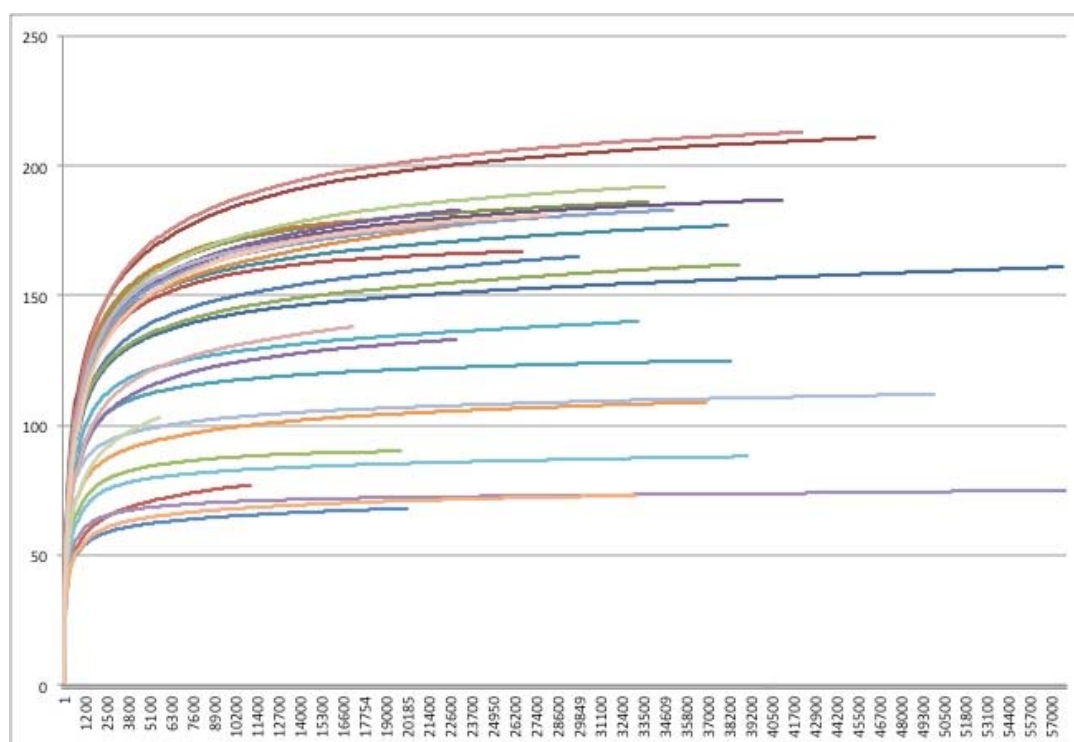


Figure 4.2. Rarefaction curve for gene sequence diversity for the *alkB* gene. Number of different sequences is shown on the y axis, and the number of sequencing reads are on the x-axis. Each station is shown in a different colour. Although additional new sequences are discovered with the addition of more sequencing depth, the curves have nearly plateaued, demonstrating that our efforts have captured the vast majority of environmental diversity.

4.4.2 Diversity overview

As depicted above, sediment samples from the GAB contained abundant and diverse gene copies of *alkB*, *c23o* and *pmoA*. An overview of this diversity is shown in Figure 4.3.

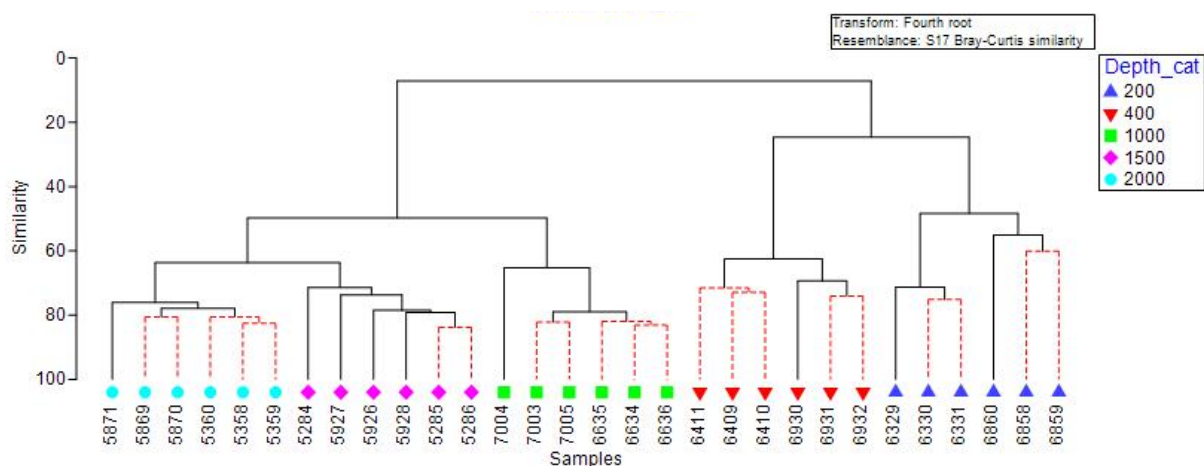


Figure 4.3. A cladogram showing sequence relatedness amongst sediment samples collected from the GAB. Percent similarity is on the vertical bar. Black lines indicate statistically significant, solid relationships. Red lines indicate relationships with limited statistical support.

Even samples collected from the same station (denoted by sequential numbers) had only 60-80% similarity in composition as determined by these functional genes. Most often, replicate samples formed a distinct cluster, and samples taken at the same depth on the different transects were more closely related than those taken at different depths. There is also a sharp contrast between samples collected in the shallower two depths (200-400 m) relative to the greater depths (1000-2000 m). The environmental factors that are correlated with these differences are shown in Figure 4.4.

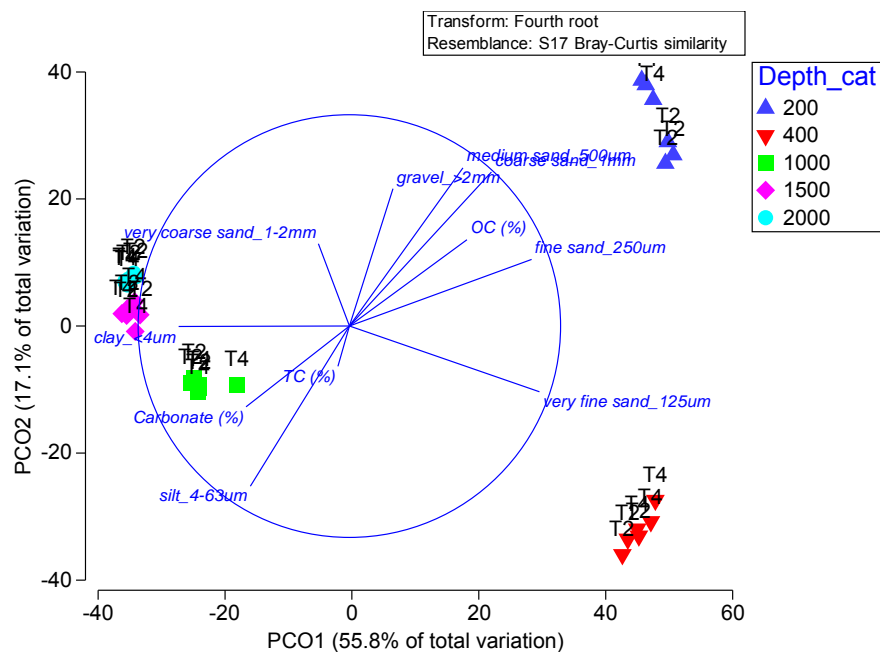


Figure 4.4. Principal co-ordinates plot showing variables that correlate with differences in diversity for the three selected functional genes, with replicates from each station plotted. For clarity, transects, not sample ID, are shown.

Again, it is apparent that station depth, not transect, is the greatest determinant of sample relatedness, as samples collected at the same depth cluster together, and there is the greatest distance between samples collected at 200-400 m and those collected at greater depths. Sediment grain size appears to be the primary environmental correlate with this difference. The differences observed in the communities of hydrocarbon degrading bacteria match the differences observed in the sediment grain size at different depths (Table 4.2). The samples collected at 200 m and 400 m have predominantly sandy sediments, whereas those collected from greater depths have a greater proportion of clay and less sand. As the levels of petroleum hydrocarbons detected in the sediments were consistently very low (Ahmed et al., 2014), we do not think they influence the microbial community composition.

Table 4.2. Sediment grain size distribution with depth for each of the sampling transects. Values are percent composition by volume up to 1 mm, and percent composition by weight for fractions > 1 mm.

Depth	clay_<4um	silt_4-63um	very sand_125um	fine sand_250um	medium sand_500um	coarse sand_1mm	very coarse sand_1-2mm	gravel_>2mm
200	0.16	25.84	19.41	20.35	16.67	17.56	2.48	5.25
	0.15	16.64	14.18	23.11	23.42	22.49	9.72	13.93
400	0.75	56.26	24.51	12.98	4.26	1.25	3.50	2.85
	0.64	44.59	23.05	19.08	8.80	3.85	0.38	0.72
1000	27.32	56.37	7.00	5.29	3.74	0.29	1.20	3.40
	36.82	52.72	6.07	2.97	1.40	0.02	6.93	6.58
1500	2.40	61.69	10.70	12.19	11.49	1.52	1.99	1.35
	40.77	40.64	5.55	5.55	6.23	1.26	7.00	4.30
2000	38.45	39.42	6.82	7.28	6.90	1.13	0.73	1.28
	37.61	42.96	5.62	6.09	6.54	1.16	12.56	8.26

The water samples collected are bottom water, and were collected from a Niskin bottle attached to the corer. The water samples had abundant gene copies for *alkB*, less abundant gene copies for *c23o*, and very few copies of *pmoA*. This is not unexpected, as methane oxidisers would be expected to be rare in oxic seawater.

The cladogram showing relatedness of different hydrocarbon degrading bacteria in water samples is shown below in Figure 4.5. For clarity, only relationships for *alkB* are shown, as these are very similar to the trends found for *c23o*. Again, the samples cluster with depth, with the water samples collected from depths of 1000 m or less forming a distinct clade relative to those collected from 1500 m or 2000 m. However, depth clustering is not as tight as was observed for the sediment samples, and the similarity between samples is lower.

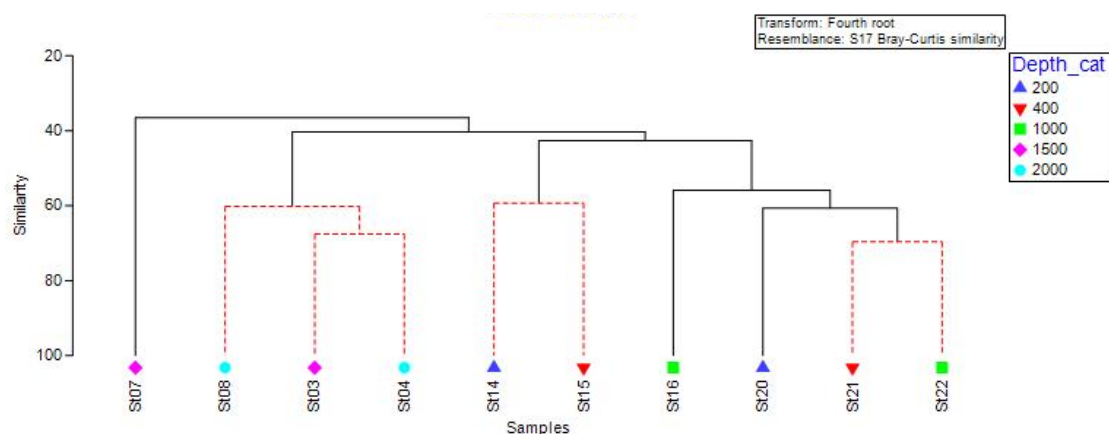


Figure 4.5. A cladogram showing *alkB* sequence relatedness amongst water samples collected from the GAB. Percent similarity is on the vertical bar. Black lines indicate statistically significant, solid relationships. Red lines indicate relationships with limited statistical support.

Similar trends are apparent if the data are grouped using principal co-ordinates analysis (Figure 4.6). Again, there is more divergence between sites at the same depth than was observed in the sediment samples, however, there is separation between the two greatest depths and the two shallowest depths, with 1000 m being somewhat in between.

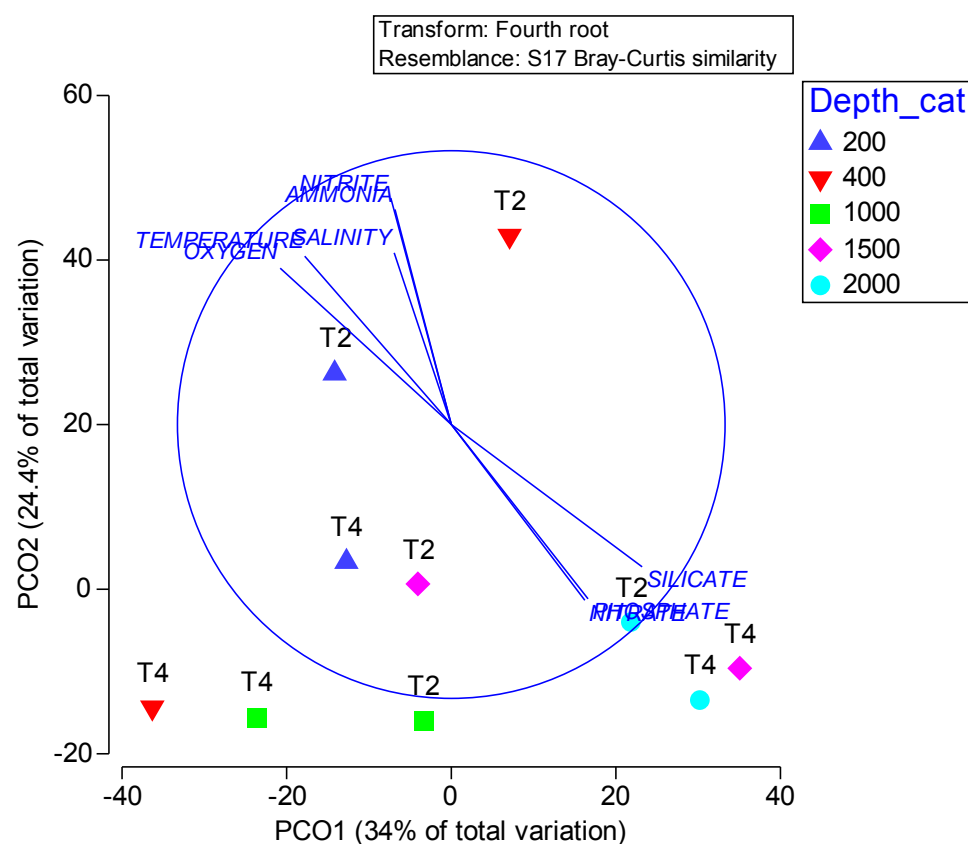


Figure 4.6. Principal co-ordinate analysis plot showing variables that correlate with differences in *alkB* sequence diversity.

These differences correspond with physico-chemical differences in the water column; dissolved oxygen, temperature, nitrite and ammonia on one hand, and silicate, phosphate and nitrate on the other. As shown in Figure 4.7, temperature and dissolved oxygen are both higher in the shallow waters, suggesting different water masses.

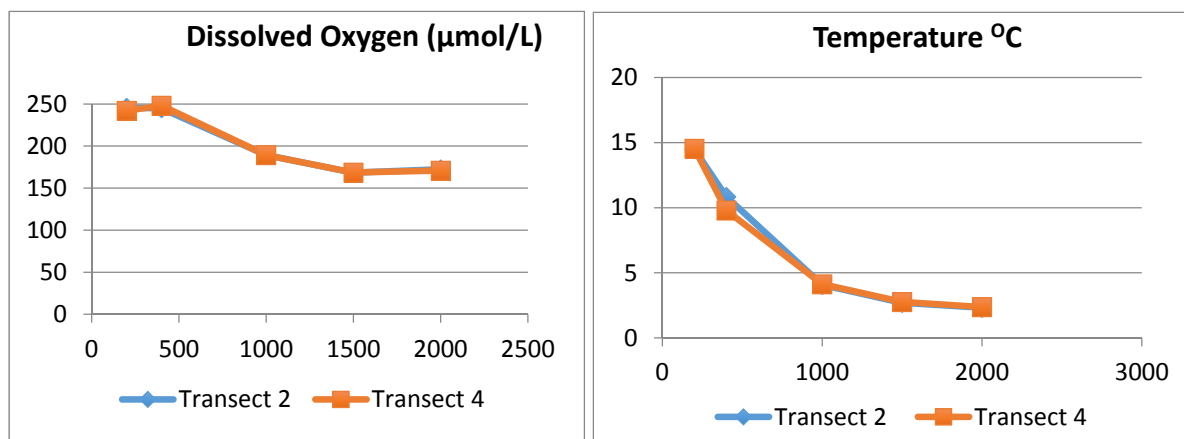


Figure 4.7. Physical characteristics of the water column for the transects surveyed.

Many of the sequences collected represent novel isoforms of the *alkB* gene that have not been previously described. A cladogram showing the relationships between the sequences identified in this study and those previously available is provided below in Figure 4.8. The total number of sequences, including the OTUs described in this study (similar sequences grouped together into an OTU at 97% similarity) is shown in black. The number of sequences identified in this project is shown in blue, and the number of sequences identified in the Gulf of Mexico is shown in brown. Clades are grouped at a high (family or above) level of taxonomic classification. It is apparent that the majority of sequences identified in this project are not found in the Gulf of Mexico, and that there are several novel high level clades that have not previously been identified from any environment. The insert shows a high diversity of novel genus level clades within one of the novel family level clades.

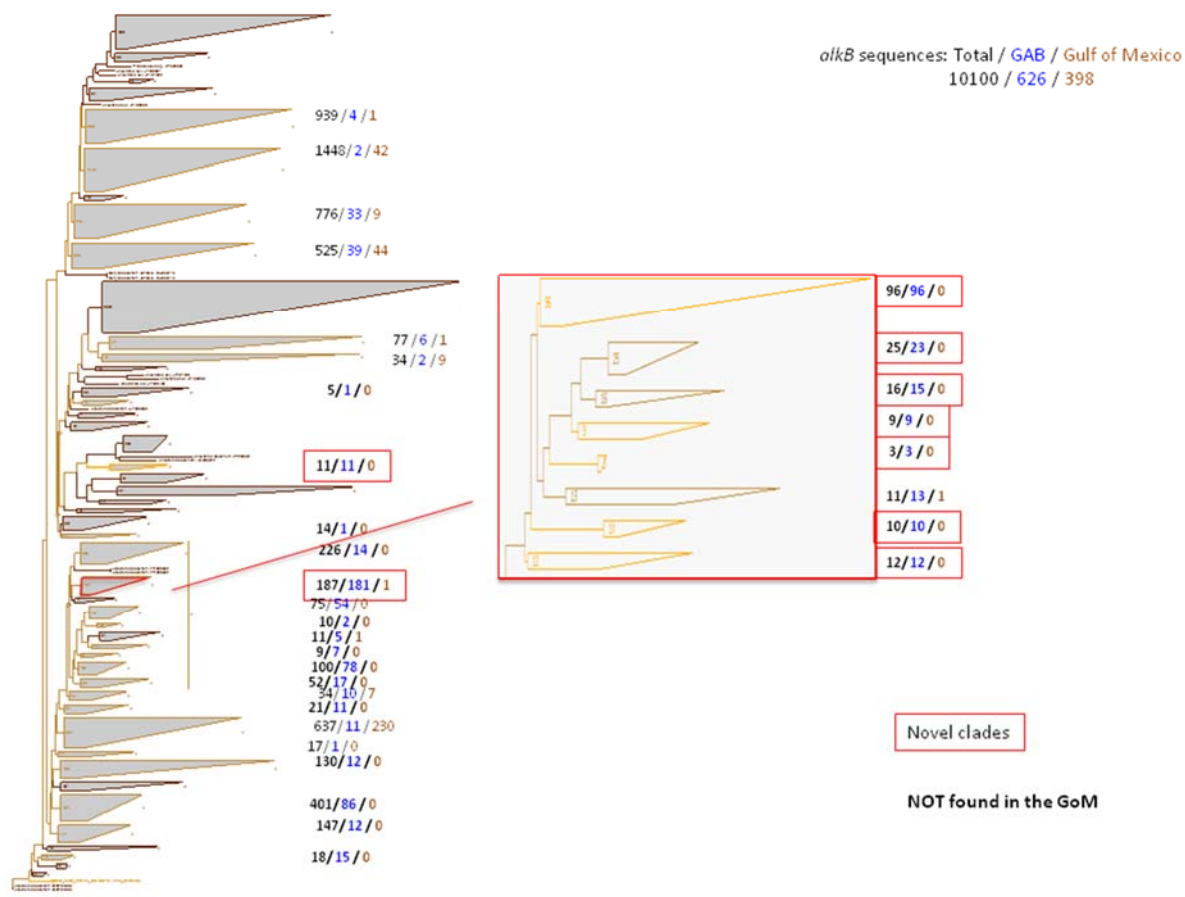


Figure 4.8. A cladogram, or hierarchical clustering tree based on phylogenetic relationships, based on nucleotide sequences of the *alkB* gene.

Similar results are observed when the phylogenetic relationships based on the nucleotide sequence of *pmoA* are examined (Figure 4.9). As *pmoA* is a comparatively well studied gene, there are many more sequences in the database. Nonetheless, four novel clades (i.e. clades that were not known at the onset of this work) were identified in this study, and there was low overlap between gene sequences for *pmoA* recorded in the GOM and those measured in this study.

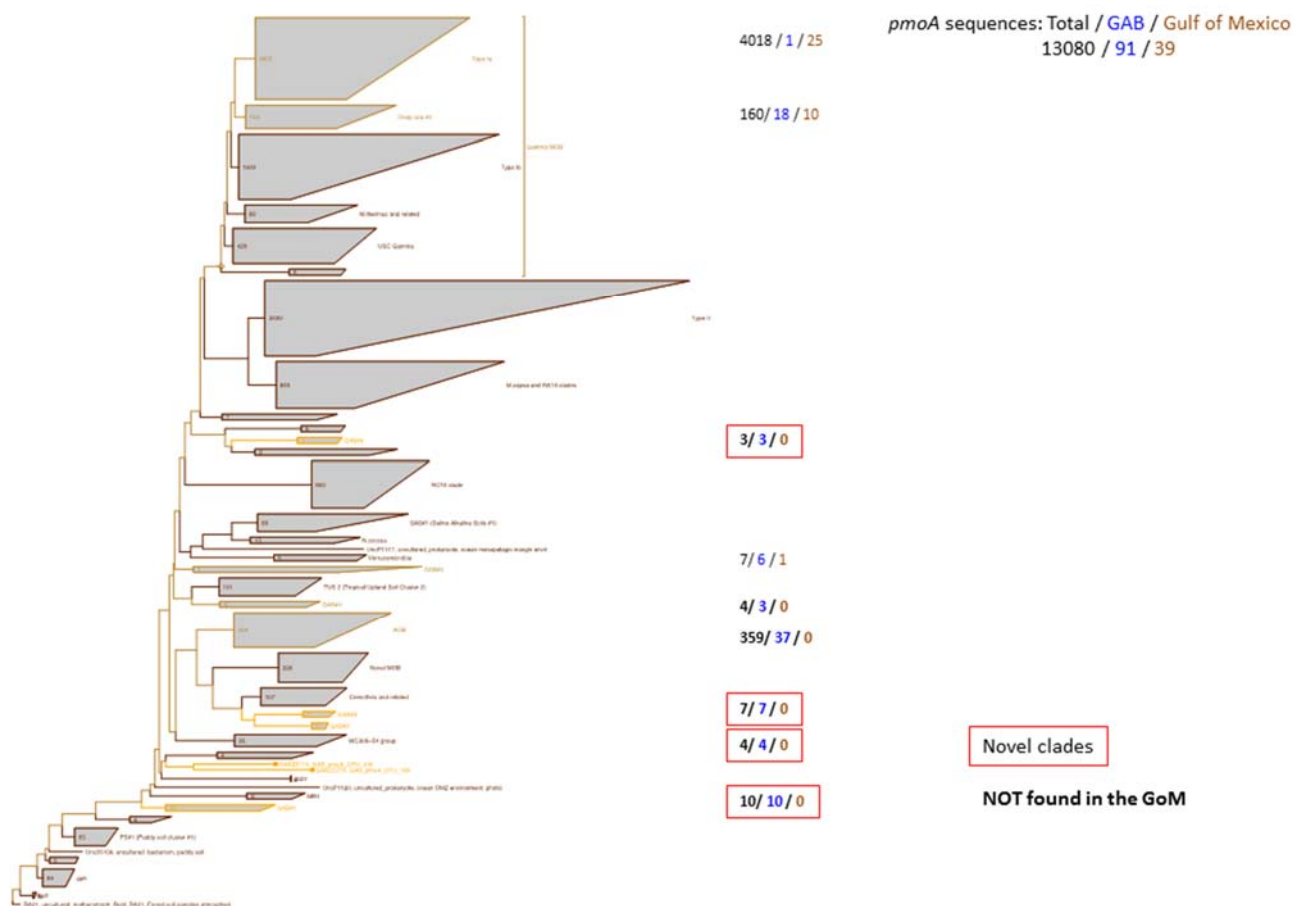


Figure 4.9. A cladogram based on nucleotide sequences of the *pmoA* gene.

The results were even more striking when the phylogenetic relationships for *c23o* are examined (Figure 4.10). Of the sequences for the gene currently in our database, one half were identified in the current project. Sixteen novel clades were identified, and none of the gene sequences had homology to those that were identified in the GOM. There are far less published sequences available for *c23o* compared to *alkB* and *pmoA*, as demonstrated by the fact that those collected in this study comprise approximately half of the entire database. That fact, combined with the overall observed novelty of *alkB* and *pmoA* sequences from this study, makes the very high level of observed novelty of *c23o* sequences not all that surprising.

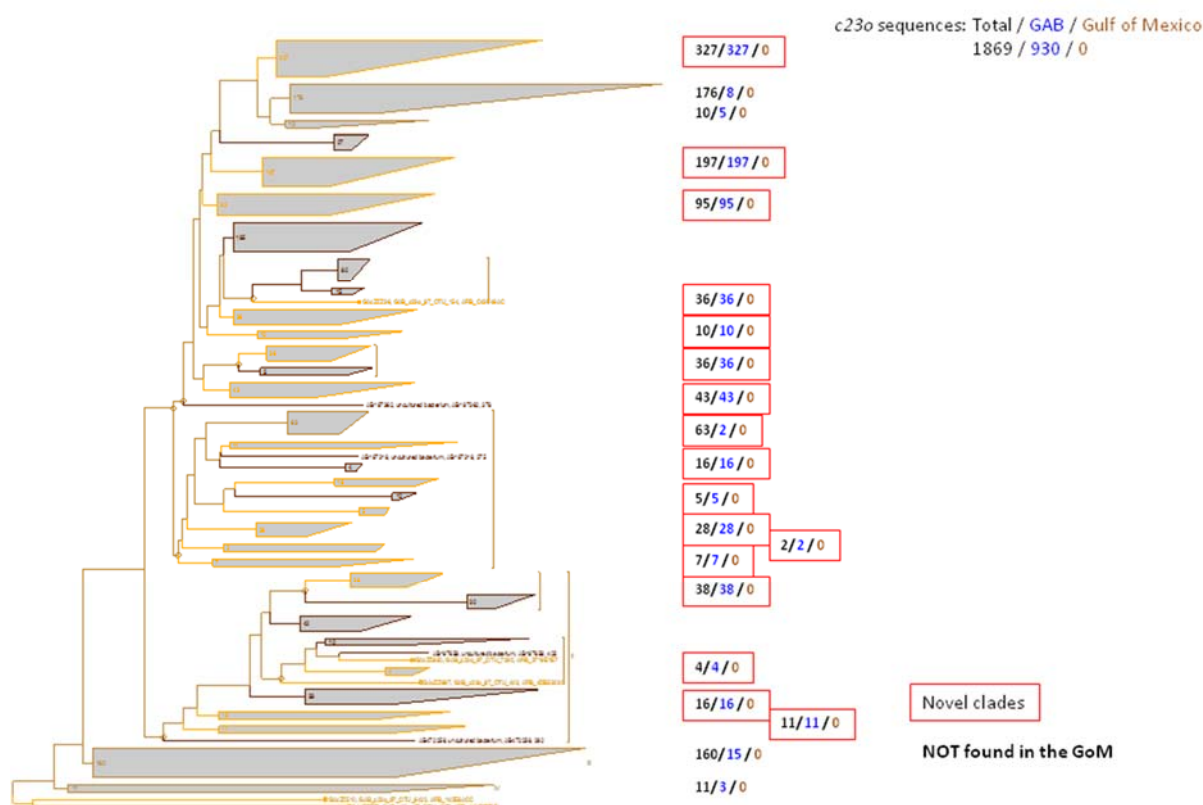


Figure 4.10. A cladogram based on nucleotide sequences of the *c23o* gene.

4.5 Discussion

These results indicate that the microbial communities present in the GAB have the capacity to metabolise the selected hydrocarbon fractions, even in the absence of an obvious petroleum source. This finding is consistent with the literature on hydrocarbon degrading bacteria. Marine ecosystems normally contain hydrocarbon degrading bacteria (Head et al. 2006, Kostka et al. 2014). Typically, hydrocarbon degrading bacteria make up 1% of the bacterial assemblage in pristine environments, and increase to about 10% in contaminated environments (Atlas 1995).

The gene sequences for all three genes identified in this study are substantially different than what has been seen before. This is consistent with other studies conducted in this region that have identified high degrees of endemism. These differences suggest that while the bacterial assemblages present in the GAB have the capacity to degrade hydrocarbons, they may have different rates of responses to sources of hydrocarbons, and thus degradation may occur at different rates to what has been documented elsewhere. There are also large differences between the GAB and the GOM – which has been well studied due to the recent *Deepwater Horizon* oil spill. The GAB has low levels of nutrients and no obvious hydrocarbon seeps, whereas the GOM has high nutrients, due in part to input from the Mississippi River, and abundant natural inputs of hydrocarbons from seeps. Given these differences, it's not surprising that the hydrocarbon degrading microbial communities may have different capacities to degrade petroleum, as evident by the high numbers of novel sequences. These differences also could reflect the novel approach we are taking using a targeted approach to do deep sequencing of functional genes. Most previous studies examining these genes have either used

traditional sequencing approaches (which produce only a few hundred sequences) (Wasmund et al. 2009, Smith et al. 2013) or have been conducted using the GeoChip (Lu et al. 2012), although some studies have used metatranscriptomic approaches (Rivers et al. 2013). We would expect that the vastly increased amounts of information provided by the approach used in this study would provide a more thorough accounting of the microbial community.

Analysis of clone libraries from sediments collected from hydrocarbon seeps in the Timor Sea also indicated a high divergence from known *alkB* sequences (Wasmund et al. 2009). This study analysed fewer depths (approximately 100-400 m), sediment only, and fewer sequences via clone libraries, but nonetheless, had similar trends. They found a high diversity of *alkB* sequences, and a high proportion of highly novel sequences. They found that the diversity of *alkB* sequences decreased with depth. In the GAB, different taxa were more or less abundant at different depths, but our initial analysis does not suggest any decreases in *alkB* diversity with depth. They did not find any correlation between *alkB* diversity and alkane concentration in the sediments, but they did find a higher abundance of *alkB* copy numbers near hydrocarbon seeps, even if the alkane concentrations in sediments were low, suggesting that this gene is a good marker for inputs of hydrocarbons into the environment (Wasmund et al. 2009). By contrast, water samples collected from the northern Gulf of Mexico prior to the *Deepwater Horizon* oil spill had fewer sequences, but again, a high proportion of novel sequences (Smith et al. 2013). These authors found no obvious trends in *alkB* community composition with depth, but their greatest depths were less than 1000 m. Also, they used a clone based approach, meaning they had much less sequencing information to work with (Smith et al. 2013).

Identifying bacteria with the capacity to degrade oil, as well as developing a means of rapidly assessing changes in their abundance, will be an important component of the response and assessment in the event of any oil spills in the area. Previous studies have identified the selected genes as good targets for this sort of an approach. Work conducted following the 2010 *Deepwater Horizon* oil spill, as well as work that has been carried out in contaminated environments elsewhere, has shown an increase in the abundance of these genes. *alkB* and unspecified genes involved in aromatic hydrocarbon degradation were determined to be more abundant in water samples collected from within the hydrocarbon plume than outside it (Lu et al. 2012). Both *alkB* and *c23o* were enriched (though not significantly) in transcriptomes from plume samples (Rivers et al. 2013). Catechol-2,3-dioxygenase was elevated in samples collected from a beach impacted by the *Deepwater Horizon* oil spill relative to an unimpacted beach from the same area (Kappell et al. 2014). A year after the spill, however, no differences in the abundance of alkane monooxygenases or ring cleaving dioxygenases were measured, even though high molecular weight alkanes were more abundant in surface sediments near the wellhead (Yergeau et al. 2015). Studies conducted after the *Deepwater Horizon* oil spill also identified increased rates of methane oxidation and an increased abundance of methane oxidising bacteria in the months immediately after the wellhead blowout, which is not surprising given the high amounts of methane released after the spill (Crespo-Medina et al. 2014). The rapid increase in methanotrophy was attributed to previously undescribed sequences. These sequences were from a clade with low affinity for methane but high oxidation rates (Crespo-Medina et al. 2014).

These genes have also been shown to be good predictors of bacterial abundances in petroleum contaminated sites that are not associated with a major oil spill. An “overabundance” of *alkB* genes was found to be the best predictor of oil contaminated sites in the Mediterranean, and chronically contaminated sites in the Mediterranean Sea had an increased abundance of catechol-2,3-dioxygenase relative to catechol-1,2-dioxygenase (Bargiela et al. 2015). Increased abundance of *alkB* and *c23o* was measured in soils collected from different oil field contaminated soils in China (Liang et al. 2011).

Although we can measure differences in the structure of these genes, we do not know if these changes would be manifested in differences in how these bacteria respond to petroleum hydrocarbons in the environment, or the rates of microbial oxidation of these hydrocarbons in the GAB as a consequence. This uncertainty is exacerbated by not having a sample near a natural source of oil to compare changes in abundance to the background samples. We would suggest that future studies could perform enrichments of water and sediment collected from the GAB to compare both the rates of hydrocarbon oxidation and the microbial response to oil to the rates and response elsewhere in the world.

4.6 References

- Abbasian F, Lockington R, Mallavarapu M, Naidu R (2015) A Comprehensive Review of Aliphatic Hydrocarbon Biodegradation by Bacteria. *Applied Biochemistry and Biotechnology* 176:670-699
- Abell GCJ, Robert SS, Frampton DMF, Volkman JK, Rizwi F, Csontos J, Bodrossy L (2012) High-Throughput Analysis of Ammonia Oxidiser Community Composition via a Novel, amoA-Based Functional Gene Array. *PloS one* 7: e51542.
- Abell GCJ, Stralis-Pavese N, Pan Y, Bodrossy L (2014) Analysis of Methanotroph Community Structure Using a pmoA-Based Microarray. In: Paulsen IT, Holmes AJ (eds) *Environmental Microbiology: Methods and Protocols*, 2nd Edition Springer, New York.
- Ahmed M, Ross A, Stalvies C, Armand S, Fuentes D, Gong S, Sestak S (2014) Organic geochemical study of seawater, seabed sediment and headspace gas samples to monitor baseline hydrocarbon levels around BP permits in the Great Australian Bight, Australia. Report to the BP-GAB alliance.
- Appleyard SA, Abell GA & Watson R (2013) Tackling microbial related issues in cultured shellfish via integrated molecular and water chemistry approaches. Final Report to the Australian Seafood CRC
- Atlas RM (1995) Petroleum biodegradation and oil spill bioremediation. *Marine pollution bulletin* 31:178-182
- Atlas RM, Hazen TC (2011) Oil biodegradation and bioremediation: a tale of the two worst spills in U.S. history. *Environmental Science & Technology* 45:6709-6715
- Bargiela R, Mapelli F, Rojo D, Chouaia B, Tornes J, Borin S, Richter M, Del Pozo MV, Cappello S, Gertler C, Genovese M, Denaro R, Martinez-Martinez M, Fodelianakis S, Amer RA, Bigazzi D, Han X, Chen J, Chernikova TN, Golyshina OV, Mahjoubi M, Jaouanil A, Benzha F, Magagnini M, Hussein E, Al-Horani F, Cherif A, Blaghen M, Abdel-Fattah YR, Kalogerakis N, Barbas C, Malkawi HI, Golyshin PN, Yakimov MM, Daffonchio D, Ferrer M (2015) Bacterial population and biodegradation potential in chronically crude oil-contaminated marine sediments are strongly linked to temperature. *Scientific Reports* 5: 11651
- Bissett A, Fitzgerald A, Meintjes T, Mele PM, Reith F, Dennis PG, Breed MF, Brown B, Brown MV, Brugger J, Byrne M, Caddy-Retalic S, Carmody B, Coates DJ, Correa C, Ferrari BC, Gupta VVSR, Hamonts K, Haslem A, Hugenholtz P, Karan M, Koval J, Lowe AJ, Macdonald S, McGrath L, Martin D, Morgan M, North KI, Paungfoo-Lonhienne C, Pendall E, Phillips L, Pirzl R, Powell JR, Ragan MA, Schmidt S, Seymour N, Snape I, Stephen JR, Stevens M, Tinning M, Williams K, Yeoh YK, Zammit CM, Young A (2016) Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. *GigaScience* 5:21

- Bray JR & Curtis J T (1957) An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27, 326-349
- Clarke KR (1993) Nonparametric Multivariate Analyses of Changes in Community Structure. *Australian Journal of Ecology*, 18, 117-143. doi: 10.1111/J.1442-9993.1993.Tb00438.X
- Clarke KR & Gorley RN (2015) PRIMER v7: User Manual/Tutorial (v7 ed.). Plymouth, UK: PRIMER-E Ltd
- Chatfield C & Collins AJ (1980) Introduction to multivariate analysis. London, UK: Chapman and Hall
- Crespo-Medina M, Meile CD, Hunter KS, Diercks AR, Asper VL, Orphan VJ, Tavormina PL, Nigro LM, Battles JJ, Chanton JP, Shiller AM, Joung DJ, Amon RMW, Bracco A, Montoya JP, Villareal TA, Wood AM, Joye SB (2014) The rise and fall of methanotrophy following a deepwater oil-well blowout. *Nature Geoscience* 7:423-427
- Dubinsky EA, Conrad ME, Chakraborty R, Bill M, Borglin SE, Hollibaugh JT, Mason OU, Piceno YM, Reid FC, Stringfellow WT, Tom LM, Hazen TC, Andersen GL (2013) Succession of Hydrocarbon-Degrading Bacteria in the Aftermath of the Deepwater Horizon Oil Spill in the Gulf of Mexico. *Environmental Science & Technology* 47:10860-10867
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26: 2460-2461. doi: 10.1093/bioinformatics/btq461
- Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10: 996-998. doi: 10.1038/NMETH.2604
- Head IM, Jones DM, Roling WFM (2006) Marine microorganisms make a meal of oil. *Nature Reviews Microbiology* 4:173-182
- Holmes AJ, Costello A, Lidstrom ME, & Murrell JC (1995) Evidence that particulate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. *FEMS Microbiology Letters*, 132: 203-208. doi: 10.1111/J.1574-6968.1995.Tb07834.X
- Inagaki F, Tsunogai U, Suzuki M, Kosaka A, Machiyama H, Takai K, Nunoura T, Nealson KH, Horikoshi K (2004) Characterization of C-1-metabolizing prokaryotic communities in methane seep habitats at the Kuroshima Knoll, southern Ryukyu arc, by analyzing pmoA, mmoX, mxaF, mcrA, and 16S rRNA genes. *Applied and Environmental Microbiology* 70:7445-7455
- Joye SB (2015) Deepwater Horizon, 5 years on. *Science* 349:592-593
- Kappell AD, Wei Y, Newton RJ, Van Nostrand JD, Zhou J, McLellan SL, Hristova KR (2014) The polycyclic aromatic hydrocarbon degradation potential of Gulf of Mexico native coastal microbial communities after the Deepwater Horizon oil spill. *Frontiers in Microbiology* 5: 205.
- Kleindienst S, Paul JH, Joye SB (2015a) Using dispersants after oil spills: impacts on the composition and activity of microbial communities. *Nature Reviews Microbiology* 13:388-396
- Kleindienst S, Seidel M, Ziervogel K, Grim S, Loftis K, Harrison S, Malkin SY, Perkins MJ, Field J, Sogin ML, Dittmar T, Passow U, Medeiros PM, Joye SB (2015b) Chemical dispersants can suppress the activity of natural oil-degrading microorganisms. *Proceedings of the National Academy of Sciences* doi/10.1073/pnas.1507380112
- Kloos, K, Munch, JC, & Schlöter, M (2006) A new method for the detection of alkane-monooxygenase homologous genes (alkB) in soils based on PCR-hybridization. *Journal of Microbiological Methods*, 66: 486-496. doi: 10.1016/j.mimet.2006.01.014
- Kostka JE, Teske A, Joye SB, Head IM (2014) The metabolic pathways and environmental controls of hydrocarbon biodegradation in marine ecosystems. *Frontiers in Microbiology* 5: 471

- Liang Y, Van Nostrand JD, Deng Y, He Z, Wu L, Zhang X, Li G, Zhou J (2011) Functional gene diversity of soil microbial communities from five oil-contaminated fields in China. *The ISME Journal* 5:403-413
- Lu Z, Deng Y, Van Nostrand JD, He Z, Voordeckers J, Zhou A, Lee Y-J, Mason OU, Dubinsky EA, Chavarria KL, Tom LM, Fortney JL, Lamendella R, Jansson JK, D'Haeseleer P, Hazen TC, Zhou J (2012) Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *The ISME Journal* 6:451-460
- Meyer S, Moser R, Neef A, Stahl U, Kampfer P (1999) Differential detection of key enzymes of polyaromatic-hydrocarbon-degrading bacteria using PCR and gene probes. *Microbiology-Uk* 145:1731-1741
- Okuta A, Ohnishi K, Harayama S (1998) PCR isolation of catechol 2,3-dioxygenase gene fragments from environmental samples and their assembly into functional genes. *Gene* 212:221-228
- Prince RC (2015) Oil spill dispersants: boon or bane? *Environmental Science & Technology* 49:6376-6384
- Rivers AR, Sharma S, Tringe SG, Martin J, Joye SB, Moran MA (2013) Transcriptional response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill. *The ISME journal* 7:2315-2329
- Schloss, PD, Westcott, SL, Ryabin, T, Hall, JR, Hartmann, M, Hollister, EB, Weber, CF (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, 75: 7537-7541. doi: 10.1128/AEM.01541-09
- Sei, K, Asano, K, Tateishi, N, Mori, K, Ike, M, & Fujita, M (1999) Design of PCR primers and gene probes for the general detection of bacterial populations capable of degrading aromatic compounds via catechol cleavage pathways. *Journal of Bioscience and Bioengineering*, 88: 542-550. doi: 10.1016/S1389-1723(00)87673-2
- Smith CB, Tolar BB, Hollibaugh JT, King GM (2013) Alkane Hydroxylase Gene (alkB) Phylotype Composition and Diversity in Northern Gulf of Mexico Bacterioplankton. *Frontiers in Microbiology* 4: 370
- Stralis-Pavese N, Abell GCJ, Sessitsch A, Bodrossy L (2011) Analysis of methanotroph community composition using a pmoA-based microbial diagnostic microarray. *Nature Protocols* 6:609-624
- Wang, Q, Quensen, JF, Fish, JA, Lee, TK, Sun, YN, Tiedje, JM, & Cole, JR (2013) Ecological Patterns of nifH Genes in Four Terrestrial Climatic Zones Explored with Targeted Metagenomics Using FrameBot, a New Informatics Tool. *Mbio*, 4(5). e00592-1310. doi:10.1128/mBio.00592-13
- Wasmund K, Burns KA, Kurtboeke DI, Bourne DG (2009) Novel Alkane Hydroxylase Gene (alkB) Diversity in Sediments Associated with Hydrocarbon Seeps in the Timor Sea, Australia. *Applied and Environmental Microbiology* 75:7391-7398
- Yergeau E, Maynard C, Sanschagrin S, Champagne J, Juck D, Lee K, Greer CW (2015) Microbial Community Composition, Functions, and Activities in the Gulf of Mexico 1 Year after the Deepwater Horizon Accident. *Applied and Environmental Microbiology* 81:5855-5866

5. SPATIAL DISTRIBUTION AND DIVERSITY IN HYDROCARBON DEGRADING MICROBES IN THE GREAT AUSTRALIAN BIGHT II: ANALYSIS OF BACTERIAL AND ARCHAEOAL COMMUNITY STRUCTURE

Sharon Hook¹, Jodie van de Kamp², Alan Williams², Jason E. Tanner³, Levente Bodrossy²

1. CSIRO Oceans and Atmosphere Lucas Heights, NSW 2234

2. CSIRO Oceans and Atmosphere Hobart, TAS 7001

3. SARDI Aquatic Sciences West Beach, SA 5024

Reprinted with minor amendments from:

Hook S, van de Kamp J, Williams A, Tanner JE, Bodrossy L (2016). Spatial distribution and diversity in hydrocarbon degrading microbes in the Great Australian Bight II: Analysis of bacterial and archaeal community structure. Great Australian Bight Research Program. GABRP Research Report Number 9, Great Australian Bight Research Program, September 2016.

5.1 Executive summary

Following an oil spill, most of the oil released into the environment is metabolised by bacteria and other microorganisms. Knowing the relationship of indigenous bacteria within the Great Australian Bight (GAB) to those known to degrade hydrocarbons elsewhere will aid in the assessment of, and response to, an unplanned release of hydrocarbons in the environment. To this end, we analysed the community composition of Bacteria and Archaea in sediment samples collected from sites within the GAB. We found a diverse community of bacteria in all samples. These included some bacteria related to those with a known capacity to degrade hydrocarbons and that have been shown to increase in abundance in response to previous spills in other locations. By contrast, the Archaea were less diverse, and seem to be dominated by fewer operational taxonomic units. The spatial patterns in the bacterial and archaeal assemblages were very similar, suggesting that both are influenced by the same environmental parameters. The sediment and water samples had different microbial communities, and the sediment samples were clustered more tightly, indicating a higher degree of relatedness. The microbial communities in sediment samples collected from the continental shelf (200 and 400 m) differed from those collected from the continental slope (depths greater than 1000 m). For the benthic samples, diversity and species evenness decreased with depth. In general, community composition related best to depth and sediment grain size in benthic samples and to depth and nutrients in pelagic samples. This report provides important information about the baseline indigenous microbial communities present in areas of the GAB without a hydrocarbon source.

5.2 Introduction

Most oil that is released into the environment is ultimately degraded by microorganisms (Head et al. 2006). For example, following the recent *Deepwater Horizon* oil spill, marine bacteria were found to respond very quickly to the spill, with known hydrocarbon degrading bacteria significantly more abundant inside the plume than outside it (Hazen et al. 2010). It is estimated that approximately half of the oil that was released during the spill was metabolised by microorganisms (Joye 2015). The Gulf

of Mexico has high natural inputs of petroleum hydrocarbons from seeps, and as a consequence, a high abundance of hydrocarbon degrading bacteria (Kappell et al. 2014). However, hydrocarbon degrading microorganisms are widespread in the environment, normally making up 1% of the microbial population in pristine environments, increasing to 10% in environments with a petroleum hydrocarbon source (Atlas 1995).

Since oil is a complex mixture of more than 10,000 different chemical compounds, different microorganisms have developed the ability to degrade different constituents of oil over evolutionary time (Head et al. 2006). Numerous bacteria can degrade hydrocarbons – oil degrading bacteria have been isolated from the α , β , δ , and γ Proteobacteria and the Firmicutes (Head et al. 2006). Alkanes, or linear arrangements of carbon and hydrogen, are the most labile fraction of oil. Alkane degraders, such as *Alcanivorax*, *Oceanospirillaceae*, *Pseudomonas* and *Marinobacter*, are typically the fastest responding microorganisms to a new source of oil. Polycyclic aromatic hydrocarbons (PAH) are also readily degraded by microorganisms, and bacteria such as *Colwellia* and *Cycloclasticus* are typically abundant in areas with aromatic hydrocarbons present (Dubinsky et al. 2013, Kleindienst et al. 2015). Gas is also associated with the breakdown of some petroleum products and may be discharged from petroleum wells. Microorganisms such as *Methylobomonas* oxidise methane to carbon dioxide in oxic environments. Following the *Deepwater Horizon* oil spill, a “phylochip” array was used to measure changes in microbial species abundance. Although all taxa were active in the hydrocarbon enriched plume throughout the response, changes in relative abundance matched patterns that would be predicted by changes in the hydrocarbon chemistry (Dubinsky et al. 2013).

Since the existing microbial community structure likely controls the rate at which an ecosystem can respond to an oil spill (Kostka et al. 2014) our overall goal is to be able to predict the capacity of indigenous organisms in the Great Australian Bight (GAB) to degrade oil – both as discharged as part of routine operations, and in case of an oil spill. The presence of a high abundance and diversity of microbial oil degraders may also point to the natural presence of oil in the system, and provide pointers to the regions prospectivity. To that end, we characterised the prokaryotic diversity using high throughput 16S rDNA gene sequencing. We used the 16S gene sequences to identify the proportion of the community that were known or expected oil degraders, and how they are distributed in both the sediment and water column prior to the commencement of drilling. Differences in the community structure are related to available environmental parameters, such as depth, sediment grain size and temperature. In a complementary report, we examine the presence of specific gene sequences encoding enzymes that are involved in key steps in conserved pathways for the degradation of different fractions of oil from the same samples (Chapter 4).

5.3 Material and methods

5.3.1 Sample collection

Sediment samples were collected from the surface of cores collected at the stations shown in Figure 5.1, taken during the *Southern Surveyor* cruise, 3-22 April 2013, and a second cruise from 30 November to 21 December 2015 conducted on the *RV Investigator*. A 6-core multicorer from KC (Denmark) was incorporated into an instrumented coring platform (ICP) that could be controlled from the vessel and allowed reliable collection of sediment samples at depths between 200 and 3000 m (Sherlock et al. 2014). Triplicate sediment cores from each deployment were subsampled for microbial analysis using 30 mm diameter minicores. The top 2 cm of each minicore was extruded, placed into a DNA free tube, and frozen until they were transferred to CSIRO laboratories for DNA extraction. A fourth core from each deployment was sampled for sediment grain size and carbon content. Water samples were collected at maximum depth at each Station using a Niskin bottle on the ICP (located ~ 1 m off bottom).

Microbial cells were collected by filtration of 2 L seawater through a 0.22 μm pore Sterivex™ GP filter (Millipore®, Massachusetts. Cat. # SVGPL10RC), using a 6 channel peristaltic pump. Pump tubing was rinsed with ~200ml seawater from the appropriate depth prior to cell collection. Pumping continued for 1 min after the sample had cleared the filter to dry. Both ends of the filter were capped, placed in individual snap-lock bags and stored at -80°.

Near bottom nutrients (nitrate, nitrite, ammonia, phosphate and silicate), as well as temperature, salinity and oxygen levels, were obtained from a separate CTD cast taken at each station.

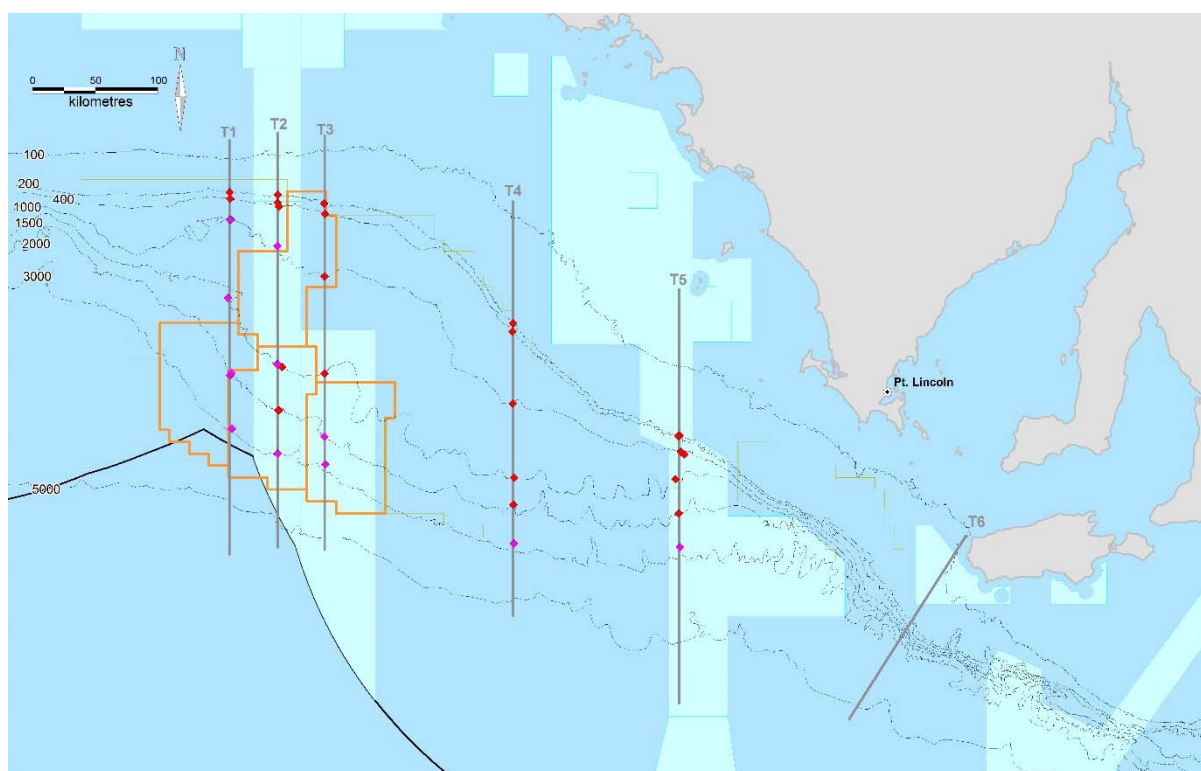


Figure 5.1. Transects and sampling locations used in the 2013 and 2015 cruises. Sampling locations used for development of the functional assays are shown in red. BP's lease blocks are shown in orange, and the light blue shading indicates Commonwealth marine reserves.

5.3.2 DNA extractions

Sediment

Ten grams of sediment were used for each DNA extraction. DNA was extracted using the PowerMax® Soil DNA Isolation Kit (Mo Bio Laboratories Inc, USA), modified as follows: 10 minutes incubation at 70 °C after adding the lysis solution (C1), and extending the incubation times to 30 minutes. Once DNA was eluted, the sample was concentrated to dry pellets in a “speed vac” vacuum concentrator, and then washed in 100% ethanol to remove excess salts.

Water samples

DNA was extracted from sterivex filters following a modified version of the PowerWater® Sterivex™ DNA Isolation Kit (Mo Bio Laboratories Inc, USA) (Appleyard et al. 2011). Filters were removed from -80 °C and brought to room temperature before adding 1875 μL lysis buffer (200 mM sodium phosphate buffer, pH 7.0 containing: 1% CTAB, 2% PVP K30, 0.3 M NaCl and lysozyme at a final concentration of 5 mg mL⁻¹) and 125 μL MT Buffer (MP Biomedicals, LLC, USA) via the inlet valve.

Sterivex filters were recapped and attached to a horizontal vortexer (Vortex-Genie 2, Mo Bio Laboratories Inc, USA) and vortexed at speed setting 6 for 1 h. A 3 mL syringe was attached to the inlet of the sterivex and using back pressure from the syringe the contents of the sterivex was removed and split into 2 x 2.0 mL microfuge tubes (approximately 1 mL per tube). 900 µL of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) was added to each tube and mixed by inverting several times. After centrifuging at 13 000 rpm x 10 min, the supernatant was removed and combined in a new 2.0 mL microfuge. Proteinase K was added (20 µL of 20mg/mL stock) followed by incubation at 60 °C for 2 h. 500 µL of Chloroform:Isoamyl Alcohol (CI) (24:1) was added and mixed by inverting several times. After centrifuging at 13 000 rpm x 10 min, the supernatant was removed to a new 2.0 mL microfuge tube and the CI extraction step was repeated. Following centrifugation the supernatant was removed to a 5.0 mL tube and 3 mL of ST4 Buffer (prewarmed to 65 °C) was added. The barrel of a 20 mL syringe was attached to a filter column that was then attached to a vacuum manifold (Vac-Man® Laboratory Vacuum Manifold, Promega Corp, USA). The contents of the 5.0 mL tube was poured into the syringe barrel and pulled through the filter column using the vacuum. Once the entire volume had been pulled through the column the 20 mL syringe barrel was removed. With the vacuum still flowing the filter was washed with 500 µL ST5 followed by 500 µL ST6 and dried by continuing vacuum flow for a further 2 min after the ST6 was entirely pulled through. The vacuum was turned off and the filter transferred to a new 2.0 mL collection tube and allowed to air dry on the bench for 10 min. To elute DNA, the filter was incubated with 80 µL of 0.1 x TE at 37 °C for 45 min, followed by a final spin at 13 000 rpm x 2 min.

The quality and quantity of all DNA was checked using a NanoDrop™ 8000 Spectrophotometer (Thermo Scientific™). DNA was aliquoted into multiple plates, vacuum dried and stored at -20 °C.

5.3.3 PCR and Sequencing

PCR and sequencing of extracted DNA was carried out at the Ramaciotti Center for Genomics (Sydney, Australia). Bacterial and archaeal assemblages were surveyed using small-subunit ribosomal RNA gene amplicon sequencing methods. Amplicons for the V1 to V3 regions of the 16S rRNA gene were prepared using bacterial primers 27F – 519R (Lane 1991, Lane et al. 1985) and archaeal primer A2F – 519R (DeLong 1992, Lane et al. 1985) and sequenced at the same facility using the Illumina MiSeq platform (Illumina, Inc., USA), with 300 bp paired reads, according to the manufacturer's directions. Details of the methodology can be found at (<http://support.illumina.com/sequencing/documentation.html>).

5.3.4 Bioinformatics

Amplicons were analysed in a strictly standardized fashion following the bioinformatics workflow established for the Bioplatforms Australia (BPA) Biome of Australian Soils (BASE) project (Bisset et al. in revision) and adopted by other Australian microbial biodiversity initiatives. This allows the microbial diversity detected in the GAB to be placed within the broader context of microbial diversity in other Australian environments. Full details of the bioinformatics workflow can be found here: (<https://ccgapps.com.au/bpa-metadata/base/information>).

5.3.5 Phylogeny

Preliminary classification was performed using MOTHUR's (Schloss et al 2009) implementation of the Wang classifier (Wang et al 2007) at 60% sequence similarity cut-off with the Green Genes database and taxonomy files (DeSantis et al 2006: <http://greengenes.lbl.gov/cgi-bin/nph-index.cgi>). OTU

sequences that did not classify to the correct lineage (i.e. bacteria or archaea) were discarded from further analysis.

5.3.6 Data Analysis and Statistics

Matrices containing relative abundance data for all OTUs across all samples were generated using the python command `uc2otutab.py` provided on the drive5 Bioinformatics and software services site (<http://www.drive5.com/>) to assist in processing OTU data. Rarefaction curves were generated using the `rarefaction.single` command in MOTHUR (Schloss et al 2009). Preliminary community composition analysis was performed using the Primer-E Multivariate Statistics for Ecologists software package (Version 7; Clarke and Gorley 2015). For community composition analysis, the relative abundance data was fourth root transformed and a Bray-Curtis dissimilarity matrix was constructed after removing taxa that accounted for <0.001% of total reads. Spatial patterns in assemblage composition for bacteria and archaea separately were examined using permutational multivariate analysis of variance (PERMANOVA) and visualised using principal co-ordinates analysis (PCO), and the *Relate* procedure in Primer was used to examine the concordance in patterns between the two groups, and between benthic and pelagic samples. Environmental variables were correlated to community composition data using distance-based linear modelling (DISTLM) with distance-based redundancy analysis (dbRDA), and Pearson correlations with environmental variables were also overlaid on the principal co-ordinates plots. For the DISTLM, the overall best model was selected from all environmental procedures using the BEST procedure and the small sample size variation of AIC (AIC_c) with 9999 permutations. All analyses involving environmental variables were conducted at the station level, as environmental data corresponding to individual cores were not collected.

5.4 Results

5.4.1 Sequencing output and quality assurance

The output from the sequencing runs, as partitioned by dataset and environment, is summarised in Table 5.1. After QC and checks for chimeras, we were able to identify 2,414,482 16S rRNA sequences from Bacteria from benthic samples (Figure 5.2) that were grouped at 97% sequence similarity into 7993 Operational Taxonomic Units (OTUs). From the pelagic samples, 2,306,778 16S rRNA sequences could be grouped into 4522 OTUs (Figure 5.3). For the Archaea, 4,920,756 16S rRNA sequences could be grouped into 496 OTUs from benthic samples (Figure 5.4) and 2,335,864 16S rRNA sequences could be grouped into 290 OTUs from pelagic samples (Figure 5.5). OTUs defined at 97% rRNA sequence similarity are practically equivalent to species level taxonomy and the standard way of looking at 16S rRNA based genomic surveys. Our results indicate that we had sufficient sequencing depth (i.e. number of reads) to discern trends, as shown in Figures 5.2-5.5. In all samples, the curve is either beginning to plateau or has neared plateauing, suggesting that additional read depth would not have substantially increased the number of taxa identified, although there is some undersampling, especially in the bacteria.

Table 5.1. The number of reads and OTUs derived from the sequencing results, divided into data sets and partitioned by environment, where 16S are bacterial sequences, and A16S are archaeal sequences.

<i>Dataset</i>	<i>Post OTU clustering (97%)</i>		<i>Partitioned by environment</i>			<i>Removing OTUs < 0.001%</i>	
	# Reads	# OTUs		# Reads	# OTUs	# Reads	# OTUs
<i>16S</i>	4,771,424	10,236	Benthic	2,414,482	7 993	2,392,189	4,838
			Pelagic	2,306,778	4 522	2,212,238	2,122
<i>A16S</i>	7,256,651	604	Benthic	4,920,756	496	4,918,504	328
			Pelagic	2,335,864	290	2,335,253	180

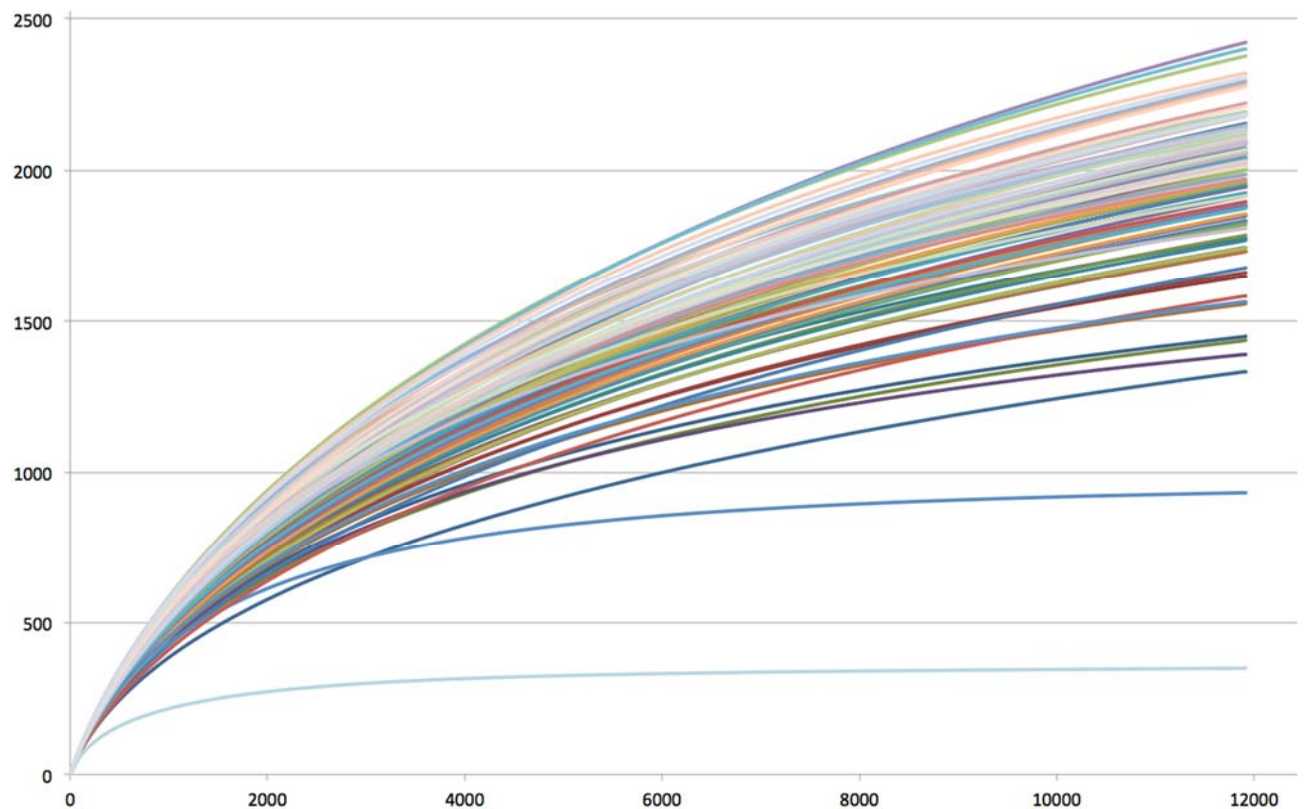


Figure 5.2. Rarefaction curve of bacterial benthic samples, showing the number of new bacterial taxa identified (y-axis) relative to the number of sequences analysed (x-axis) for each station.

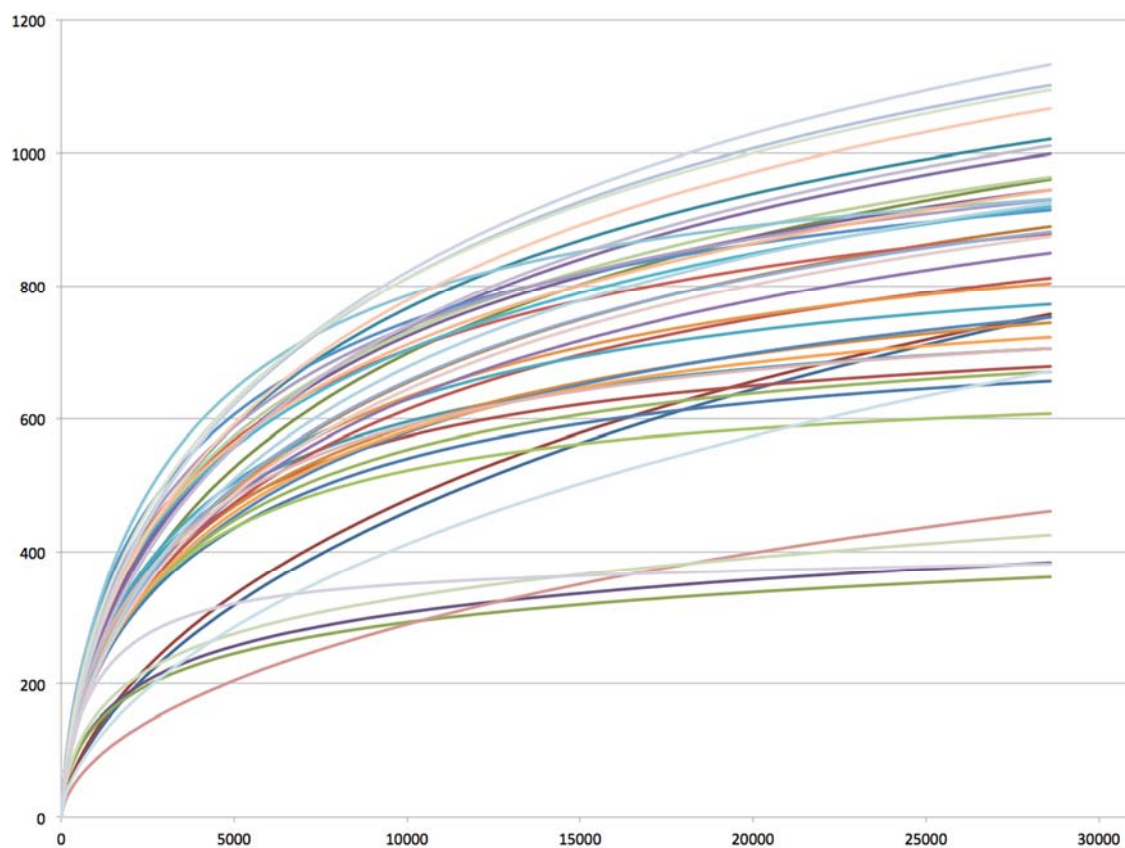


Figure 5.3. Rarefaction curve of bacterial pelagic samples, showing the number of new bacterial taxa identified (y-axis) relative to the number of sequences analysed (x-axis) for each station.

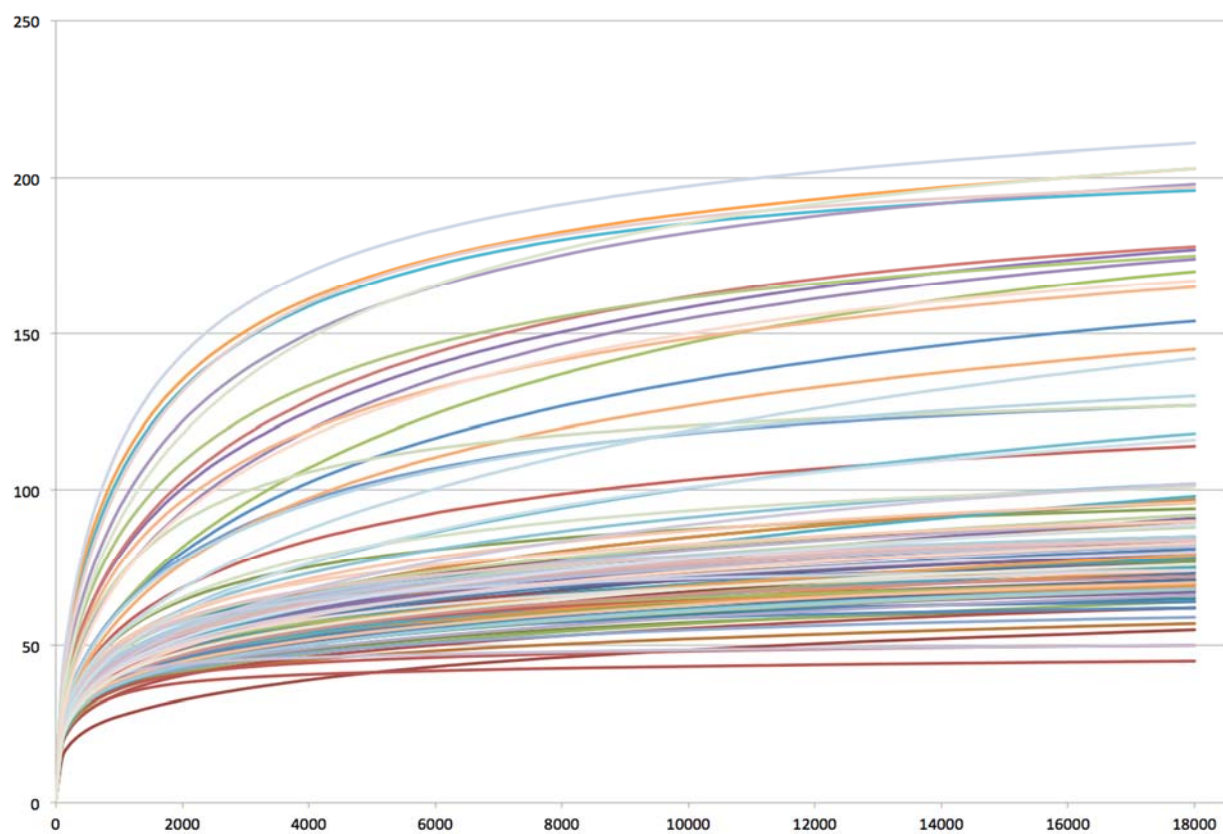


Figure 5.4. Rarefaction curve, showing the number of new archaeal taxa identified in benthic samples (y-axis) relative to the number of sequences analysed (x-axis) for each station.

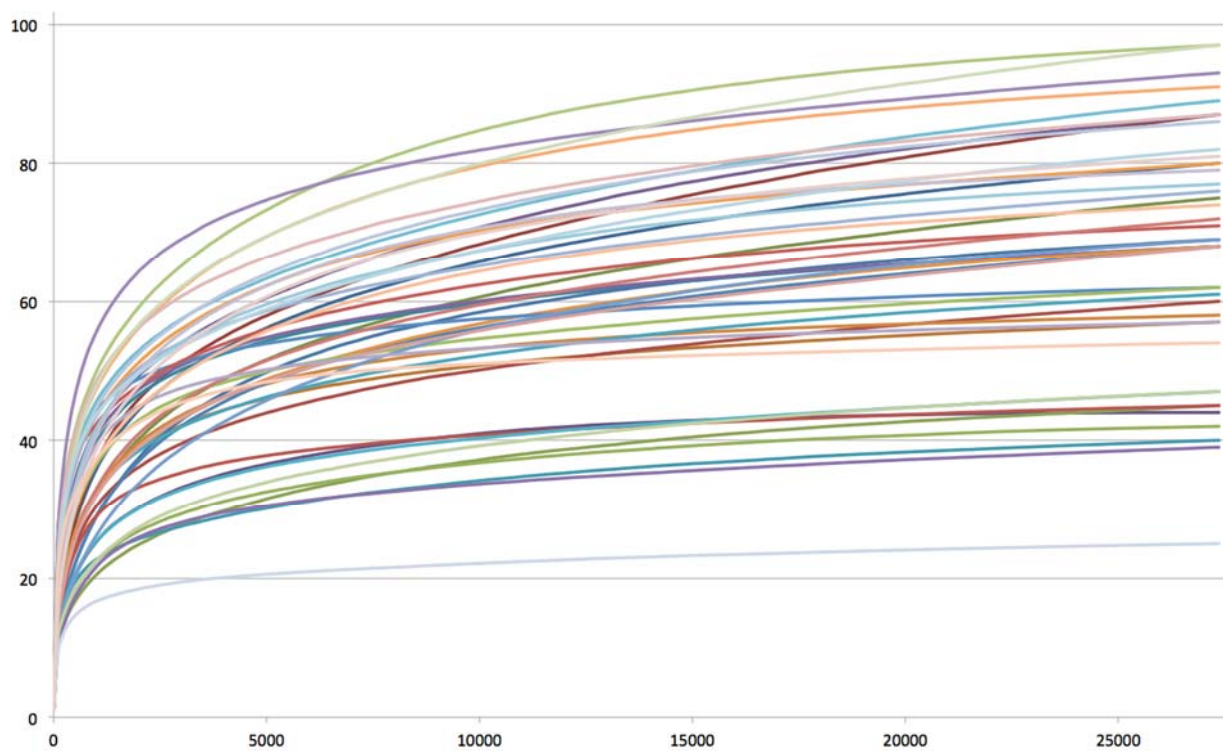


Figure 5.5. Rarefaction curve, showing the number of new archaeal taxa identified in pelagic samples (y-axis) relative to the number of sequences analysed (x-axis) for each station.

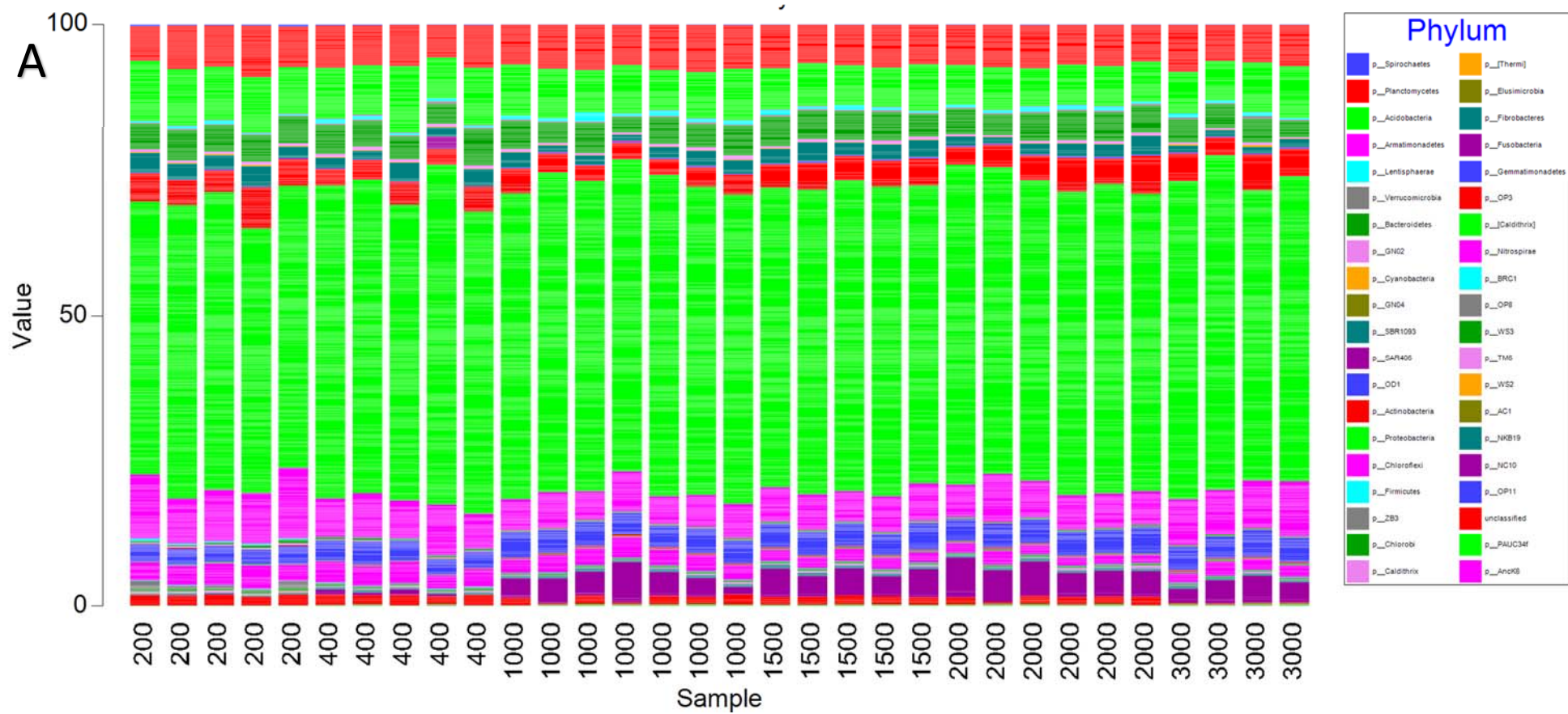
The rarefaction curves in Figure 5.4 and Figure 5.5 demonstrate that we have much lower diversity for Archaea than for bacteria, even though we had equivalent or greater depth of coverage (Table 5.1). In most stations, the diversity begins to plateau at 500 OTUs. By contrast, the bacteria are much more diverse, and typically contain more than 5000 OTUs per station. For all taxa, our sequencing depth greatly exceeds what would be required to describe the diversity of the dominant Archaea present in the regions of the GAB included in the survey, although the rarefaction curves indicate that there will still be a number of rarer unsampled species present.

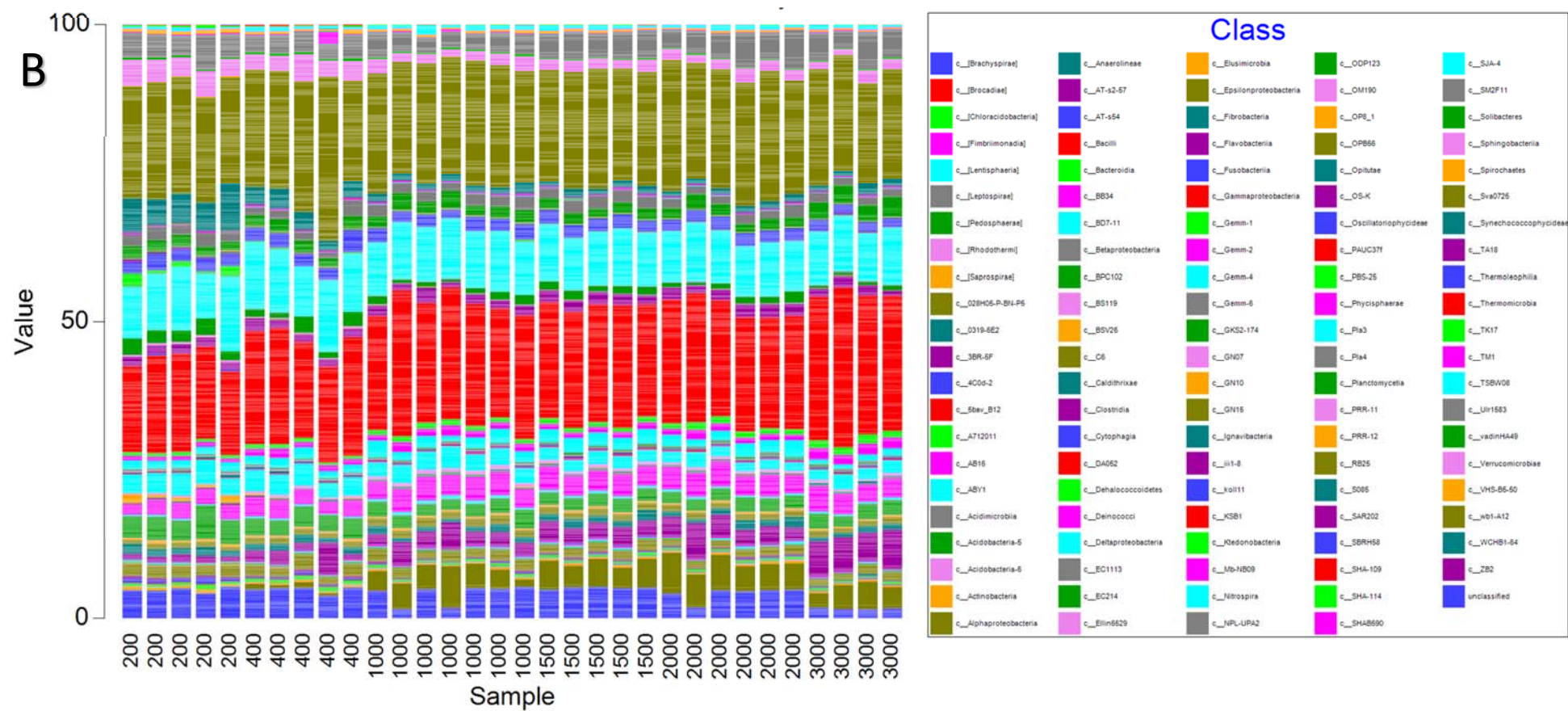
5.4.2 Community composition

Bacteria

As shown in Figure 5.6, bacteria in benthic samples were detected from a wide variety of taxonomic groups. There are few obvious visual differences in the community composition with transect or depth, with samples taken in the same location having very similar composition. When these data are examined at the phylum level (Figure 5.6A), Proteobacteria (bright green) are the most abundant taxon, making up approximately 50% of analysed sequences in all samples. When examined at the class level (Figure 5.6B), the most abundant taxa are the gamma proteobacteria (shown in red), the delta proteobacteria (shown in aqua), and the alpha proteobacteria (shown in khaki). Many of the OTUs could not be classified at the order level (Figure 5.6C, shown in grey), indicating that many of the bacteria found in the GAB are novel and have not been described before. The most abundant orders were: *Rhodospirillales* (in fuschia), *Thiotrichales* (in pink), *Rhizobiales* (in turquoise), *Acidimicrobiales* (in turquoise), the NB1-j clade and the *Nitrospirales* (also shown in pink).

OTUs related to bacteria with a known capacity to degrade hydrocarbons and which were enriched in petroleum containing samples collected during the response to the *DeepWater Horizon* oil spill were identified, as described in Table 5.2. These taxa were not abundant, as would be expected without a measurable hydrocarbon source (Ahmed et al., 2015). Even though these taxa are rare, they were evenly distributed across all samples and consistently present (Figure 5.7). There may be a slight increase in the abundance of families containing known hydrocarbon degrading organisms with depth (Figure 5.7B), but without further functional analysis, these relationships are difficult to interpret on the basis of phylogeny alone.





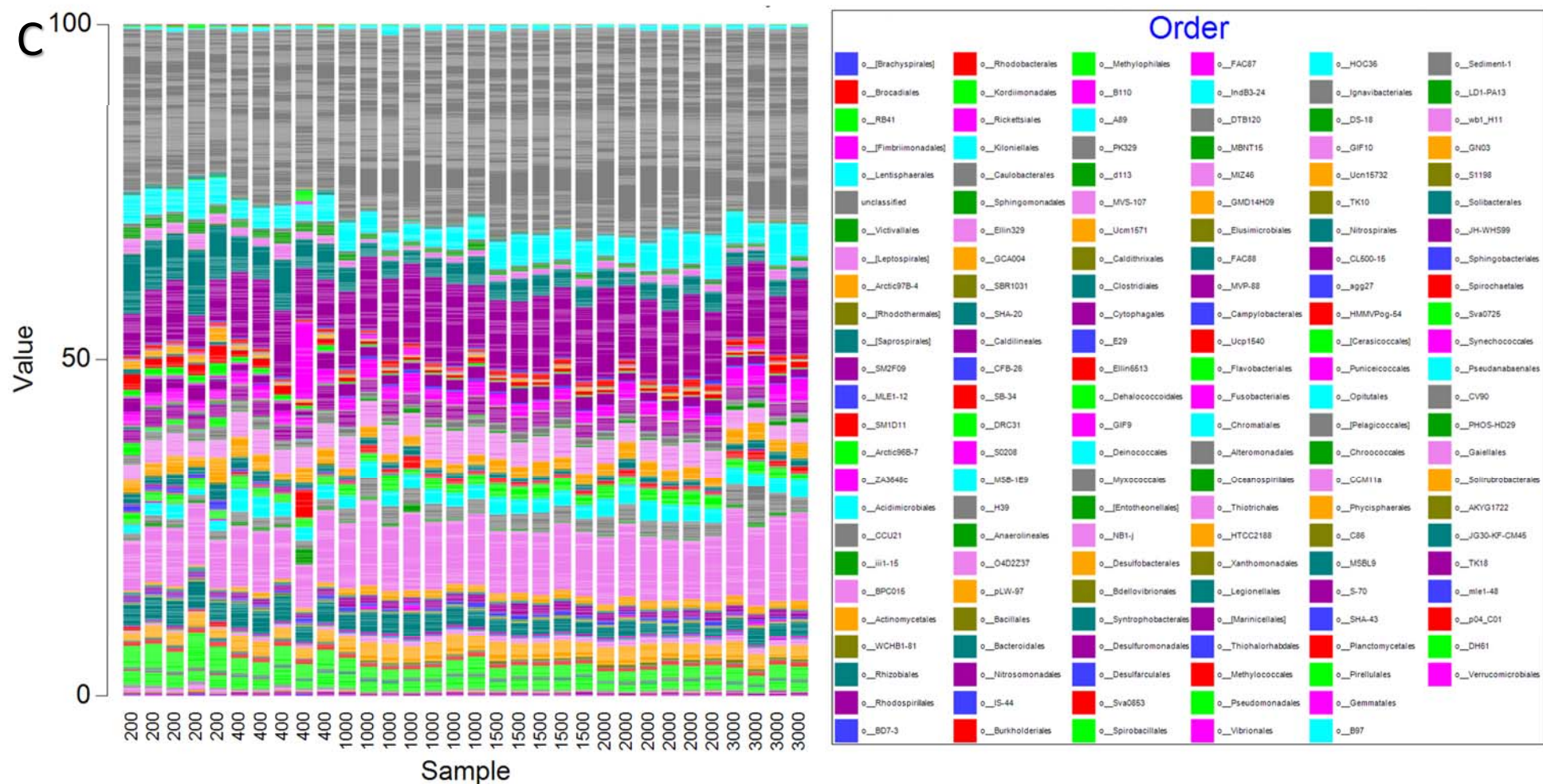


Figure 5.6. The representation of different taxonomic groups of bacteria in benthic samples collected at each station. Depth and transect are on the x axis, and % contribution of each taxonomic group to the overall composition of the sample are on the y axis. Only taxonomic groups that make up more than 0.001% of the community are shown, chloroplast and mitochondrial rRNA have been removed. Each vertical line represents a different sample, but for clarity, only depths are indicated in the x-axis labels. The taxonomic resolution in panel A is to phylum, panel B shows class, panel C shows order.

Table 5.2. A comparison of the taxonomic identification of microbes identified in the Great Australian Bight to those enriched in samples originating from the response to the *DeepWater Horizon* oil spill. Closed symbols denote families, open symbols genera.

Taxa found enriched in oil containing samples from the GOM	Reference	Taxa found in the GAB	Number of OTUs from this taxonomic group found in the GAB	Number of reads aligning to this group	Proportion of reads (of 2.3 M) aligning to this group
Pseudomonadales	1,4	Pseudomonadales (order)	3	96	4.1×10^{-5}
Oceanospirillales	1, 3, 4	Oceanospirillales (order)	91	23328	0.01
		• Oceanospirillaceae (family)	• 26	• 8408	• 0.004
		○ Marinomonas (genus)	○ 1	○ 266	○ 1.1×10^{-4}
		• Oleiphilaceae (family)	• 2	• 105	• 4.6×10^{-5}
		• Halomonadaceae (family)	• 1	• 239	• 1×10^{-4}
		• Alcanivoracaceae (family)	• 9	• 1516	• 6.6×10^{-5}
		○ Alicanivorax (genus)	○ 2	○ 473	○ 2.1×10^{-4}
Methylococcales	1	Methylococcales (order)	2	242	1.1×10^{-4}
Alteromonadales	1,3,4	Alteromonadales (order)	139	45096	0.02
		• Colwelliaceae (family)	• 10	• 1301	• 5.5×10^{-4}
		• Moritellaceae (family)	• 6	• 1778	• 7.7×10^{-4}
		• Shewanellaceae (family)	• 4	• 323	• 1.4×10^{-4}
		• Alteromonadaceae (family)	• 67	• 36158	• 0.015
		○ Marinobacter (genus)	○ 4	○ 3821	○ 1.6×10^{-3}
Flavobacteriales	1	Flavobacteriales (order)	152	38994	0.015
Sphingobacteriales	1	Sphingobacteriales (order)	79	11084	0.005
		• Saprospiraceae (family)	• 43	• 3157	• 0.0013
Desulfuromonadales	2	Desulfuromonadales (order)	20	3725	0.0016
Desulfobacterales	2	Desulfobacterales (order)	79	30977	0.013

1. Dubinsky et al., 2013, enriched in the plume; anomaly; 2. Kimes et al., 2013; 3. Chakraborty et al., 2012; 4. Hazen et al., 2010

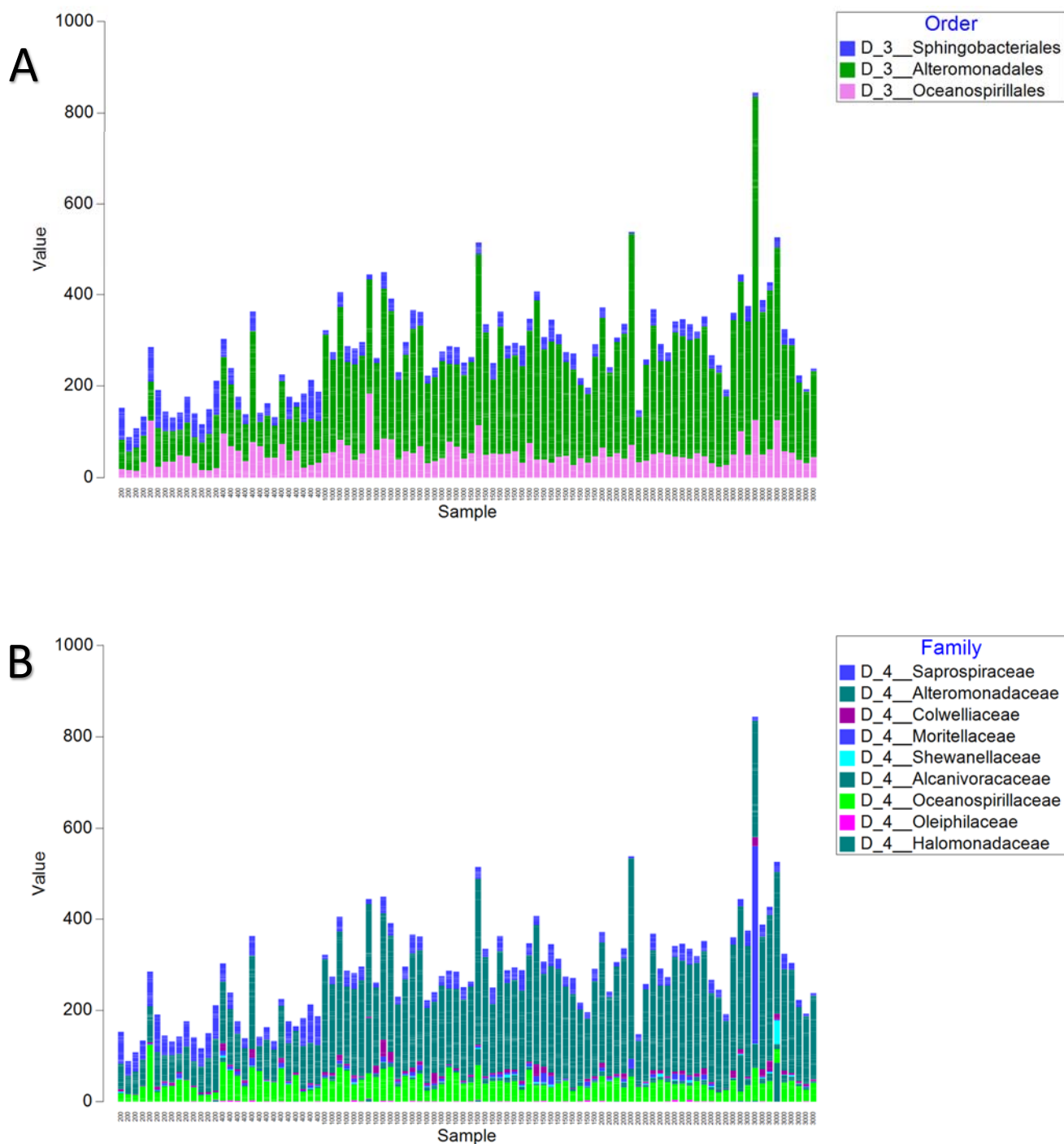
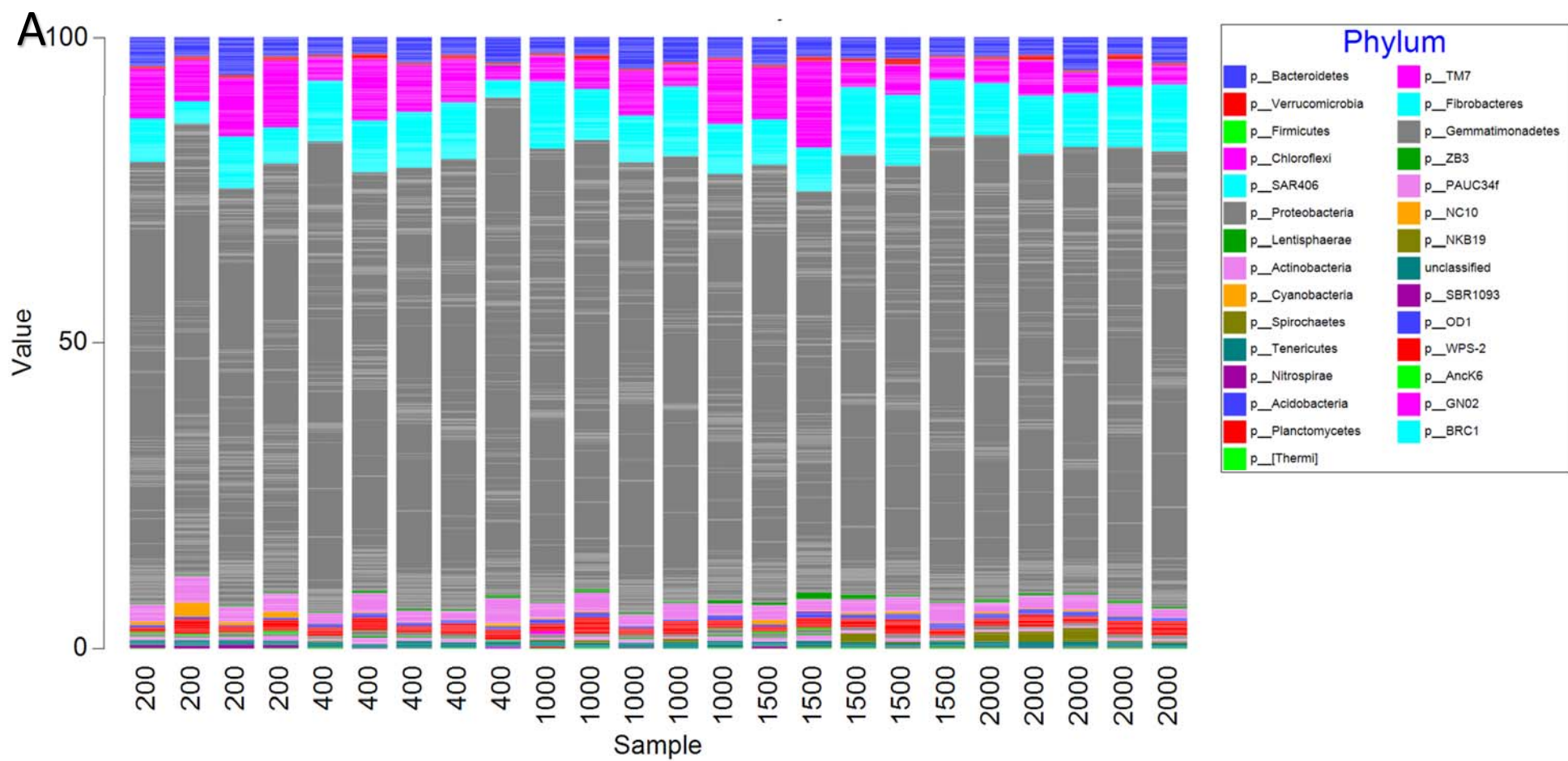
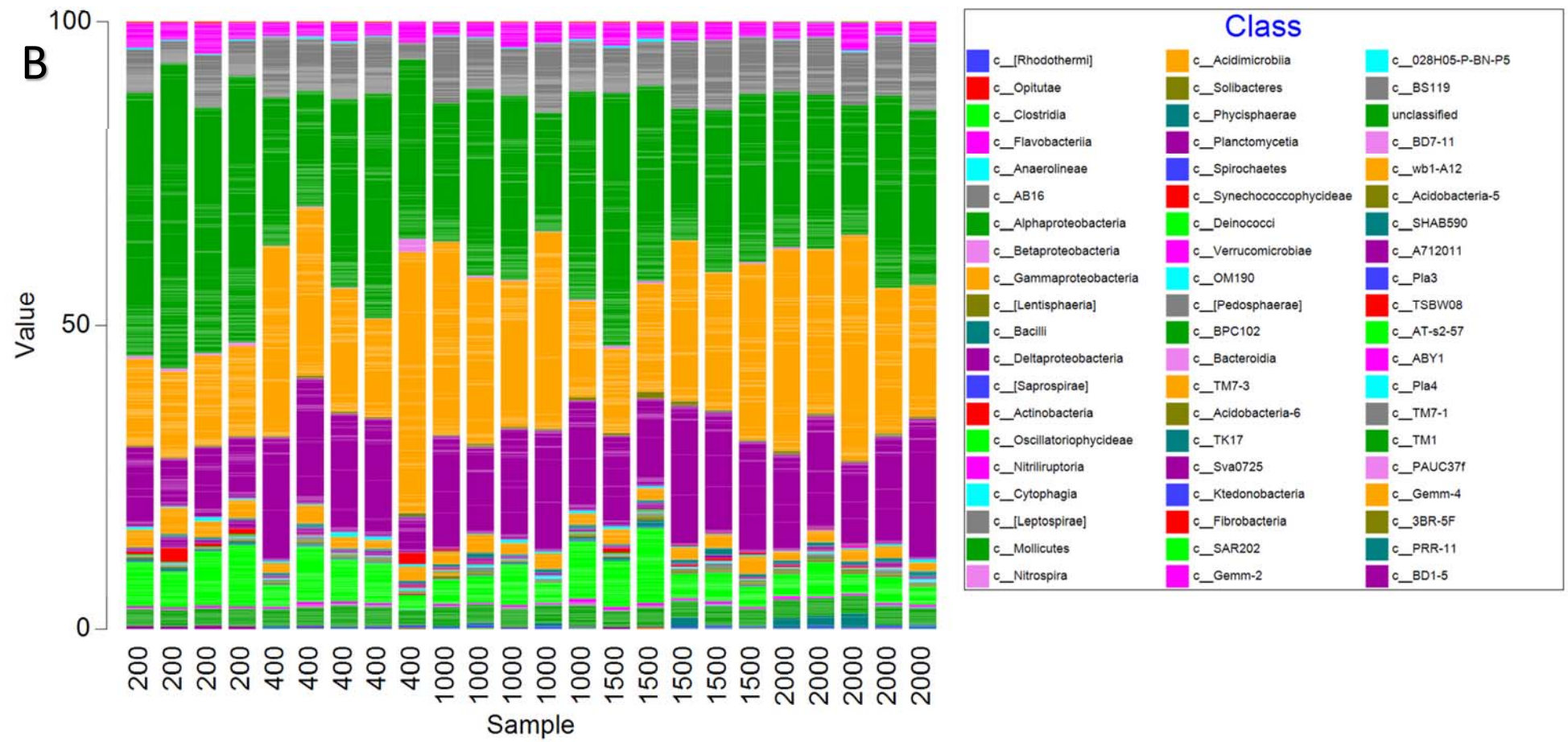


Figure 5.7. The relative abundance of bacteria in orders (panel A) or families (panel B) known to contain hydrocarbon degrading organisms. The station (in order of transect) and depth are shown on the x axis, and the abundance on the y axis. For clarity, only depth is labelled.

When bacteria found in pelagic samples are examined, a somewhat different taxonomic distribution is apparent (Figure 5.8). Like the benthic samples, few differences in composition are apparent with either transect or depth. In both the benthic and pelagic samples, the dominant phylum is the proteobacteria (Figure 5.8A). Again, at the level of class (Figure 5.8B), the alpha proteobacteria (shown in bright green), the gamma proteobacteria (shown in orange) and the delta proteobacteria (shown in purple), make up the most abundant taxa. At the taxonomic level of order (Figure 5.8C), about 15% of the OTUs (shown in red) are unclassified, again indicating that many of the bacteria found in the GAB are novel and have not been described before. The *Rickettsiales* (shown in lilac), the *Altermonadales* (shown in fuchsia) and the *Acidomicrobiales* (shown in aqua) are among the most abundant orders in the pelagic samples.





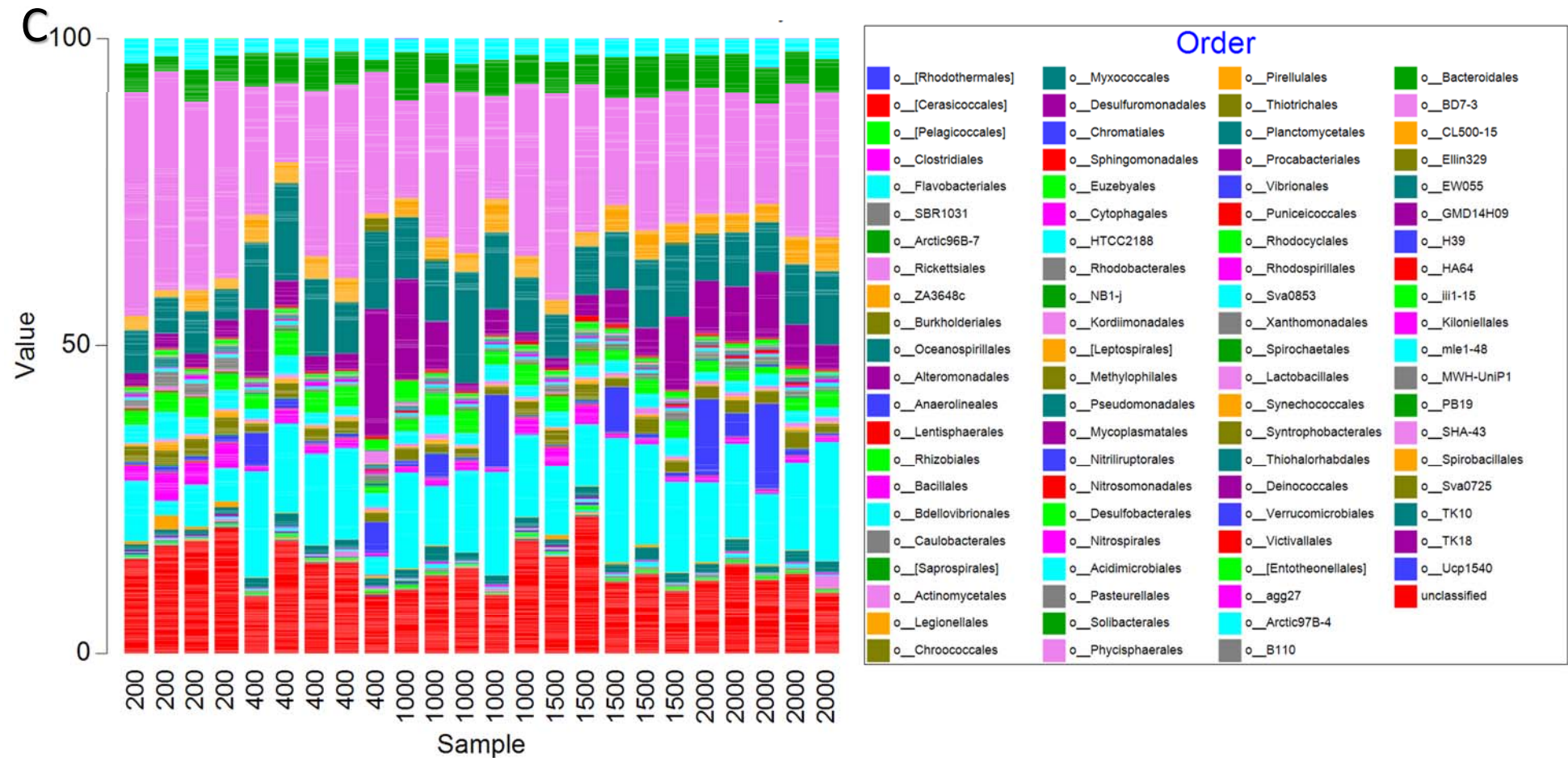
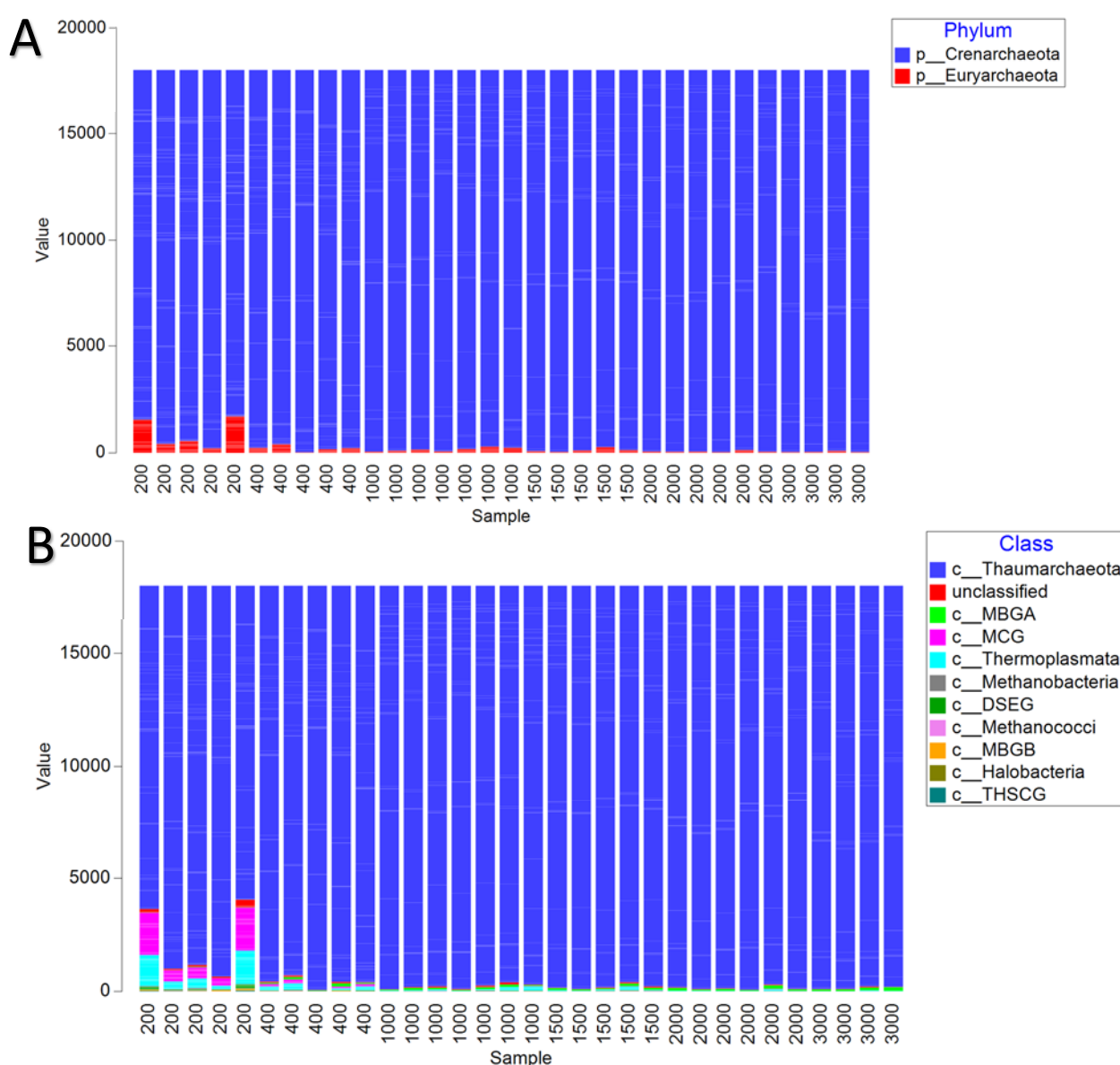


Figure 5.8. The representation of different taxonomic groups of bacteria in pelagic samples collected at each station. Depth and transect are on the x axis, and % contribution of each taxonomic group to the overall composition of the sample is on the y axis. Only taxonomic groups that make up more than 0.001% of the community are shown, chloroplast and mitochondrial rRNA have been removed. Each vertical line represents a different sample, but for clarity, only depths are labelled. The taxonomic resolution in panel A is to phylum, panel B shows class, panel C shows order.

5.4.3 Archaea

The comparative abundances of different taxonomic groups of Archaea in benthic samples are shown in Figure 5.9. Both Archaeal phyla are present at all depths, however, the Euryarchaeota are more abundant at the 200 m depths than in the remaining samples (Figure 5.9A). This increased diversity at the shallower depths is apparent at the class and family level as well (Figure 5.9B, C), although in all samples, a single genus of ammonia oxidising archaea, the genus *Nitrosopumilus* (phylum Crenarchaeota, class Thaumarchaeota, order Cenarchaeales, all three shown in blue) was most abundant. This genus contributed over two thirds of the total observed sequences across the whole dataset.



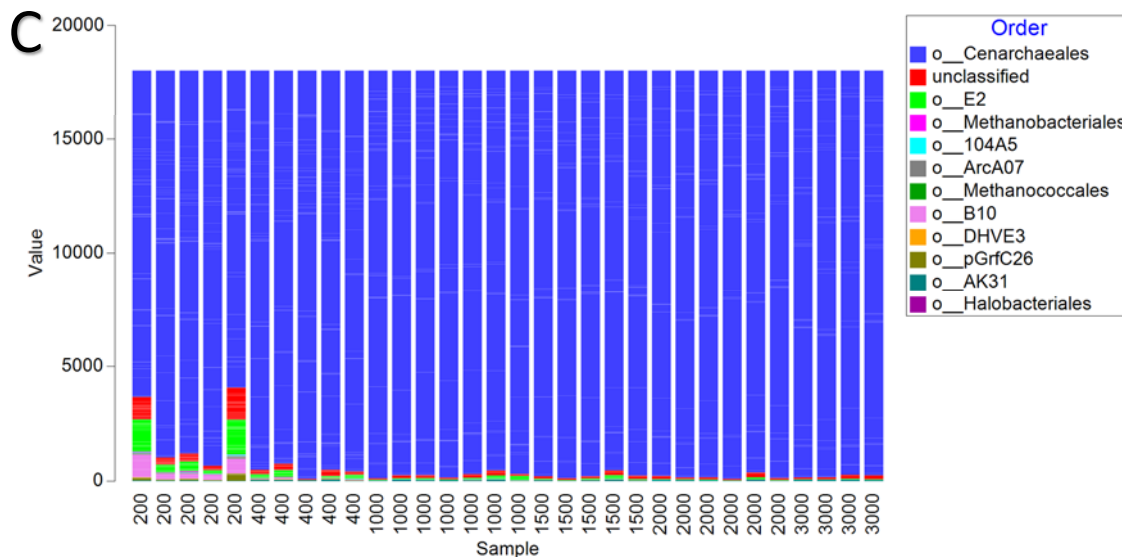


Figure 5.9. Comparative abundances of different groups of *Archaea* collected from benthic samples at the different depths. Depth and transect are on the x axis, and % contribution of each taxonomic group to the overall composition of the sample are on the y axis. Only taxonomic groups that make up more than 0.001% of the community are shown. For clarity, station numbers are not shown. Panel A is taxonomic resolution to phylum, panel B shows class, panel C shows order.

The comparative abundances of different orders of *Archaea* from pelagic samples are shown in Figure 5.10. There was greater variability between different samples in pelagic *Archaeal* composition relative to the benthic or bacterial samples, with the shallow water depths from Transect 5 being particularly different from the rest (i.e. Figure 5.10A). In most of these samples, the distribution of phyla is more even between the *Euryarchaeota* (shown in red) and the *Crenarchaeota* (shown in blue) (Figure 5.10A). There is a greater proportion of members of the class *Thermoplasmata* (shown in green) in pelagic samples relative to benthic, although the *Thaumarchaeota* (shown in blue) are more than 50% of many of the pelagic samples (Figure 5.10B). Of the sequences that can be classified to order, the *Cenarchaeles* are most abundant (Figure 5.10C). Again, similar to the benthic samples, over two thirds of the archaea detected in the pelagic samples belong to the genus *Nitrosopumilus*.

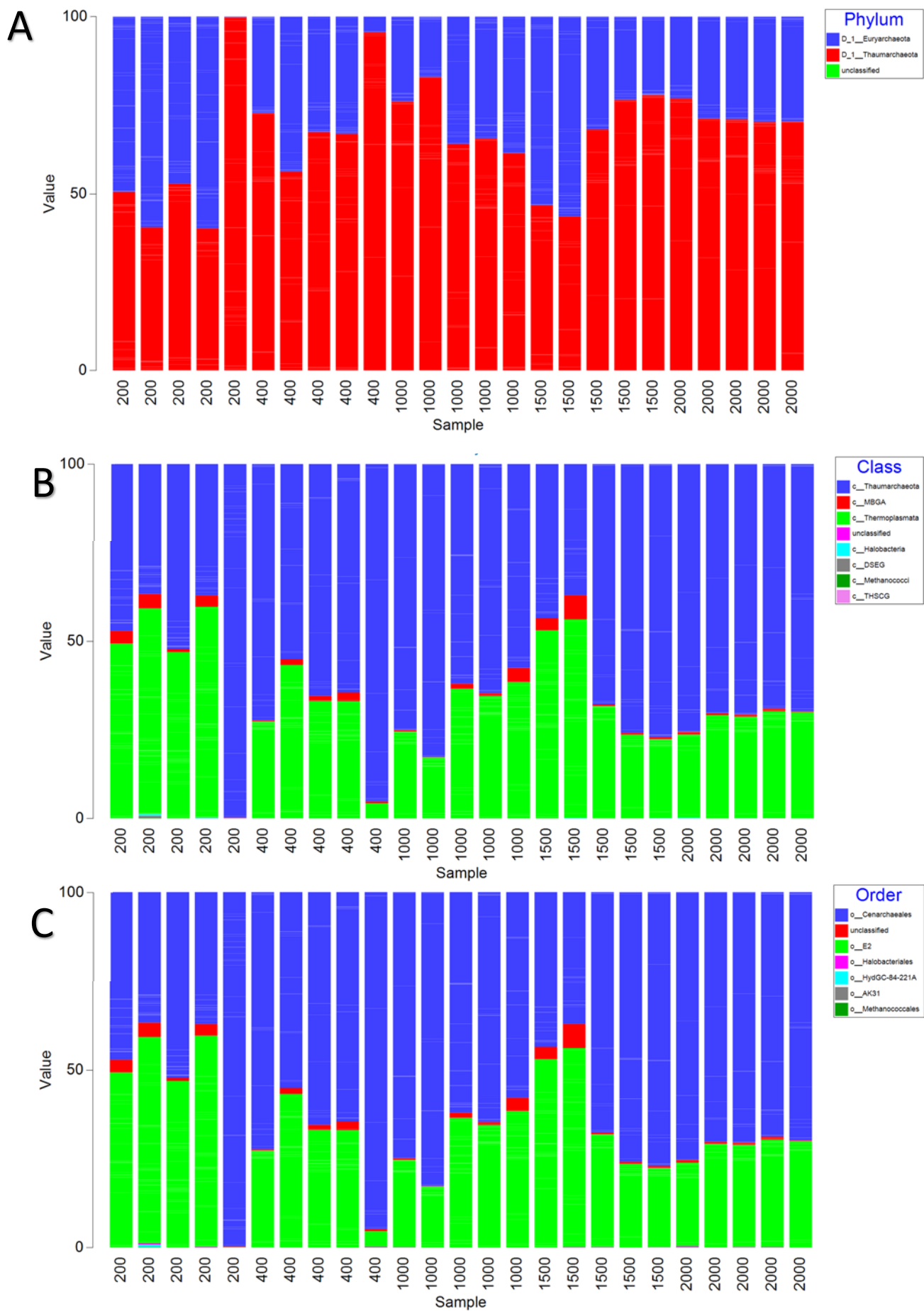


Figure 5.10. (Prev page) Comparative abundances of different groups of *Archaea* collected from pelagic samples at the different depths. Depth and transect are on the x axis, and % contribution of each taxonomic group to the overall composition of the sample are on the y axis. Only taxonomic groups that make up more than 0.001% of the community are shown. For clarity, station numbers are not shown. Panel A is taxonomic resolution to phylum, panel B shows class, panel C shows order.

5.4.4 Diversity Indices

Trends in the biodiversity of microbial communities were explored further using Shannon and Simpson indexes of biodiversity, and by exploring the dominance and evenness of the distribution of different OTUs. As shown in Figure 5.11, there is a general trend towards less bacterial diversity in the sediment with depth (Figure 5.11A). This is reflected in a greater degree of dominance of some OTUs in the samples collected at deeper stations (Figure 5.11B). In bacterial communities from pelagic samples, however, there is more variability in both diversity (Figure 5.12A) and evenness (Figure 5.12B), obscuring any trends in ecological parameters with depth.

For the Archaeal communities, the decrease in diversity and species evenness with depth is even more pronounced, as depicted in Figure 5.13. The differences between the 200 m stations and the deeper communities are most noticeable. However, as shown in Figure 5.14, variability between different stations obscures any differences in either diversity or dominance with depth in pelagic samples.

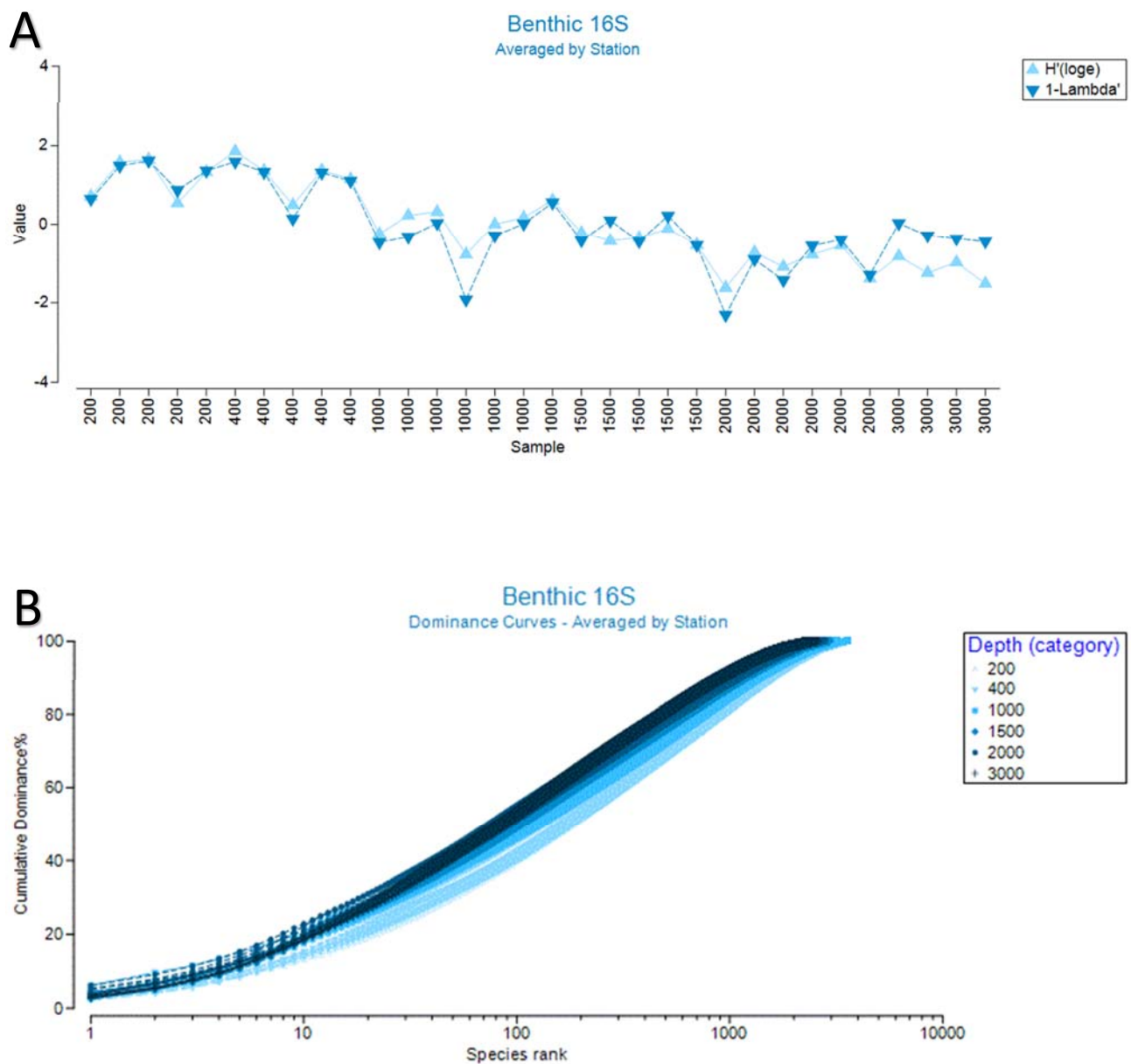


Figure 5.11. Biodiversity metrics in the benthic bacterial samples. Panel A shows diversity as determined by either the Shannon (H) or Simpson (1-lambda) metrics; Panel B shows evenness for each station.

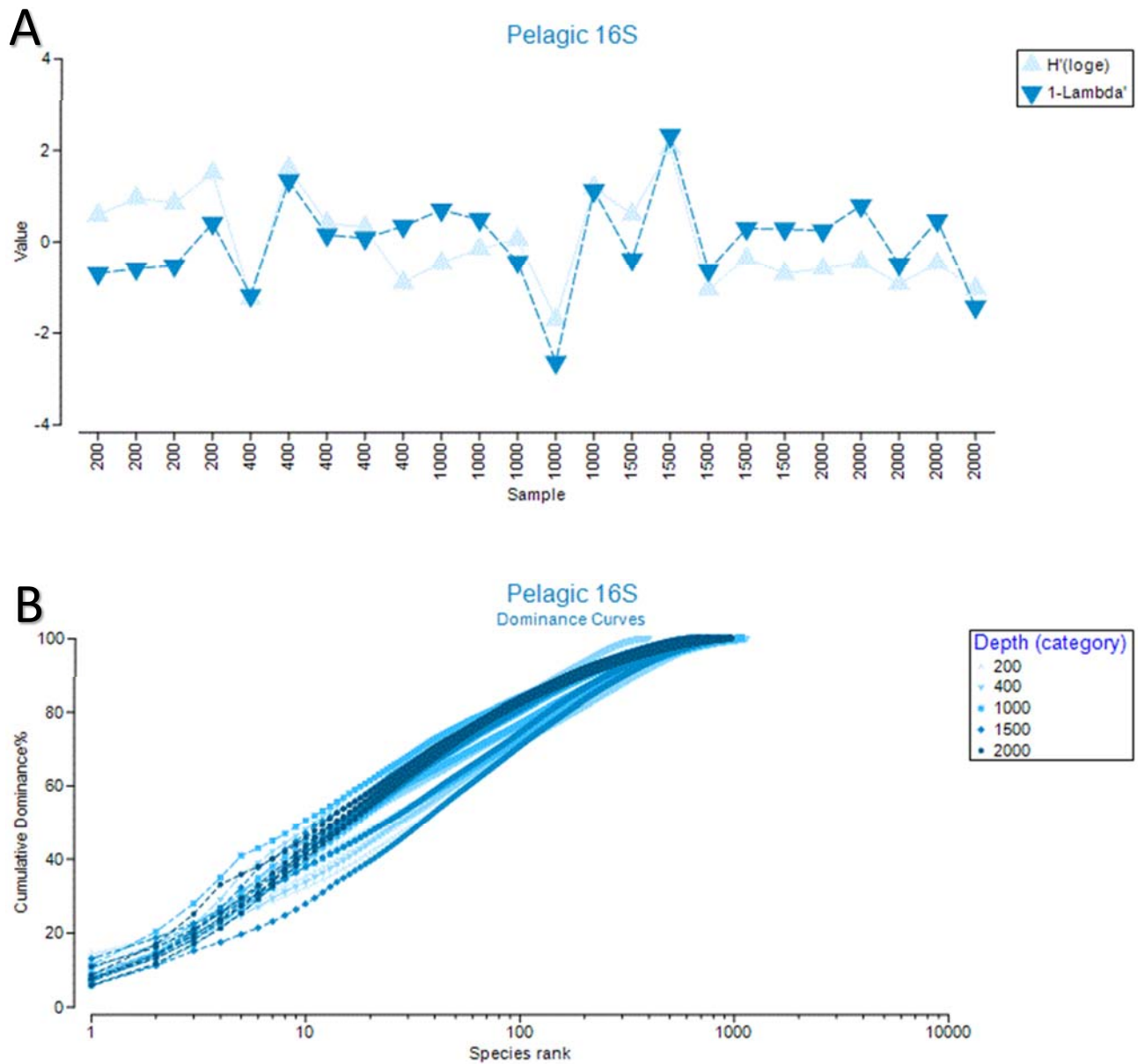


Figure 5.12. Biodiversity metrics in the pelagic bacterial samples. Panel A shows diversity as determined by either the Shannon (H) or Simpson ($1-\lambda$) metrics; Panel B shows evenness for each station.

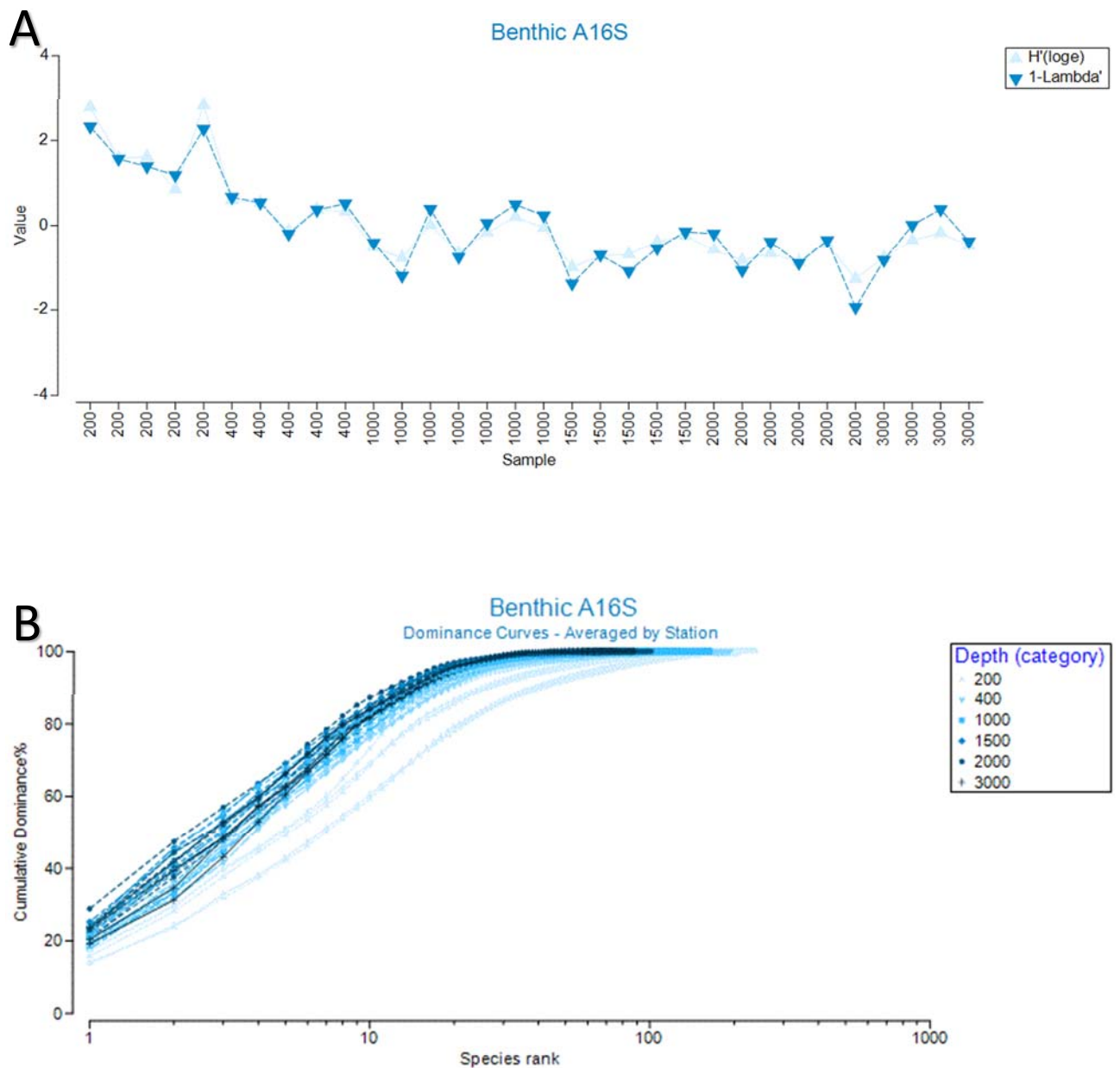


Figure 5.13. Biodiversity metrics in the benthic archaeal samples. Panel A shows diversity as determined by either the Shannon (H) or Simpson (1-lambda) metrics; Panel B shows evenness for each station.

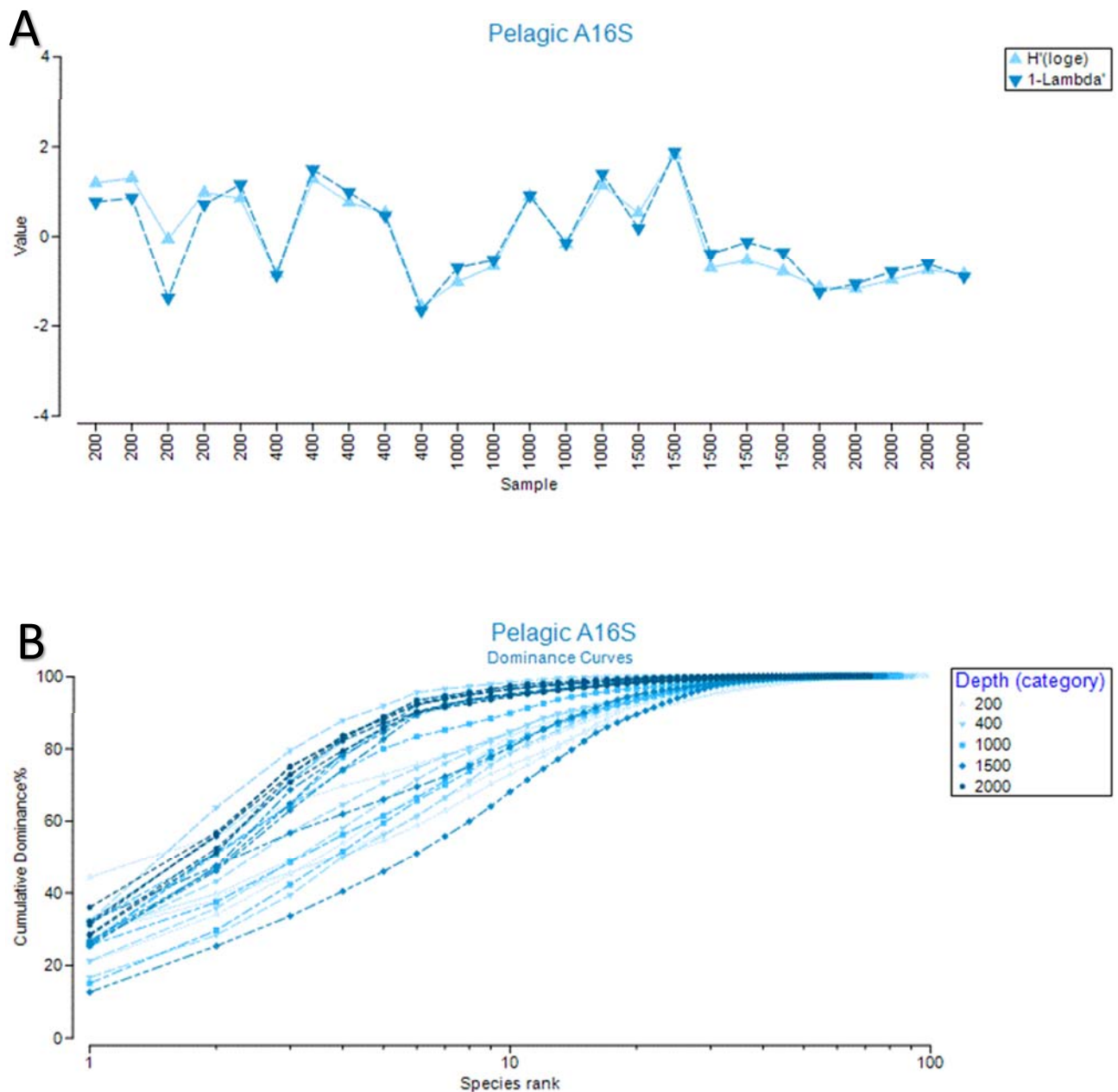


Figure 5.14. Biodiversity metrics in the pelagic archaeal samples. Panel A shows diversity as determined by either the Shannon (H) or Simpson (1-lambda) metrics; Panel B shows evenness for each station.

5.4.5 Relationships between community composition and environmental variables

The relationships between the taxonomic compositions of different samples were visualised using principle co-ordinates analysis. As shown in Figure 5.15, the benthic bacterial samples cluster tightly with depth, with little variation either between different replicates in the same transect or different transects, although there was an interaction between depth and transect (PERMANOVA: $F_{19,67}=1.36$, $p<0.001$). Pairwise tests indicated that Transects 1 and 2 both differed to Transect 5 at 1000 m only. Samples from 200 m were consistently different to all other depths, except on Transects 3 and 4, where they were the same as the 400m samples. Similarly, 400 m samples differed to all deeper

samples except for the 1000 m samples on Transect 3. Samples from 1000 m consistently differed from those from 3000 m, and also the 2000 m samples on Transects 1 and 2, and 1500 m samples differed from 3000 m on Transects 1 and 4. When only those taxonomic families known to contain hydrocarbon degrading taxa are analysed (Figure 5.16), the same trends are apparent both spatially and with regards to the clustering of depths. This suggests that the differences in abundance of the taxa related to known hydrocarbon degraders are determined by the same environmental factors influencing the overall bacterial community. Pelagic bacterial assemblages also separated out by depth (PERMANOVA: $F_{4,15}=2.99$, $p=0.011$), but not transect (PERMANOVA: $F_{4,15}=0.60$, $p=0.87$), although there was higher variation between samples at each depth (Figure 5.17). Pairwise tests indicated that 200 m samples differed to all other depths except for 1500 m, but all other depth pairs were equal. These differences could reflect that the deep waters are a more variable environment than the deep sediment. Although there was a significant relationship between the benthic and pelagic bacterial assemblages at the station level, the correlation was only weak (RELATE: $\rho=0.352$, $p=0.009$).

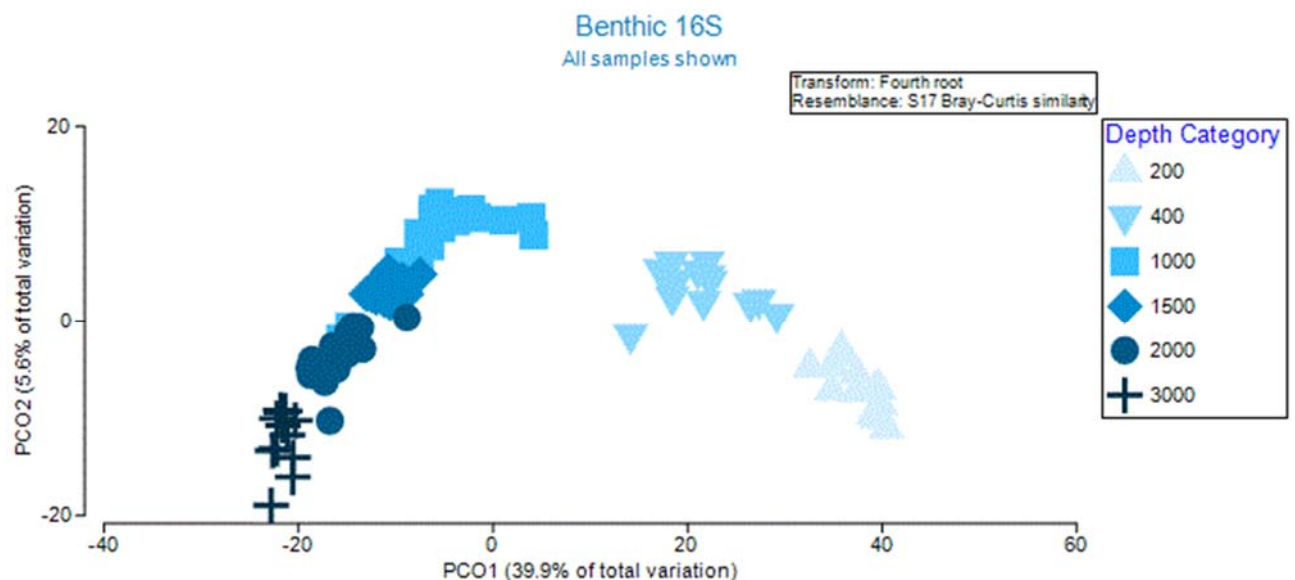


Figure 5.15. Relationships between bacterial benthic samples as visualised via principle co-ordinates analysis.

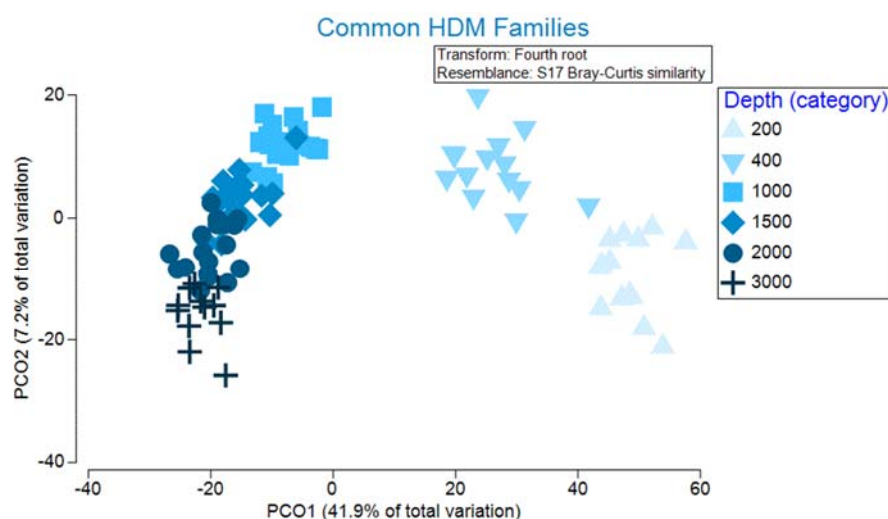


Figure 5.16. Relationships between families containing known hydrocarbon degrading taxa in benthic samples as visualised via principle co-ordinates analysis.

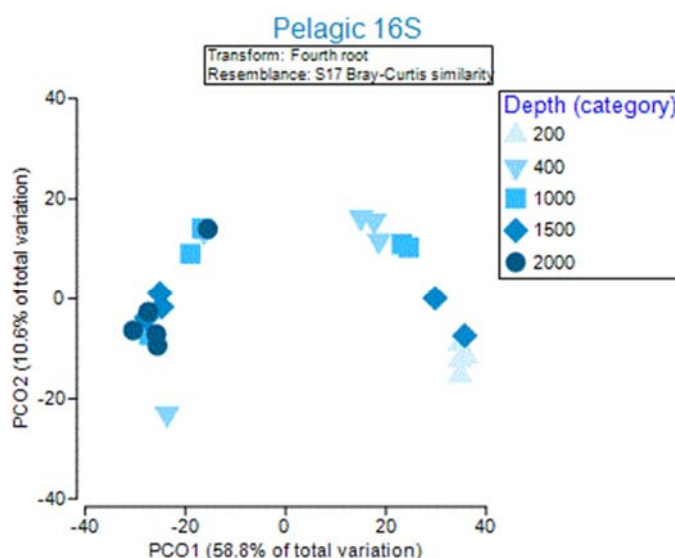


Figure 5.17. Relationships between the bacterial composition of different pelagic samples as visualised via principle co-ordinates analysis.

The relationships amongst the Archaeal communities in benthic samples are shown in Figure 5.18. Although the depth clusters are not as tight as observed in the benthic bacterial samples, close relationships amongst the same depth and among shelf (200-400 m) or slope (>1000 m) samples are apparent. There is less spatial resolution between the samples collected from depths greater than 1000 m, as shown in Figure 5.18, but this may be a result of fewer OTUs being present in the sample. Again, there is an interaction between depth and transect (PERMANOVA: $F_{19,70}=2.20$, $p<0.001$). Pairwise tests indicated that at shallower depths (200-1000 m) Transect 5 is generally different to the other transects, at 1500 m, Transect 1 is different, and at 2000 and 3000 m there are no differences between transects. The composition of the archaeal assemblage is highly correlated with that of the bacterial assemblage (RELATE: $\rho=0.895$, $p<0.001$).

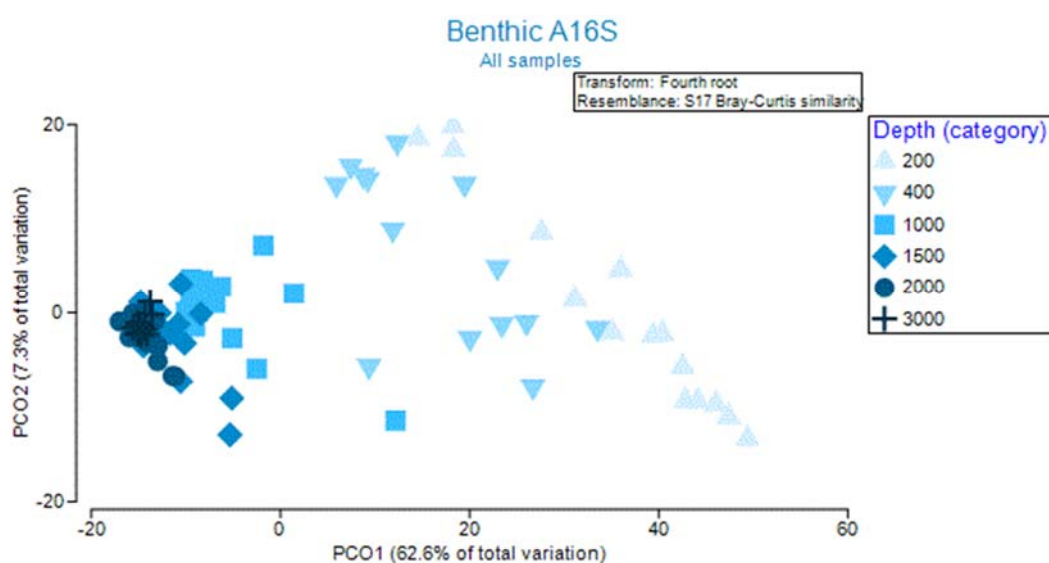


Figure 5.18. Relationships between the archaeal composition of different benthic samples as visualised via principle co-ordinates analysis.

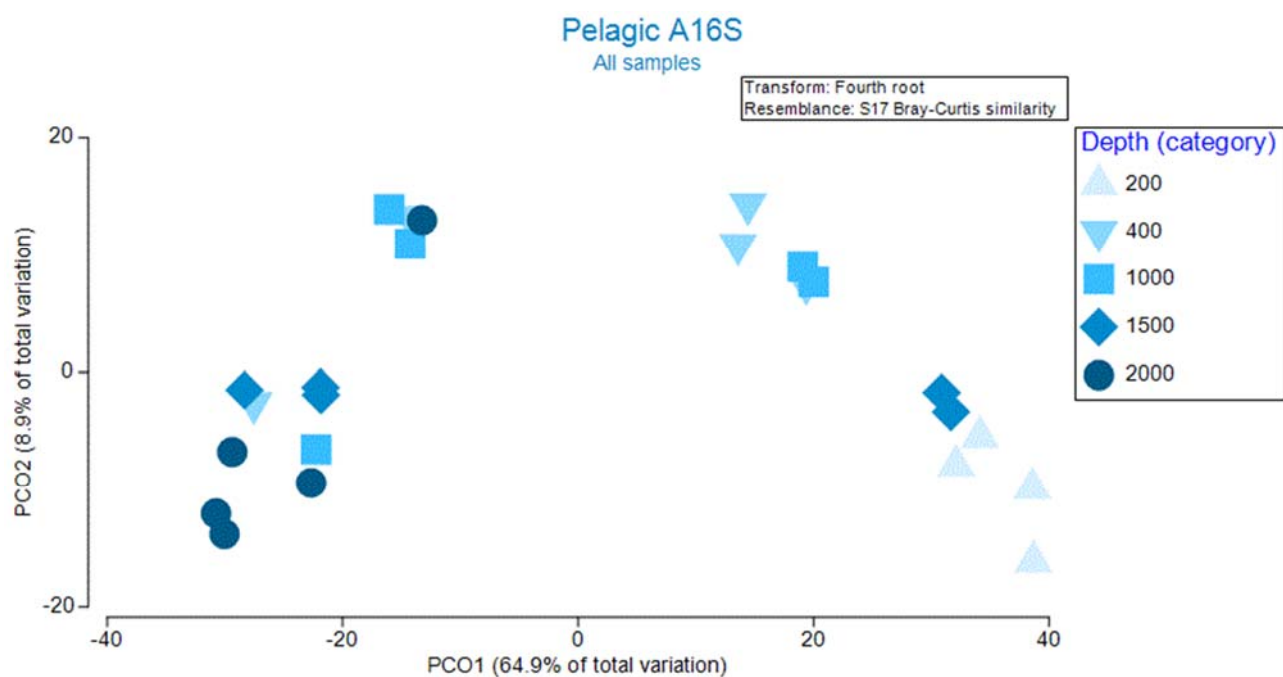


Figure 5.19. Relationships between the archaeal composition of different pelagic samples as visualised via principle co-ordinates analysis.

The relationships amongst the Archaeal communities in pelagic samples are shown in Figure 5.19. Again, assemblages separated out by depth (PERMANOVA: $F_{4,16}=3.44$, $p=0.005$), but not transect (PERMANOVA: $F_{4,16}=0.673$, $p=0.70$), and there was higher variation between samples at each depth than for benthic assemblages, suggesting that there may be greater environmental variability in

oceanic waters than in sediment. The composition of the archaeal assemblage is very highly correlated with that of the bacterial assemblage (RELATE: $\rho=0.965$, $p<0.001$). Again, while the relationship between benthic and pelagic assemblages was significant, it was only weak (RELATE: $\rho=0.419$, $p=0.005$).

The relationship between community composition and environmental variables was determined via distance-based linear modelling. For benthic bacteria, the final selected model ($r^2=0.60$) only contained temperature and depth, with none of the sediment variables being important (Table 5.3). For benthic archaea, the best model also included coarse sand ($r^2=0.73$). Much less of the variation in pelagic microbial assemblages was explained by the measured environmental variables, with the best model for bacteria containing temperature and oxygen ($r^2=0.31$), and for archaea temperature only ($r^2=0.21$). Transect never had a significant relationship with community composition ($F < 1$; $P > 0.312$).

Table 5.3. Relationships between community composition and environmental parameters as determined by distance-based linear modelling. Only variables contributing to the final selected model are included. Values for Pseudo-F and P represent values for marginal tests for each variable individually.

Variable	Pseudo-F	P value
16S Benthic		
Temperature	35.27	< 0.001
Depth	23.22	< 0.001
A16S Benthic		
Temperature	58.52	< 0.001
Depth	27.42	< 0.001
Coarse sand (0.5-1 mm)	21.783	< 0.001
16S Pelagic		
Temperature	6.46	0.003
Oxygen	4.16	0.015
A16S Pelagic		
Temperature	6.24	0.002

To further explore these relationships, they were visualised with both principal co-ordinates analysis and distance based redundancy analysis. Both approaches showed similar relationships between environmental variable towards and community composition. For instance, for benthic bacterial samples (Figure 5.20), both analyses indicated that depth was the major environmental variable separating stations from 1000-3000 m, while temperature and sediment grain size composition played

a greater role in separating out 200-1000 m stations. For reference, the sediment grain size distributions are provided in Table 5.4. There are slight differences in the influence of clay between the two analyses. There are also strong relationships between the pelagic bacteria assemblage and environmental variables (Figure 5.21), particularly nutrient concentrations and depth. The environmental factors that are important are similar in both analyses, but the relationships between them are different depending on the analysis type.

Similar relationships are seen between archaeal community composition and environmental variables. For instance, as shown in Figure 5.22, depth, temperature, and sediment grain size are the primary determinants of archaeal community composition, although the relationships between these variables differ depending on the analysis tool used. For pelagic samples, depth, nutrient levels and temperature are primarily influencing differences in community composition, as shown in Figure 5.23. Again, while both types of analysis identify the same environmental variables as being important influencers of Archaeal biodiversity, the relationships between these parameters differs depending on the analysis type used.

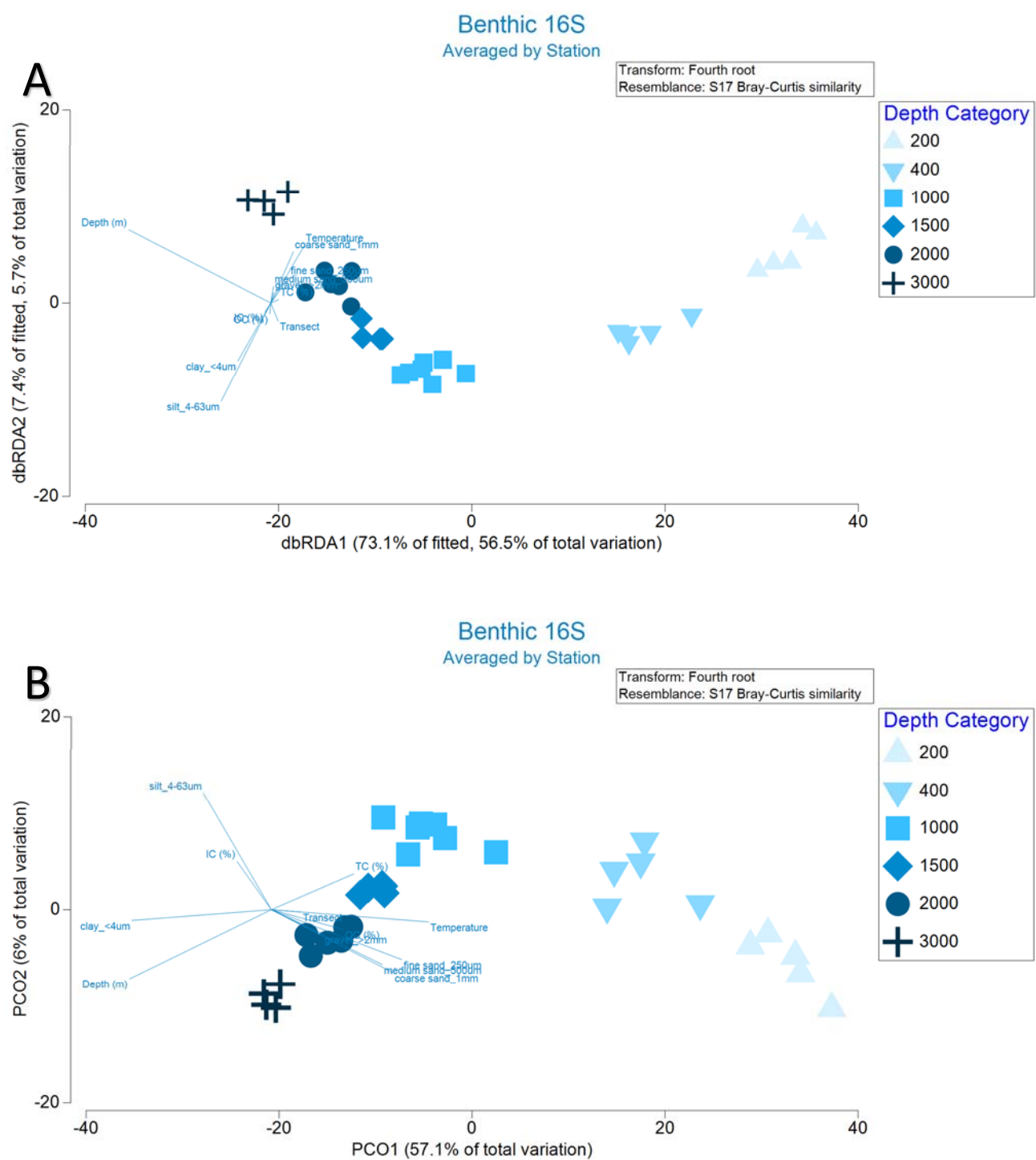


Figure 5.20. Relationships between community composition in benthic bacterial samples and environmental variables as determined via distance-based redundancy analysis (panel A) or principle co-ordinates analysis (panel B).

Table 5.4. Sediment grain size distribution (%) with depth for each of the sampling transects. Values are percent composition by volume up to 1 mm, and percent composition by weight for fractions > 1 mm.

Transect	Depth (m)	Clay <4 μ m	Silt 4-63 μ m	Very fine sand 63-125 μ m	Fine sand 125-250 μ m	Medium sand 250-500 μ m	Coarse sand 0.5-1mm	Very coarse sand 1-2mm	Gravel >2mm
1	200	13.86	0.16	25.42	22.43	24.47	17.02	10.51	0.74
2	200	14.5	0.16	25.84	19.41	20.35	16.67	17.56	2.48
3	200	15.22	0.13	23.83	18.84	23.19	20.26	13.76	4.52
4	200	14.55	0.15	16.64	14.18	23.11	23.42	22.49	9.72
5	200	13.23	0.15	21.88	16.16	16.02	16	29.8	17.59
1	400	10.5	19.27	41.94	21.22	13.52	3.56	0.49	0.21
2	400	10.84	0.75	56.26	24.51	12.98	4.26	1.25	3.5
3	400	11.48	15.13	43.39	23.34	13.6	4.05	0.49	0.74
4	400	9.8	0.64	44.59	23.05	19.08	8.8	3.85	0.38
5	400	11.08	0.61	55.03	23.96	11.98	5.47	2.94	0.51
1	1000	4.31	30.02	53.14	9.02	5.01	2.68	0.14	3.49
2	1000	4.36	27.32	56.37	7	5.29	3.74	0.29	1.2
3	1000	4.07	32.08	44.48	7.78	8.59	6.63	0.44	12.39
4	1000	4.13	36.82	52.72	6.07	2.97	1.4	0.02	6.93
5	1000	4.83	25.54	47.05	16.6	8.24	2.09	0.48	0.21
1	1500	2.71	37.53	47.11	5.55	5.29	4.12	0.4	2.96
2	1500	2.69	2.4	61.69	10.7	12.19	11.49	1.52	1.99
3	1500	2.73	33.24	45.06	8.17	7.61	5.35	0.57	1.34
4	1500	2.77	40.77	40.64	5.55	5.55	6.23	1.26	7
5	1500	2.74	31.6	47.13	7.99	6.52	6.08	0.67	6.55
1	2000	2.28	34.16	45.15	6.4	6.94	6.55	0.8	11.84
2	2000	2.31	38.45	39.42	6.82	7.28	6.9	1.13	0.73
3	2000	2.3	31.34	30.66	8.36	12.91	14.5	2.23	0.07
4	2000	2.38	37.61	42.96	5.62	6.09	6.54	1.16	12.56
5	2000	2.34	45.84	42.66	4.68	3.5	2.96	0.36	6.75
1	3000	1.78	53.43	35.63	6.95	3.21	0.77	0	0
2	3000	1.89	51.81	32.3	8.33	5.64	1.89	0.04	0
3	3000	1.89	28.61	36.51	20.55	13.35	0.98	0	0
4	3000	1.81	36.85	39.29	11.81	9.73	2.33	0	0

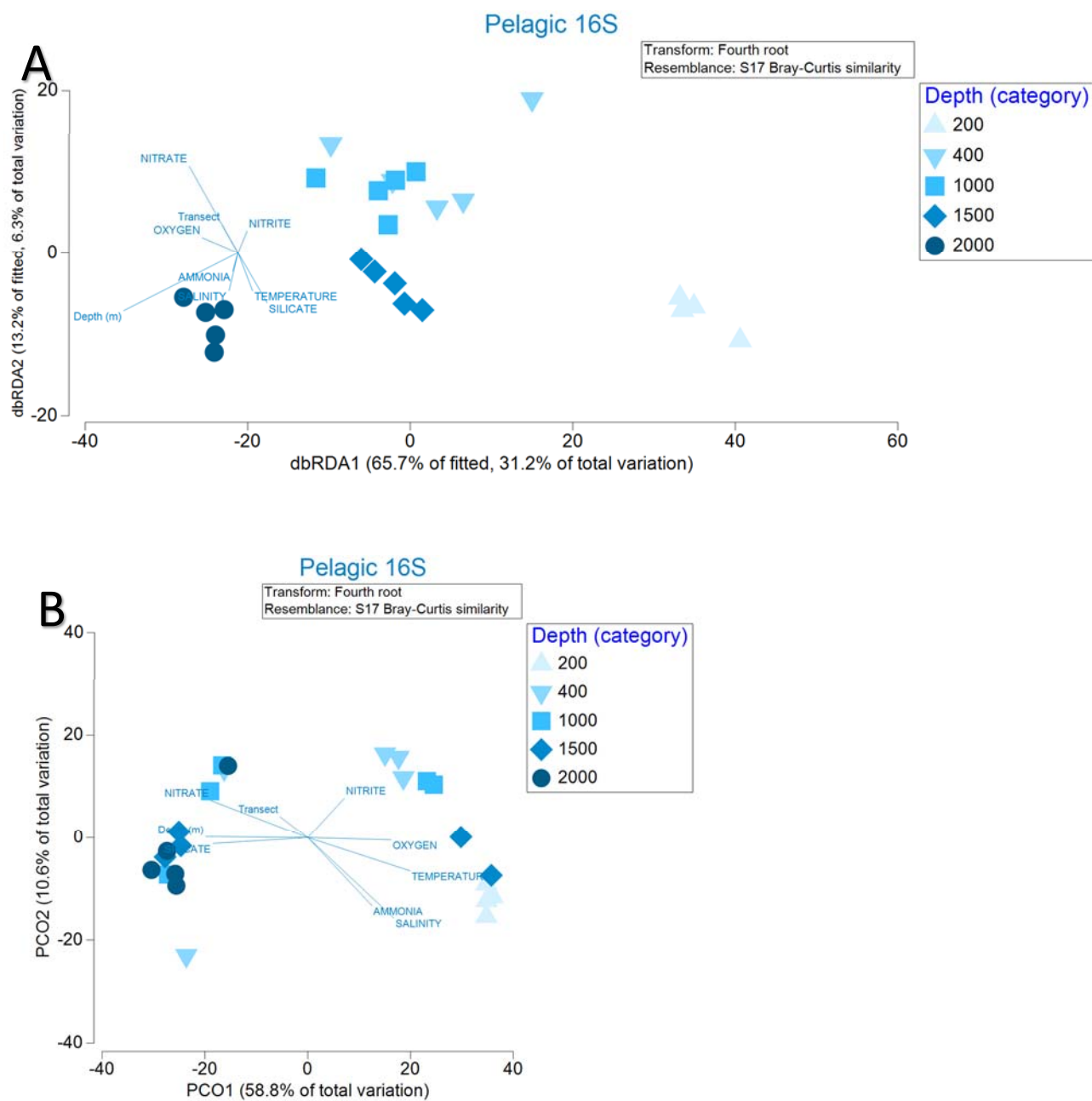


Figure 5.21. Relationships between community composition in pelagic bacterial samples and environmental variables as determined via distance based redundancy analysis (panel A) or principle co-ordinates analysis (panel B).

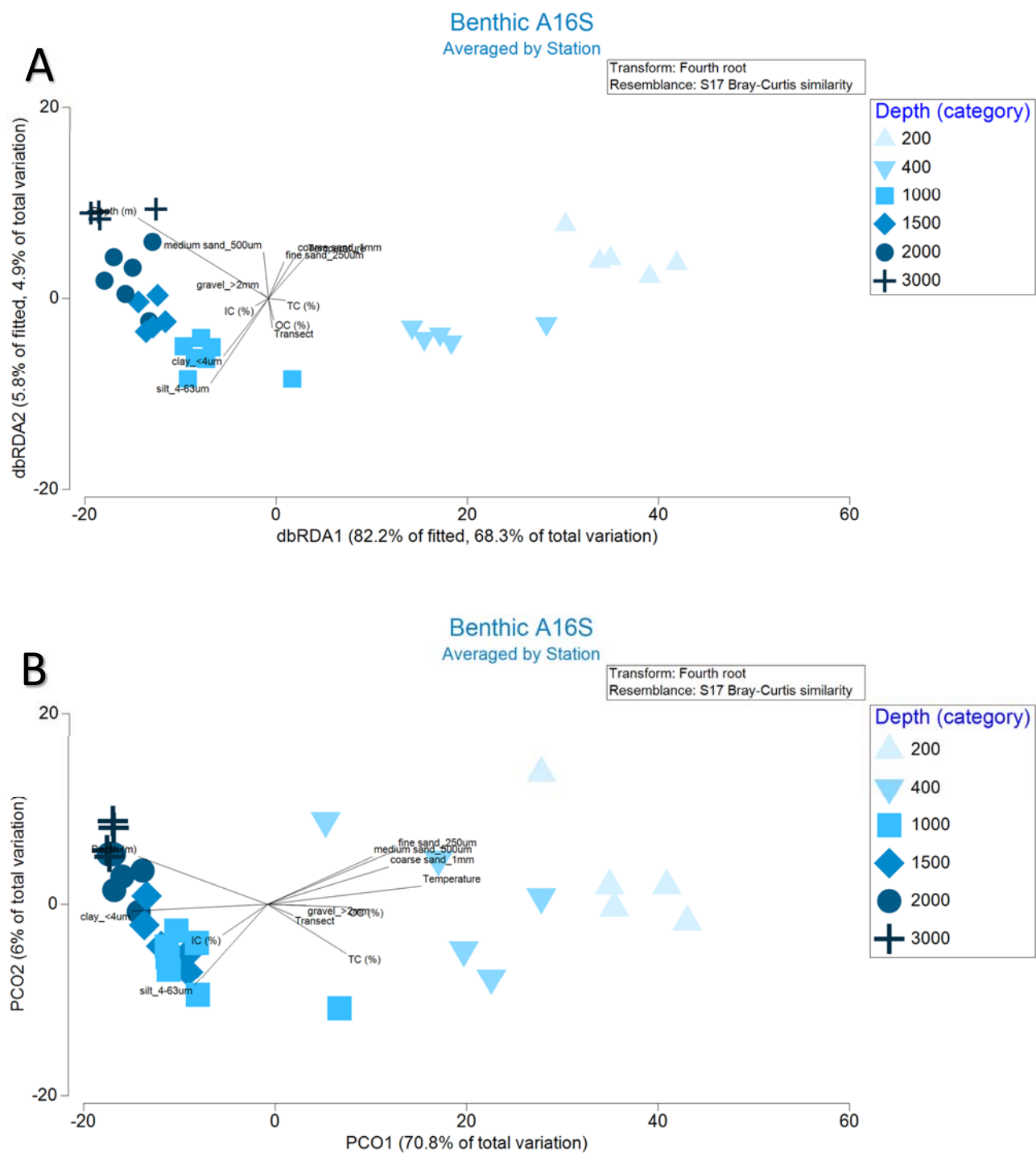
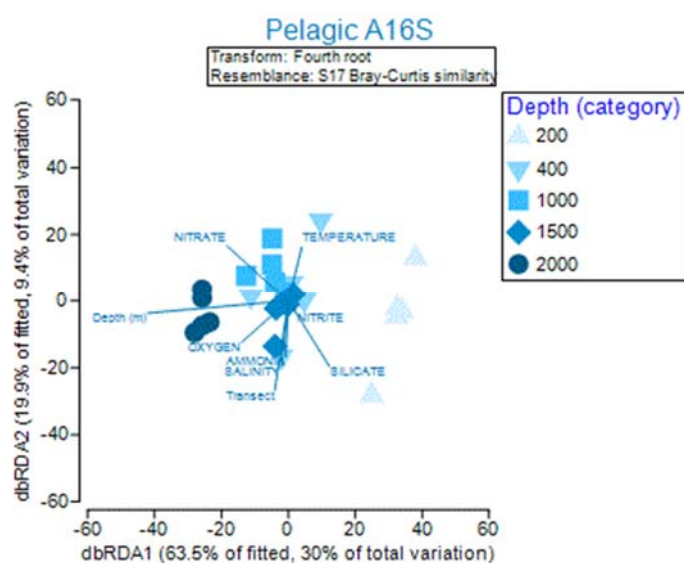


Figure 5.22. Relationships between community composition in benthic archaeal samples and environmental variables as determined via distance based redundancy analysis (panel A) or principle co-ordinates analysis (panel B).

A



B

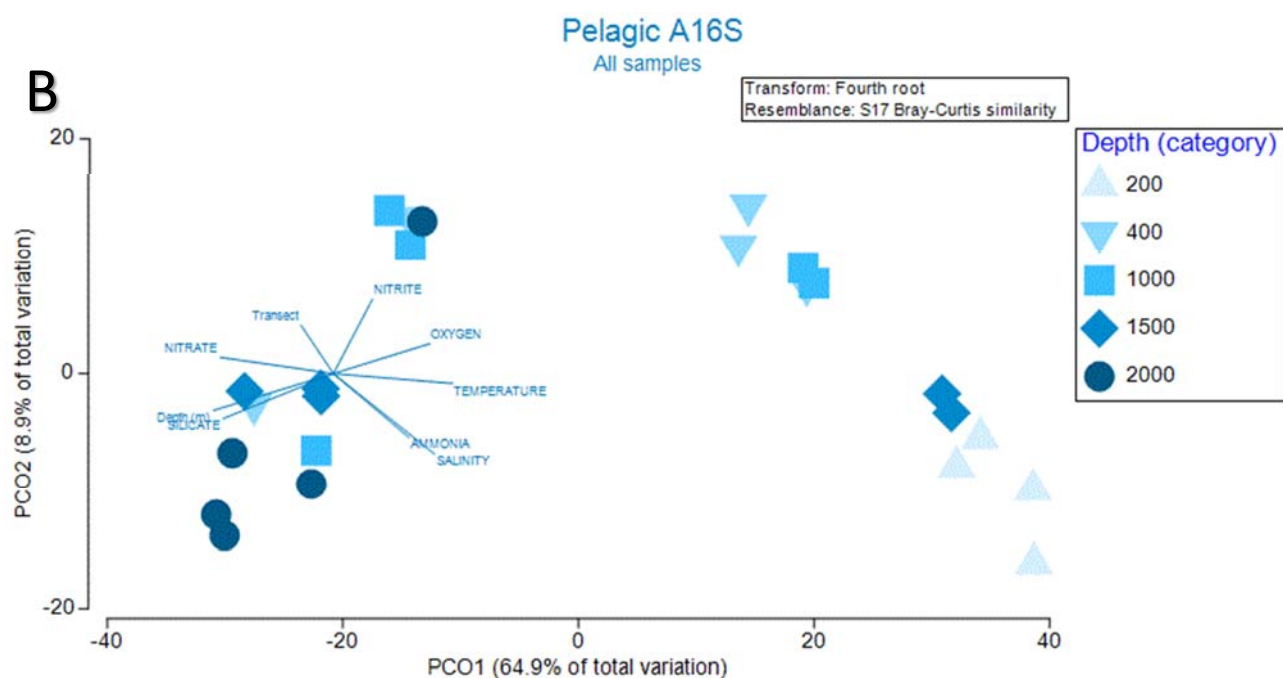


Figure 5.23. Relationships between community composition in pelagic archaeal samples and environmental variables as determined via distance based redundancy analysis (panel A) or principle co-ordinates analysis (panel B).

5.5 Discussion

The study presented in this report provides important baseline information about the indigenous microorganisms in the GAB. At the stations surveyed, there is a greater diversity in the bacterial community than in the archaeal community. The major taxonomic groups present in both benthic and

pelagic samples were identified. The composition of the benthic samples was more homogenous in samples collected along different transects but at the same depths than were the pelagic samples, as indicated by the relatedness in the PCO plots. There was an overall decrease in species diversity and evenness with depth in the benthic samples, but not the pelagic samples, likely due to the increased inter-station variability. Community composition of both Bacteria and Archaea differs between the shelf and the slope sites. For benthic samples, differences in community composition were best predicted by depth and temperature, although there were also strong relationships to sediment grain size, although these variables were poorly represented in the final models as they were highly correlated to depth and/or temperature. For the pelagic samples, temperature and oxygen level were the best predictor variables, although nutrient concentrations and depth were highly correlated to these, and also good predictors in their own right. At all stations surveyed, bacteria related to known hydrocarbon degrading taxa were identified. However, bacteria related to hydrocarbon degrading taxa likely do not all have the capacity to degrade hydrocarbons, and may instead fill other ecological roles. The distribution of microorganisms with the capacity to degrade petroleum will be best understood using a weight of evidence approach – incorporating both the community structure discussed in this report as well as the functional genomics discussed in Chapter 4. The GAB is relatively pristine, consequently about 1% of taxa would be expected to have the ability to degrade oil (Atlas 1995).

The rate at which these indigenous bacteria would respond to a hydrocarbon source such as an oil spill is not known. Since the GAB has few known sources of hydrocarbons (either as seeps or from anthropogenic contamination), the indigenous microbes may not respond as quickly to an oil spill as was seen in the GOM (e.g. Hazen et al. 2010) or as would be expected in other, more contaminated environments (Bargiela et al. 2015). Bacteria in environments with continuous inputs of crude oil are expected to respond more quickly to pollution events (Bargiela et al. 2015). Mesocosm studies or controlled releases may be the best means of determining the rate of response of the indigenous microbial community within the GAB.

Known hydrocarbon degrading bacteria, including members of the γ - Proteobacteria, were more abundant in sediment and water samples impacted by the *Deepwater Horizon* oil spill (Hazen et al. 2010, Kostka et al. 2011, Kimes et al. 2013). In addition, following the *Deepwater Horizon* oil spill, there was a change in the abundance of hydrocarbon degrading bacteria that corresponded with the predominant petroleum hydrocarbons in the water column (Chakraborty et al. 2012, Dubinsky et al. 2013). Although multiple modes of hydrocarbon metabolism were occurring simultaneously, the relative dominance of each, inferred from taxonomic composition as measured using the Phylochip, changed as the oil was captured and the well shut in. When oil was freely flowing, *Oceanospirillaceae* and *Pseudomonas* were abundant in the water column, but their abundance declined after the wellhead was shut (Dubinsky et al. 2013). The PAH degrading bacterial groups including *Colwellia*, *Cycloclasticus*, and *Pseudoalteromonas*, as well as the methanotroph *Methylomonas*, were abundant for a longer time period. Distance from the wellhead, temperature and nutrient availability did not seem to change the bacterial composition within the plume – only the chemical composition of the oil itself did (Chakraborty et al. 2012, Dubinsky et al. 2013).

Bacteria from related taxonomic groups to those found after the *Deepwater Horizon* oil spill have been identified in our initial analysis of samples from the GAB. Potential hydrocarbon degraders in the α , δ , and γ Proteobacteria were all identified, as were methane degrading bacteria. This suggests that the microorganisms present within the GAB have at least some indigenous metabolic capacity to oxidise hydrocarbons. This could be important, as previous studies have shown that the microbial community structure likely controls the potential of an ecosystem to degrade oil (Kostka et al. 2014).

However, measuring relatedness to hydrocarbon degrading bacteria alone may not predict all bacteria with the capacity to degrade hydrocarbons, as some studies have shown that the capacity to degrade PAHs in particular is encoded on a plasmid and may be conferred by horizontal gene transfer (Louvado et al. 2015).

Determining which bacteria will respond to a rapid influx of oil, as occurs in an oil spill, will be more challenging. Some reports from the *Deepwater Horizon* oil spill indicate that the microbial community was dominated temporarily by a few, rare, specialist oil degrading taxa (Yang et al. 2016). The bacteria that responded had been rare before the spill (Kleindienst et al. 2015) and were rare once increased concentrations of hydrocarbons could no longer be detected in the water column. Known hydrocarbon degrading bacteria, including *Oceanospirillales*, *Alcanivorax*, *Marinobacter*, as well as *Methylococcaceae* and *Methylocystaceae*, were detected but were not differentially abundant between sites near the well head and reference sites one year after the *Deepwater Horizon* oil spill (Yergeau et al. 2015). How the microbial assemblage present in the GAB will actually respond to the presence of oil can only be determined by measuring changes in their abundance following an input of oil.

5.6 References

- Atlas RM (1995) Petroleum biodegradation and oil spill bioremediation. *Marine pollution bulletin* 31:178-182
- Bargiela R, Mapelli F, Rojo D, Chouaia B, Tornes J, Borin S, Richter M, Del Pozo MV, Cappello S, Gertler C, Genovese M, Denaro R, Martinez-Martinez M, Fodelianakis S, Amer RA, Bigazzi D, Han X, Chen J, Chernikova TN, Golyshina OV, Mahjoubi M, Jaouanil A, Benzha F, Magagnini M, Hussein E, Al-Horani F, Cherif A, Blaghen M, Abdel-Fattah YR, Kalogerakis N, Barbas C, Malkawi HI, Golyshin PN, Yakimov MM, Daffonchio D, Ferrer M (2015) Bacterial population and biodegradation potential in chronically crude oil-contaminated marine sediments are strongly linked to temperature. *Scientific Reports* 5
- Bray, JR, & Curtis, J.T (1957) An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4), 326-349
- Chakraborty R, Borglin SE, Dubinsky EA, Andersen GL, Hazen TC (2012) Microbial Response to the MC-252 Oil and Corexit 9500 in the Gulf of Mexico. *Frontiers in Microbiology* 3:357
- Chatfield C & Collins AJ (1980) Introduction to multivariate analysis. London, UK: Chapman and Hall
- Clarke, KR (1993) Nonparametric Multivariate Analyses of Changes in Community Structure. *Australian Journal of Ecology*, 18(1), 117-143. Doi 10.1111/J.1442-9993.1993.Tb00438.X
- Delong, EF (1992) Archaea in Coastal Marine Environments. *Proceedings of the National Academy of Sciences of the United States of America*, 89(12), 5685-5689. Doi 10.1073/Pnas.89.12.5685
- DeSantis, TZ, Hugenholtz, P, Larsen, N, Rojas, M, Brodie, EL, Keller, KAndersen, GL (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72(7), 5069-5072. doi: 10.1128/AEM.03006-05
- Dubinsky EA, Conrad ME, Chakraborty R, Bill M, Borglin SE, Hollibaugh JT, Mason OU, Piceno YM, Reid FC, Stringfellow WT, Tom LM, Hazen TC, Andersen GL (2013) Succession of Hydrocarbon-Degrading Bacteria in the Aftermath of the Deepwater Horizon Oil Spill in the Gulf of Mexico. *Environmental Science & Technology* 47:10860-10867
- Edgar, RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19),

- Edgar, RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10(10), 996-+. doi: 10.1038/NMETH.2604
- Hazen TC, Dubinsky EA, DeSantis TZ, Andersen GL, Piceno YM, Singh N, Jansson JK, Probst A, Borglin SE, Fortney JL, Stringfellow WT, Bill M, Conrad ME, Tom LM, Chavarria KL, Alusi TR, Lamendella R, Joyner DC, Spier C, Baelum J, Auer M, Zemla ML, Chakraborty R, Sonnenthal EL, D'Haeseleer P, Holman HY, Osman S, Lu Z, Van Nostrand JD, Deng Y, Zhou J, Mason OU (2010) Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science* 330:204-208
- Head IM, Jones DM, Roling WFM (2006) Marine microorganisms make a meal of oil. *Nature Reviews Microbiology* 4:173-182
- Jorgensen, BB (1982) Mineralization of Organic-Matter in the Sea Bed - the Role of Sulfate Reduction. *Nature*, 296(5858), 643-645. Doi 10.1038/296643a0
- Joye SB (2015) Deepwater Horizon, 5 years on. *Science* 349:592-593
- Kappell AD, Wei Y, Newton RJ, Van Nostrand JD, Zhou J, McLellan SL, Hristova KR (2014) The polycyclic aromatic hydrocarbon degradation potential of Gulf of Mexico native coastal microbial communities after the Deepwater Horizon oil spill. *Frontiers in Microbiology* 5: 205.
- Kimes NE, Callaghan AV, Aktas DF, Smith WL, Sunner J, Golding BT, Drozdowska M, Hazen TC, Suflita JM, Morris PJ (2013) Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. *Frontiers in Microbiology* 4:50
- Kleindienst S, Grim S, Sogin M, Bracco A, Crespo-Medina M, Joye SB (2015) Diverse, rare microbial taxa responded to the Deepwater Horizon deep-sea hydrocarbon plume. *The ISME journal* Doi: 10.1038/ismej.2015.121
- Kostka JE, Prakash O, Overholt WA, Green SJ, Freyer G, Canion A, Delgadio J, Norton N, Hazen TC, Huettel M (2011) Hydrocarbon-Degrading Bacteria and the Bacterial Community Response in Gulf of Mexico Beach Sands Impacted by the Deepwater Horizon Oil Spill. *Applied and Environmental Microbiology* 77:7962-7974
- Kostka JE, Teske A, Joye SB, Head IM (2014) The metabolic pathways and environmental controls of hydrocarbon biodegradation in marine ecosystems. *Frontiers in Microbiology* 5:471.
- Lane, DJ, Pace, B, Olsen, GJ, Stahl, DA, Sogin, ML, & Pace, NR (1985) Rapid-Determination of 16s Ribosomal-Rna Sequences for Phylogenetic Analyses. *Proceedings of the National Academy of Sciences of the United States of America*, 82(20), 6955-6959. Doi 10.1073/Pnas.82.20.6955
- Louvado A, Gomes NCM, Simoes MMQ, Almeida A, Cleary DFR, Cunha A (2015) Polycyclic aromatic hydrocarbons in deep sea sediments: Microbe-pollutant interactions in a remote environment. *Science of the Total Environment* 526:312-328
- Miralles, G, Grossi, V, Acquaviva, M, Duran, R, Bertrand, JC, & Cuny, P (2007) Alkane biodegradation and dynamics of phylogenetic subgroups of sulfate-reducing bacteria in an anoxic coastal marine sediment artificially contaminated with oil. *Chemosphere*, 68(7), 1327-1334. doi: 10.1016/j.chemosphere.2007.01.033
- Schloss, PD, Westcott, SL, Ryabin, T, Hall, JR, Hartmann, M, Hollister, EB, Weber, CF (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, 75(23), 7537-7541. doi: 10.1128/AEM.01541-09
- Wang, Q, Garrity, GM, Tiedje, JM, & Cole, JR (2007) Naive Bayesian classifier for rapid assignment of

- rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73(16), 5261-5267. doi: 10.1128/AEM.00062-07
- Yang T, Nigro LM, Gutierrez T, D'Ambrosio L, Joye SB, Highsmith R, Teske A 2016. Pulsed blooms and persistent oil-degrading bacterial populations in the water column during and after the Deepwater Horizon blowout. *Deep Sea Research Part II: Topical Studies in Oceanography*. 129: 282-291.
- Yergeau E, Maynard C, Sanschagrin S, Champagne J, Juck D, Lee K, Greer CW (2015) Microbial Community Composition, Functions, and Activities in the Gulf of Mexico 1 Year after the Deepwater Horizon Accident. *Applied and Environmental Microbiology* 81:5855-5866
- Yeung CW, Lee K, Cobanli S, King T, Bugden J, Whyte LG, Greer CW (2015) Characterization of the microbial community structure and the physicochemical properties of produced water and seawater from the Hibernia oil production platform. *Environmental Science and Pollution Research International* 22:17697-17715
- Zhang, W, Song, LS, Ki, JS, Lau, CK, Li, XD, & Qian, PY (2008) Microbial diversity in polluted harbor sediments II: Sulfate-reducing bacterial community assessment using terminal restriction fragment length polymorphism and clone library of *dsrAB* gene. *Estuarine Coastal and Shelf Science*, 76(3), 682-691. doi: 10.1016/j.ecss.2007.07.039

6. DEVELOPMENT OF A FUNCTIONAL GENOMICS ASSAY TO MEASURE THE ABUNDANCE OF HYDROCARBON DEGRADING BACTERIA IN THE GREAT AUSTRALIAN BIGHT

Sharon Hook¹, Jodie van de Kamp², Alan Williams², Jason E. Tanner³, Levente Bodrossy²

1. CSIRO Oceans and Atmosphere Lucas Heights, NSW 2234

2. CSIRO Oceans and Atmosphere Hobart, TAS 7001

3. SARDI Aquatic Sciences West Beach, SA 5024

Reprinted with minor amendments from:

Hook, S., van de Kamp, J., Williams, A., Tanner, J.E., Bodrossy, L. 2016. Development of a functional genomics assay to measure the abundance of hydrocarbon degrading bacteria in the Great Australian Bight. Great Australian Bight Research Program.

6.1 Executive summary

Being able to quantify hydrocarbon loss to microbial degradation is important for both the risk assessment for marine oil wells, and to determine the amount of petroleum released into the environment either from a spill or routine discharges. With that in mind, we used previously reported high throughput sequencing results to design quantitative PCR primers specific to the microbial taxa indigenous to the Great Australian Bight (GAB). These primers were intended for use in functional gene assays that would quantify the changes in abundance of bacteria involved in alkane, methane or aromatic hydrocarbon degradation. Using these primers, we were able to measure the abundance of hydrocarbon degrading bacteria from environmental samples collected from two surveys in the GAB, the first on the RV Southern Surveyor in April 2013, and the second on RV Investigator in December 2015. Although these bacteria were rare (between 10^{-5} to 0.1% of the total bacterial population), they were present at every station and every sample we analysed. Although the previously conducted sequencing efforts were instrumental in identifying the sequences of the selected genes in bacteria indigenous to the GAB, the tool is not adequately high throughput (e.g. low cost, quick turnaround) for routine monitoring. By contrast, these qPCR assays can be used as both a high throughput screening tool to measure the spatial and temporal duration of impact from routine discharges of petroleum and to monitor environmental fate and persistence following an oil spill, especially if coupled with a Fluidigm system or similar technology.

6.2 Introduction

When oil is released into the environment, its composition and concentration changes through a process known as weathering (RSC, 2015). Different hydrocarbons within oil weather differently, and consequently experience a different environmental fate and persistence. Some of these compounds will evaporate, some will be subject to photolysis, and others will adhere to sediment particles and sink (reviewed in RSC, 2015). Ultimately, much of the oil that is released will be degraded by bacteria. Oil degrading bacteria have been detected in all environments studied to date (Atlas, 1995a, b), including the Great Australian Bight (GAB) (Hook et al., 2016 a,b). Microbial degradation is thought to be the fate of most oil released into the environment via natural processes

such as seeps (Atlas and Hazen, 2011; Head et al., 2006), and the fate of approximately half of the oil released from the *Deepwater Horizon* oil spill (Kleindienst et al., 2015).

When oil is accidentally released into the environment, being able to measure the rates of oil degradation is useful when planning the response (Dubinsky et al., 2013; Hazen et al., 2010; Lamendella et al., 2014; Lu et al., 2012). Knowing rates of oil degradation enables accurate predictions of fate and persistence, and assists assessment of environmental risk. It also assists calculation of how much oil was released into the environment. Finally, being able to measure rates of oil degradation enables determination of the efficacy of response technologies deployed in the event of a spill, such as chemical oil dispersion or bioremediation.

Exploitation of the petroleum resources in the GAB carries with it the risk of an oil spill, either via a well head blowout or from transport of the resources. Companies involved in petroleum resource development will also be required to monitor the impacts of any routine discharges, such as those due to any minor leaks around the well head. A high throughput functional genomics based assay to monitor changes in the abundance of hydrocarbon degrading bacteria will be useful in either scenario. The abundance of functional genes associated with hydrocarbon degradation increases in samples with a source of oil, and returns to “background” levels once that oil source is degraded (Chakraborty et al., 2012; Dubinsky et al., 2013; Lu et al., 2012). Consequently, we have designed a series of functional gene-based assays to quantify the abundance of bacteria in the environment with the capacity to degrade specific fractions of oil. Since the diversity of hydrocarbon degrading gene sequences from hydrocarbon degrading bacteria in environmental samples collected from the GAB could not be sufficiently described using sequences in NCBI’s GenBank, we first used massively parallel high throughput sequencing to determine the gene variants indigenous to the region, as described in a previous report (Hook et al., 2016a). In this study, we describe quantitative PCR assays that we use to measure the abundance of different variants of these genes in environmental samples. These assays can be used as a tool to monitor both the environmental fate of routine discharges of oil as well as petroleum compounds released during a spill.

6.3 Materials and methods

6.3.1 Study sites and sampling design

Sediment samples were collected from two different research surveys: a survey from 3-22 April 2013 conducted on the *RV Southern Surveyor*, and a second survey from 30 November to 21 December 2015 conducted on the *RV Investigator*. Sampling locations are shown below in Figure 6.1. During each survey, samples were collected using a 6-core multicorer from KC (Denmark) that was controlled from the vessel and allowed reliable collection of sediment samples at depths between 200 and 3,000 m (Sherlock et al., 2014). A 30 mm diameter minicore was extracted from each of 3 cores at each location for microbial analysis. Sediment from the top 20 mm of each minicore was collected into a DNA free tube and frozen. Samples were shipped frozen to CSIRO laboratories for DNA extraction.

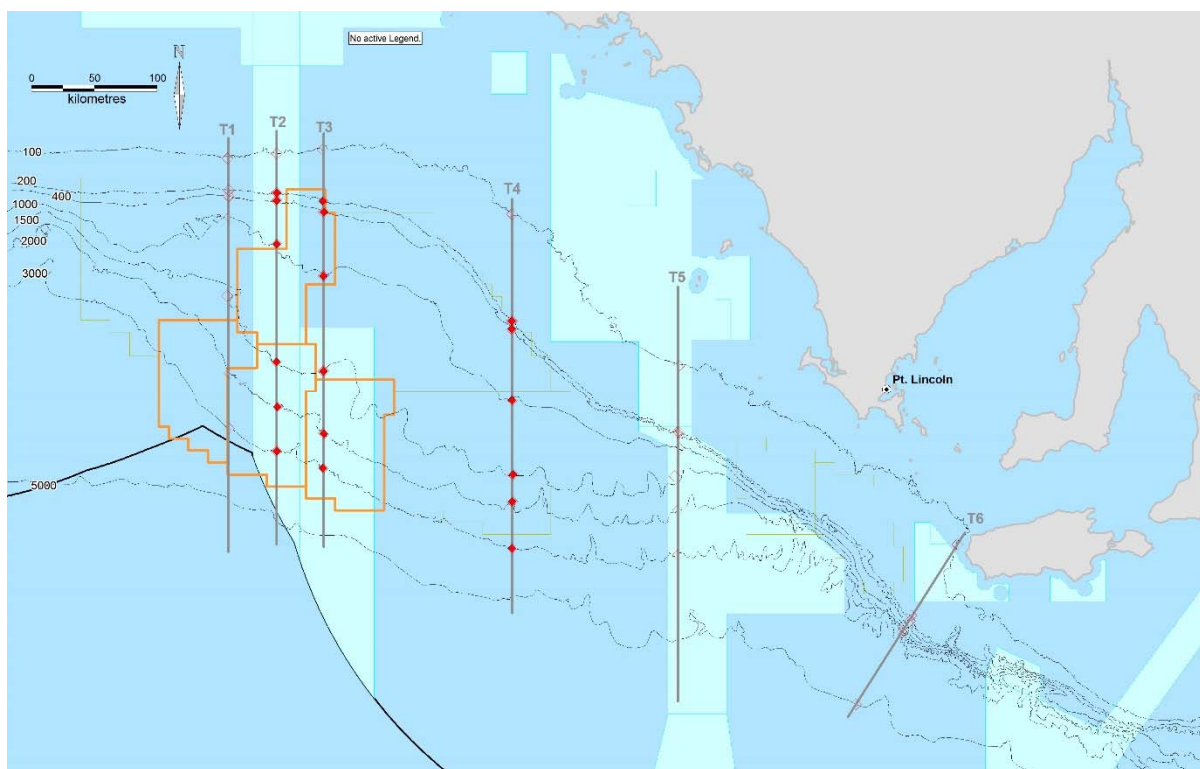


Figure 6.1. Transects and sampling locations used in the 2013 and 2015 surveys. Sampling locations used for development of the functional assays are shown in red. BP's lease blocks are shown in orange, and the light blue shading indicates Commonwealth marine reserves.

6.3.2 DNA extractions

DNA was extracted as described previously (Hook et al., 2016a). DNA was extracted from 10 g of sediment using the MoBio Powermax kit (Mo Bio Laboratories Inc, USA), with the following modifications: after adding the lysis solution (C1), samples were incubated for 10 minute at 70 °C, and all other incubation times were extended to 30 minutes. Once DNA was eluted, the samples were concentrated to dry pellets in a “speed vac” vacuum concentrator, and then washed in 100% ethanol to remove excess salts. A NanoDrop™ 8000 Spectrophotometer (Thermo Scientific™) was used to verify the quality and quantity of all DNA. DNA was aliquoted into multiple tubes, vacuum dried and stored at -20 °C.

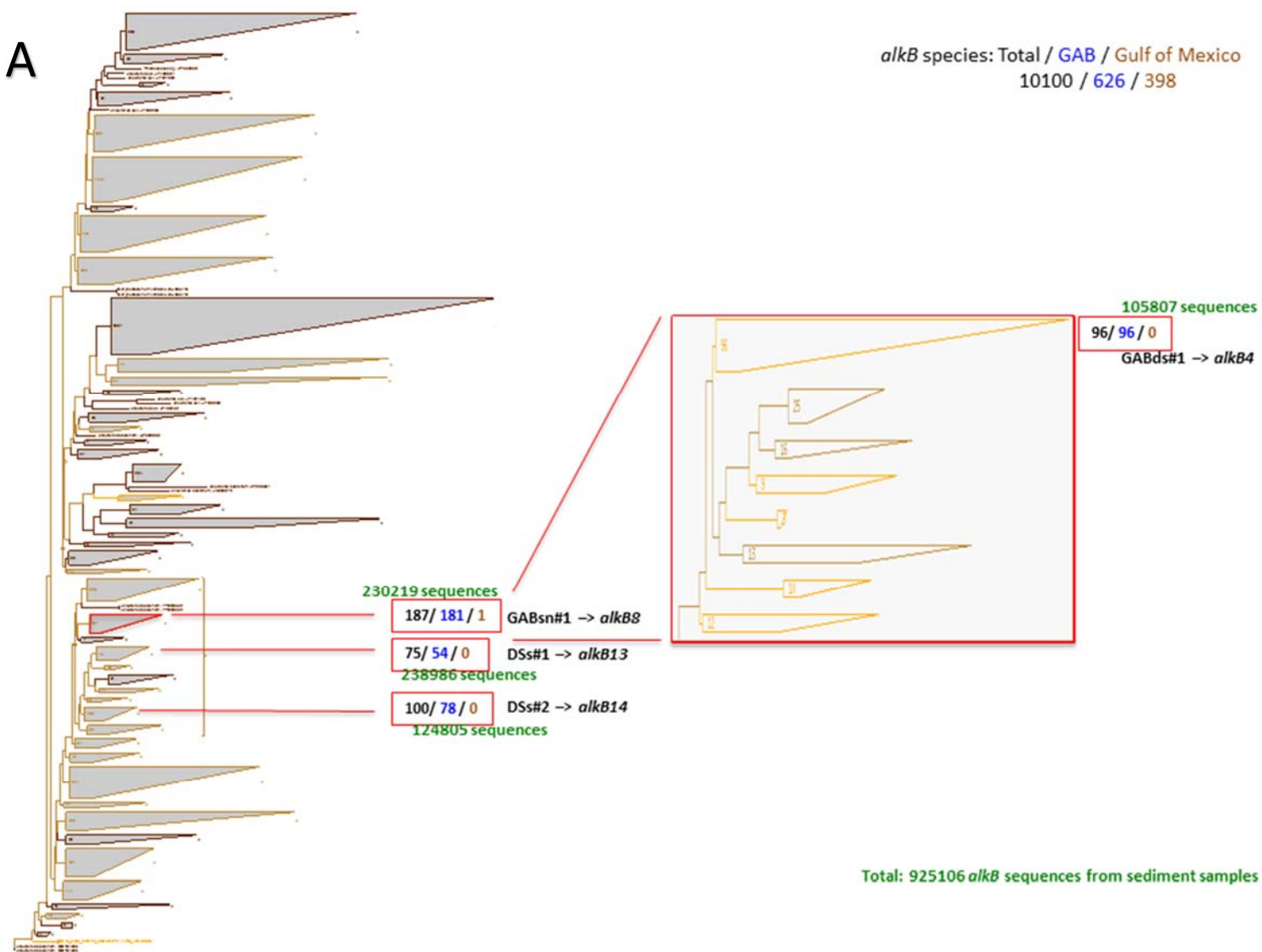
6.3.3 PCR primer design

Databases for *alkB*, *pmoA* and *c23o* that contain all publicly available sequences obtained from the NCBI GenBank databases, including those deposited after the *DeepWater Horizon* oil spill, as well as the representative sequences of OTUs generated from Illumina high throughput sequencing of GAB sediment samples (Hook et al., 2016a) were constructed. Sequences in these databases were aligned and phylogenetic trees were created using the ARB phylogenetic software package (<http://www.arb-home.de/>). Regions of the phylogenetic trees selected for primer design are illustrated in Figure 6.2. Primer pairs were designed to target specific clades of abundant OTUs detected in the GAB sediments (Hook et al., 2016a). The numbers of OTUs in each clade (the diversities within the clades) are given in the numbers in blue next to each clade name. These numbers can be compared to how the same clades have been represented in previous studies as a whole (in black) and specifically in the Gulf of Mexico (in brown). The numbers of amplicons from bacteria (i.e. the relative abundance of each taxon) in each clade are in green. For *alkB*, 3 major

clades containing between them 50% of the novel diversity discovered within the GAB sediment samples were selected, as shown in Figure 6.2A. These clades also contained >60% of all the *alkB* containing bacteria detected in the GAB sediments. For *pmoA* (Figure 6.2B), clade GABds#1 was chosen as it contained most of the *pmoA* sequences found in the GAB. The Deep Sea #2 clade was also chosen as it has been known for a longer time and includes sequences from a range of deep sea environments. This group was relatively abundant in samples from the GAB. For *c23o*, clades GABsA and GABsD (shown in Figure 6.2C) contained >50% of the novel diversity of *c23o* discovered in the GAB. They also contained >90% of all the *c23o* containing bacteria detected in the GAB sediments.

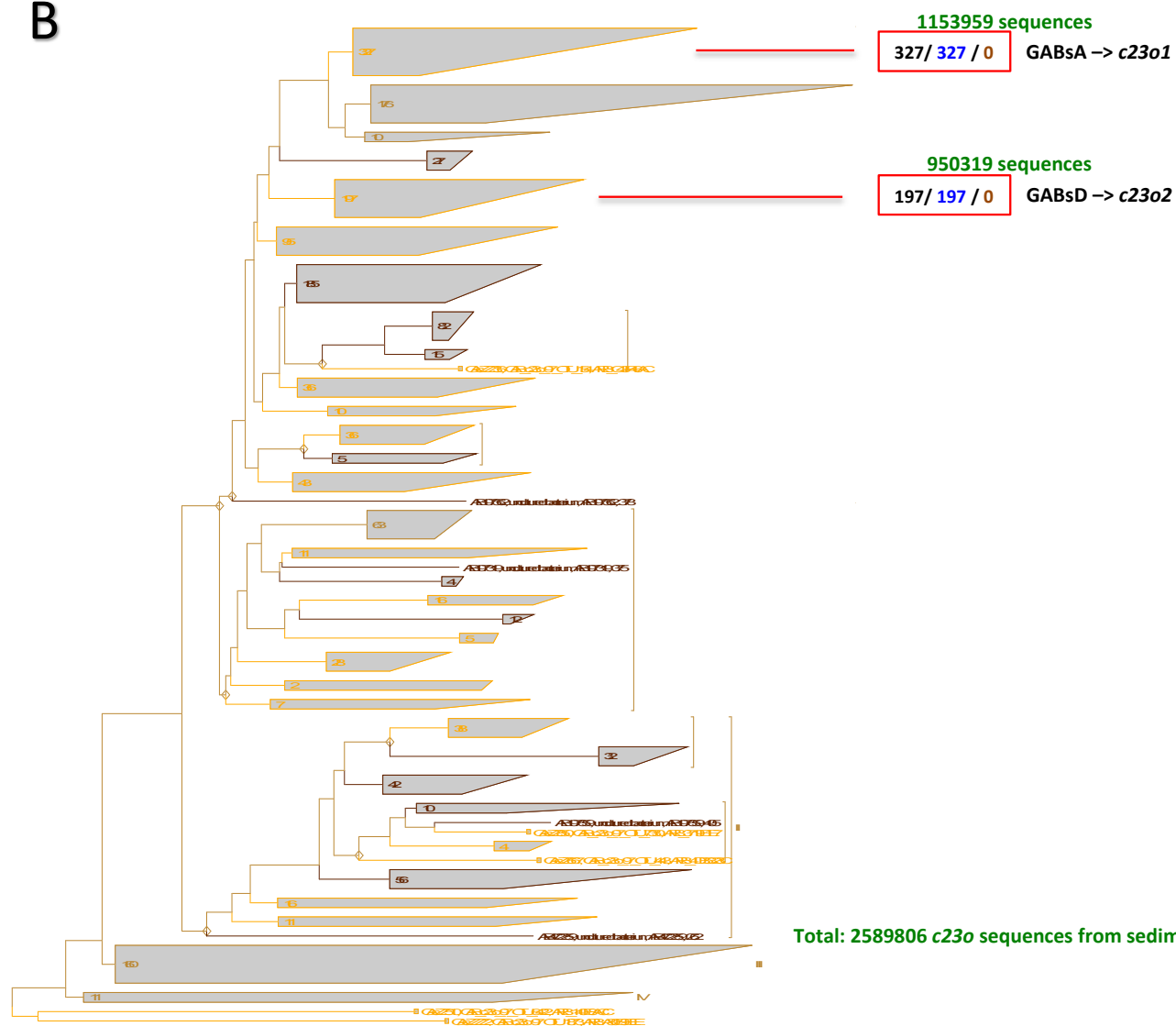
Primer pairs were designed to match common criteria for quantitative PCR (e.g. Mackay, 2004). These were: 1) Similar annealing temperatures, 2) amplicon length between 70 and 300 nucleotides, 3) maximized specificity and coverage for the targeted clade. Primer details are given in Table 6.1. Generic (broad specificity, targeting known diversity of the functional marker genes) qPCR primers from the literature were adapted and tested as well.

A



B

c23o sequences: Total / GAB / Gulf of Mexico
1869 / 930 / 0



C

pmoA sequences: Total / GAB / Gulf of Mexico
13080 / 91 / 39

27859 sequences

160 / 18 / 10 Deep Sea #2 → *pmoA1*

85325 sequences

7 / 7 / 0 GABs#1 → *pmoA3*

Total: 155796 *pmoA* sequences from sediment samples

27859 sequences

160/ 18 / 10	Deep Sea #2 → <i>pmoA1</i>
--------------	----------------------------

160/ 18 / 10 | Deep Sea #2 → *pmoA1*

7/7/0 | GABds#1 → *pmoA3*

Total: 155796 *pmoA* sequences from sediment samples

Figure 6.2. (prev pages) Clades selected for designing *alkB* (panel A), *pmoA* (panel B), and *c23o* (panel C) qPCR primers. The relative abundance of each taxon is given in the numbers in blue next to each clade name. These abundances can be compared to their representation in the literature as a whole (in black) and to their abundance in the Gulf of Mexico (in brown). The numbers of amplicons from bacteria in each clade are in green. Clades outlined in light brown contain sequences that have only been found in the Great Australian Bight. Clades outlined in dark brown contain sequences that have been found in the GAB as well as sequences that have been described in other studies. Clades outlined in black only contain sequences from studies in the literature and do not have members that were identified in this study.

Table 6.1. Primers used in qPCR reactions

Gene	Assay	Forward Primer (5'-3')	Reverse Primer (5'-3')	Annealing Temp (°C)	Acquisition Temp (°C)	Amplicon Size (bp)
alkB	Broad ¹	AAYACNGCNCAYGARCTNGGNCAYAA	GCRTGRTGRTCNGARTGNCGYTG	55	81	548
	4 – GABds#1	CTTTTCCTGGTGGGTGCTNAC	GGTTCGAGTATTTGTGNTTNGC	55	81	136
	8 – GABsn#1	ACCAGACTGGAAACCACNYT	GTTCAATCATCCAGGCCCG	55	84	203
	13 – DSs#1	GUCGGRAUCUGCCCRUAUG	AACCRAACTGCAATAYYGCAGA	55	75	80
	14 – DSs#2	GCTGGATGATGATTCCRTTYT	AACAGCACCAGRTTGCTGA	55	83	183
pmoA	Broad ²	GGNGACTGGGACTTCTGG	GAASGCNGAGAAGAASGC	59	84	531
	1 – Deep Sea #2	GTTAYTTCAAYTTYTGGGGWTGGAC	ACCTTTCTCRATCATMCKRATGTAYTC	59	84	288
	3 – GABds#1	GGTGGGTCGTGGTRACACC	GAACGTMRCCGTGATCCAYTTT	59	84	125
c23o	Broad ³	AAGAGGCATGGGGGCGCACCGTTTCGATCA	CCAGCAAACACCTCGTTGCGGTTGCC	57	85	363
	1 - GABsA	TTYCTGACCGTGTCNATGAARGC	TCGTGCAYGGNNATMAGGTC	57	83	128
	2 - GABsD	GAGCTGACCGARAARGTNACNT	GCATGNCCRACRTCNTCCCA	57	83	132

1. Primers from Kloos et al., 2006
2. Primers from Holmes et al., 1995
3. Primers from Sei et al., 1999

6.3.4 Detection and quantification of *alkB*, *c23o* and *pmoA* genes via qPCR

SybrGreen-based quantitative real-time PCR (qPCR) was performed with the ABI 7500 Real Time PCR System (Applied Biosystems, Foster City, CA, USA). PCR reactions were a total of 15 µL consisting of 7.5 µL of 2x SensiFAST™ SYBR® Lo-Rox Mix (Bioline, London, UK), 0.1 µM each of forward and reverse primers and 1 µL template DNA (environmental template DNA was at a concentration of 10 ng/µL). The temperature at which acquisition of the fluorescence signal occurred was optimized for each assay (as listed in Table 6.1) to decrease the influence of primer dimers. Annealing temperatures were also gene specific (Table 6.1). After an initial activation of 95 °C x 10 min, cycling parameters consisted of 40 cycles of 95 °C for 40 sec, followed by annealing for 30 sec, extension at 72 °C for 60 sec and acquisition for 30 sec. Dissociation curves were run at 95 °C x 15 sec, 60 °C x 60 sec, 95 °C x 30 sec and 60 °C x 15 sec.

DNA standards used in qPCR assays consisted of serial dilutions of purified PCR products derived from GAB sediment samples for each of the gene fragments. Trial PCRs were performed for each primer pair on 10 ng of sediment DNA from Transect 2 and 4. Positive amplifications for each gene were pooled and run on a 1% agarose gel in 1 x TAE buffer. Gene fragments of the correct size were excised with sterile scalpels and gel purified using the Wizard® SV Gel and PCR Clean-Up System (Promega Corporation, Madison, Wisconsin, USA) according to the manufacturers' instructions. The concentration of purified PCR products was measured and serial dilutions made from 10⁻¹ to 10⁻⁸ ng/µL. Each qPCR run consisted of a five-point standard curve, environmental DNA samples and no template controls run in triplicate. Standard curves were linear ($r^2 > 0.98$) and amplification efficiency was >83. Results were analysed using the ABI 7500 Software package V2.3 (Applied Biosystems, Foster City, CA, USA). Final copy numbers of each gene in environmental samples were calculated assuming 100% DNA extraction efficiency, and the results are expressed as copy numbers per gram of sediment.

6.3.5 Statistical Analysis

All statistical analyses were carried out in SigmaPlot (version 12.5) and R (version 3.3.0). Data were log transformed to meet the assumption of normality, then 2 factor ANOVA tests were run with depth and transect as fixed factors, and $\alpha=0.05$. Pairwise comparisons were carried out via Tukey's HSD test, where required, to identify stations that had significantly different abundances of copies for *alkB*, *pmoA*, or *c23o*. Differences between samples collected at the same station in different years were determined via a t-test ($\alpha=0.05$). For each sample, the results of the triplicate qPCR runs were averaged for all analyses.

6.4 Results

In presenting the results, we assumed that each cell had a single copy of the selected genes, as is normal for bacteria, allowing us to directly convert copies per mL sediment to cells per gram sediment. Bacterial genomes that have been sequenced to date typically have a 1 or 2 copies of *Alk-B* or *pMOA* (Wang et al., 2010, Knief, 2015), suggesting that this assumption is robust. In general, bacteria with genes encoding enzymes involved in alkane degradation (*alkB*), methane degradation (*pmoA*) and aromatic hydrocarbon degradation (*c23o*) were detected at every site, even though these areas did not have detectable levels of hydrocarbons (Ahmed et al., 2014). As a consequence of the lack of detectable hydrocarbons, the differences between sites (analysed both as depth and transect) were subtle, but occasionally statistically significant. The significance of these differences in abundance with regards to the environmental processes ongoing in the GAB is difficult to interpret.

6.4.1 *alkB*

Trends in abundance of *alkB* are shown in

Figure 6.3. Microbial cells with the capacity to degrade alkanes are comparatively abundant, as measured with the broad primers (panel A) and GAB specific primers from clade 8 (panel C). Both primer sets detected

between 10^4 and 10^5 copies of the gene for *alkB* per gram of sediment. This is slightly different from the original sequencing results, which detected 230,219 sequences in clade 8 (Figure 6.2A) out of about one million total *alkB* sequences. For reference, a gram of sediment contains 10^8 - 10^9 bacterial cells as a generalisation (Bargiela et al., 2015). The *alkB* 4 primer pair, targeting a subgroup of those targeted by the *alkB* 8 pair, also showed abundant copies of this gene, as shown in Figure 6.3B. The original sequencing efforts detected 105,807 copies of *alkB* that aligned to this clade. By contrast, the primers designed based on sequences that were abundant in other regions as well as in the GAB, *alkB* 13 and *alkB* 14, detected far fewer copies of this gene in the same samples (

Figure 6.3D and E). However, the previously conducted sequencing results indicated that *alkB* 13 and 14 had abundances comparable to *alkB* 8 and *alkB* 4, respectively (Figure 6.2A). For all primers, the number of gene copies was significantly influenced by an interaction between transect and depth (ANOVA: $F_{10,38} > 2.16$, $p < 0.042$).

Comparisons between results from high throughput sequencing and the qPCR findings need to be interpreted with caution. The high throughput sequencing results are based on a generic PCR, which is prone to bias. PCR biases have been reported to cause an order or more of magnitude over- or underrepresentation of relative abundances of subgroups (see Acinas et al., 2005; Engelbrektson et al., 2010; O'Donnell et al., 2016). Quantitative PCR based numbers are generally more reliable, especially those focusing on smaller, more closely related groups. It should also be noted that the qPCR methods developed here are aimed at detecting substantial changes in the abundance of hydrocarbon degrading bacterial clades in case of increasing hydrocarbon loads. While a qPCR method can also be inaccurate, reporting up to 2-3x higher or lower abundance than the actual abundance, this inaccuracy is consistent and would thus not impact on the quantification of changes.

Although small differences were occasionally detected between different stations, no statistically significant differences were detected between samples taken at Transect 2, 1,000 m depth in different years (broad *alkB*: $t_4 = 2.088$, $p = 0.105$; *alkB*4: $t_4 = 1.608$, $p = 0.183$; *alkB*8: $t_4 = 2.562$, $p = 0.0625$; *alkB*13: $t_4 = 0.616$, $p = 0.571$; *alkB*14: $t_4 = 0.616$, $p = 0.293$).

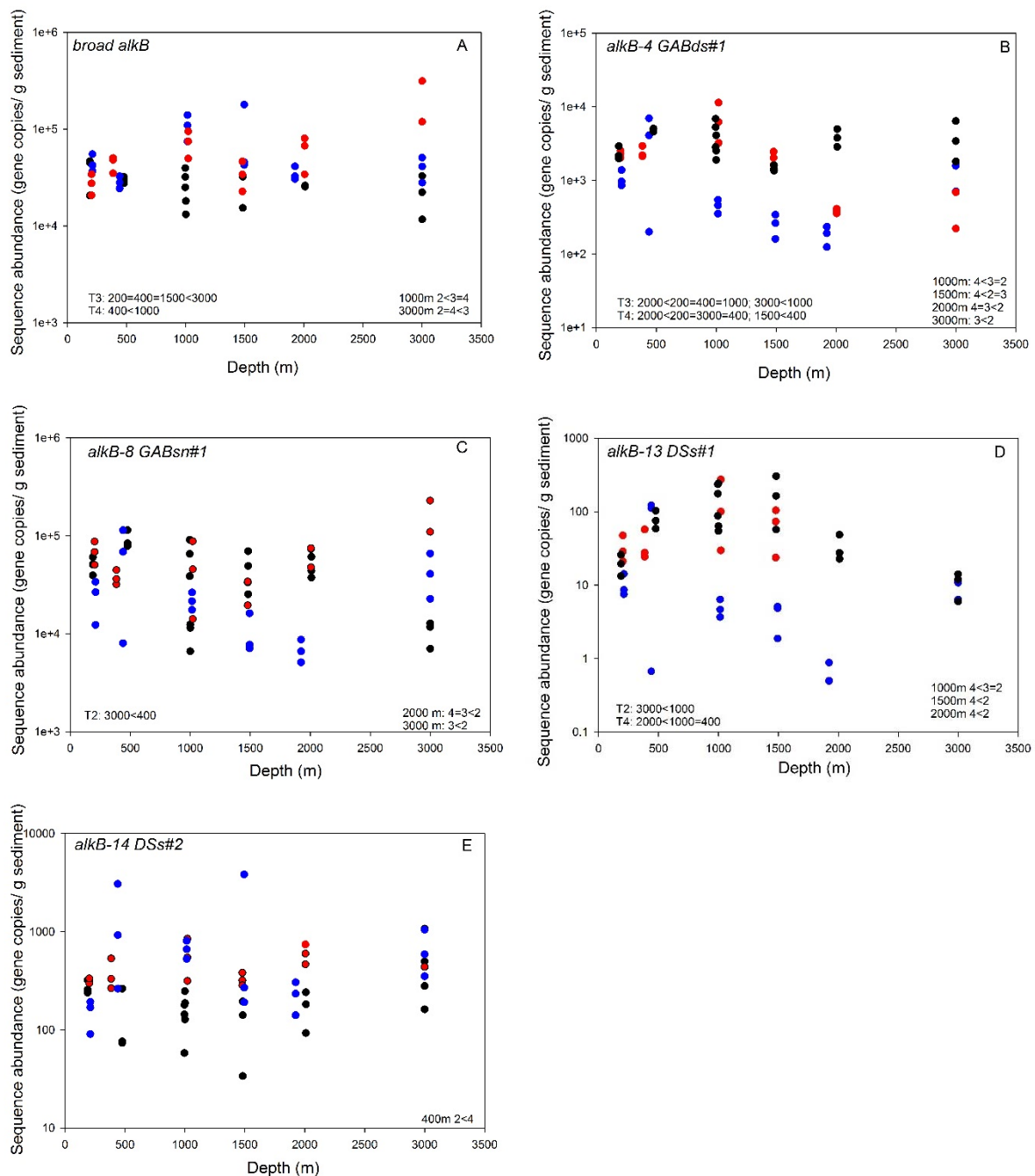


Figure 6.3. Number of copies of *alkB* detected with the different primer sets in sediment samples collected from the GAB. Transect 2 is shown in black, Transect 3 in red, and Transect 4 in blue. Inequalities in the bottom left of each panel indicate statistically significant differences between depths within a transect, while those in the bottom right indicate differences between transects within a depth (ANOVA followed by a Tukey's HSD pairwise comparison, $p < 0.05$).

6.4.2 *pmoA*

The abundance of cells containing the gene sequence for *pmoA* is shown in Figure 6.4. In comparison to *alkB*, cells containing *pmoA* were much less abundant, with roughly 10^2 to 10^3 copies of the gene per

gram sediment measured if the broad *pmoA* (Figure 6.4A) or the primers designed for clade *pmoA* GABDS#1 (*pmoA3*, Figure 6.4C) are used. By contrast, the primers for the *pmoA* DS#1 clade (*pmoA1*) detected between 1 to 100 copies of the gene per gram sediment. In the original sequencing efforts, members of the GABDS#1 clade were more abundant (83525 copies) than members of the DS#1 clade (27859 copies) (Figure 6.2B). For the broad *pmoA* primers (Figure 6.4A), the number of gene copies detected varied between transects (ANOVA: $F_{2,10}=3.32$, $p=0.047$) and depths ($F_{5,10}=3.24$, $p=0.016$), although there was no interaction ($F_{10,34}=0.88$, $p=0.56$). For the primers specific to a single clade (Figure 6.4B and C), there were interactions between transect and depth (ANOVA: $F_{10,38}>3.03$, $p<0.012$).

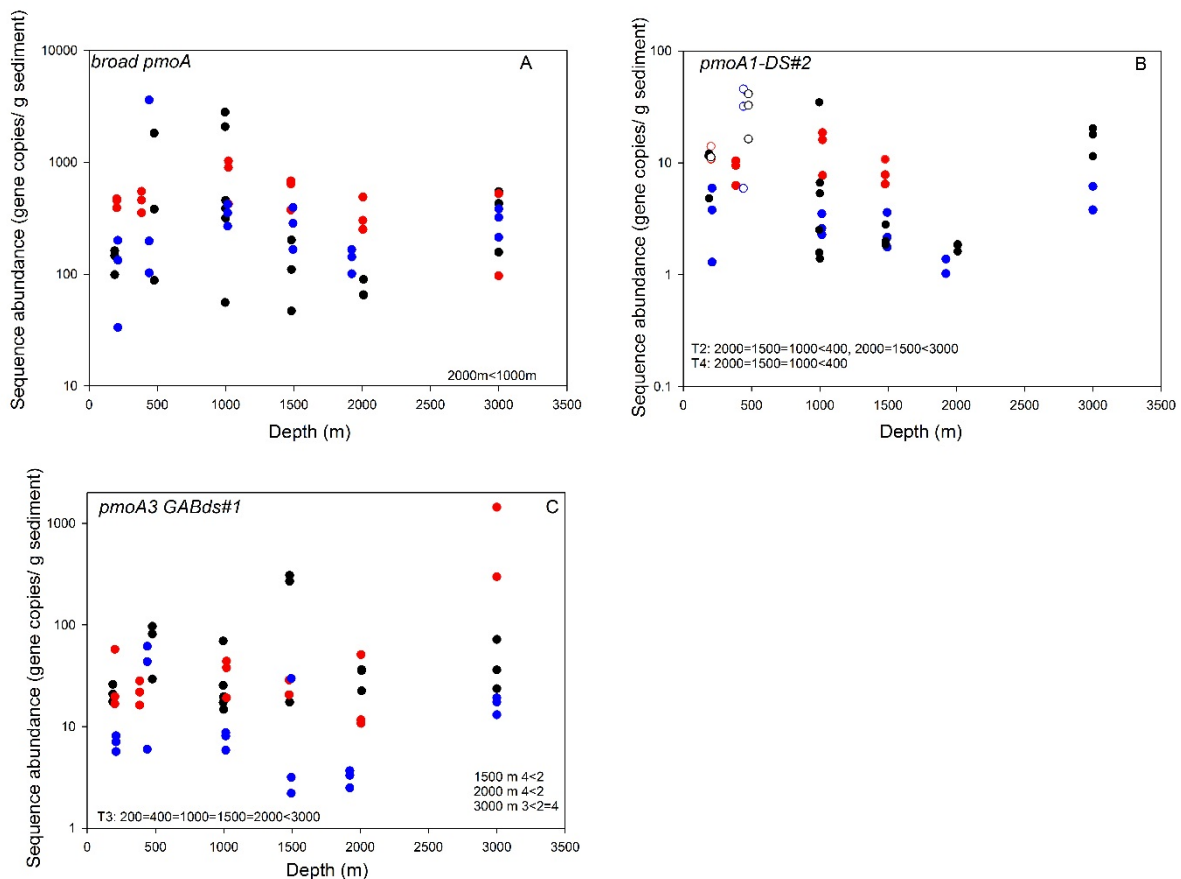


Figure 6.4. Number of copies of *pmoA* detected with the different primer sets in sediment samples collected from the GAB. Transect 2 is shown in black, Transect 3 in red, and Transect 4 in blue. Inequalities in the bottom left of each panel indicate statistically significant differences between depths within a transect, while those in the bottom right indicate differences between transects within a depth except for panel A, where differences are between the main effect of depth (ANOVA followed by a Tukey's HSD pairwise comparison, $p<0.05$).

Although small differences in *pmoA* abundance were detected at different stations, no statistically significant differences were detected between samples taken at Transect 2, 1,000 m depth from the different voyages (t-test, broad *pmoA*: $t_4 = 1.531$, $p = 0.201$; *pmoA1*: $t_4 = 1.531$, $p = 0.484$; *pmoA3*: $t_4 = 1.193$, $p = 0.299$).

6.4.3 c23o

The abundance of cells with a copy of the gene for *c23o* is shown in Figure 6.5. Like *pmoA*, cells containing *c23o* were less abundant than those containing *alkB*, with roughly 1,000 copies of the gene

detected per gram sediment as measured by all primer pairs (Figure 6.5A-C). Both clades had equivalent abundance as measured through both the sequencing efforts (1.15 M and 950,000, respectively, Figure 6.2C) and the qPCR assays (Figure 6.5B and C). There were detectable differences in abundance due to the interaction between transect and depth for *c23o* and *c23o2* (ANOVA: $F_{10,34} > 2.1$, $p < 0.046$), but no significant effects for *c23o1*.

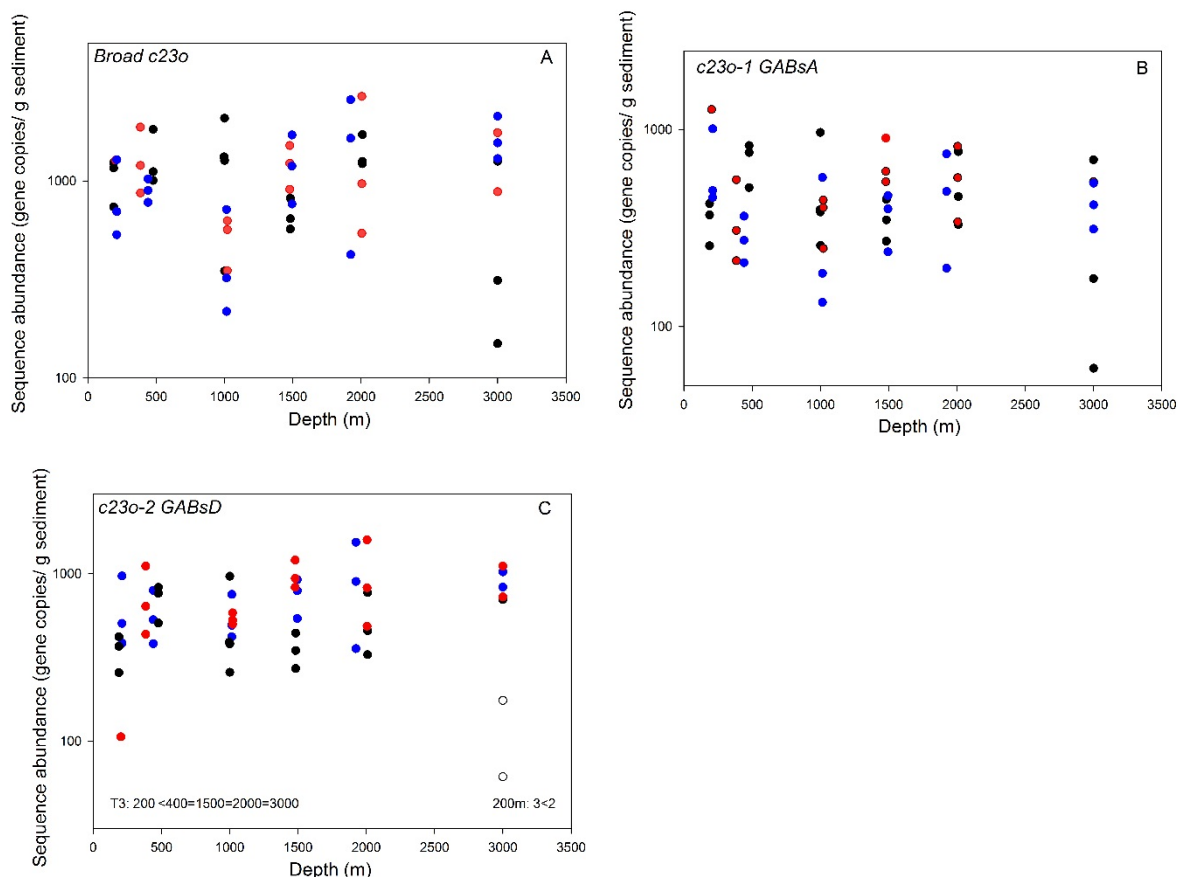


Figure 6.5. Number of copies of *c23o* detected with the different primer sets in sediment samples collected from the GAB. Transect 2 is shown in black, Transect 3 in red, and Transect 4 in blue. Where shown, inequalities in the bottom left of each panel indicate statistically significant differences between depths within a transect, while those in the bottom right indicate differences between transects within a depth (ANOVA followed by a Tukey's HSD pairwise comparison, $p < 0.05$).

Despite measuring small changes in *c23o* abundance at different stations, no statistically significant differences (t-test, broad *c23o*: $t_4 = -0.704$, $p = 0.520$; *c23o-1*: $t_4 = 0.854$, $p = 0.441$; $t_4 = -0.659$, $p = 0.546$) were detected between samples taken at Transect 2, 1,000 m depth in 2013 and 2015.

6.5 Discussion

Our goal was to develop a high throughput functional genomics based screening tool to measure changes in the abundance of hydrocarbon degrading bacteria. The tool has application to monitoring routine discharges during oil and gas production and for evaluating the persistence of oil in the event of an unplanned release.

We focussed our efforts on three genes in three different functional pathways: 1) *alkB*, which is the first step in the breakdown of alkanes (Abbasian et al., 2015), 2) *pmoA*, which is involved in aerobic methane oxidation (Inagaki et al., 2004), and 3) *c23o* which is the conserved ring breaking step in the

degradation of different polycyclic aromatic hydrocarbons (PAH) (Meyer et al., 1999). Because of the high levels of diversity in gene sequences from the GAB, we designed primers from the literature as well as from sequencing results we obtained in an earlier portion of this study (Hook et al., 2016a). Using these primers, we were able to identify genes in the chosen oil degrading pathways at every site, indicating the presence of bacteria capable of hydrocarbon degradation. In our analyses, we assumed one gene copy per cell, allowing us to convert between our copy number per ng/DNA added to the reaction to cells/g sediment. The genes, and in turn the bacteria carrying them, were present at 10^1 to 10^5 cells per gram sediment, making up $1/10^8$ to $1/10^3$ of the total bacterial community. Previous studies have also identified a low but consistent presence of hydrocarbon degrading bacteria even in environments without a local hydrocarbon source (Atlas, 1995b). While $1/10^8$ to $1/10^3$ is a low proportion of the bacterial community, it should be noted that following the *DeepWater Horizon* oil spill, the bacteria that increased in abundance had also been rare previously (Kleindienst et al., 2015). By using primer pairs targeted to different clades, in the case of increased hydrocarbon load, we will be able to better follow the progress of, and differences in, the types of alkane, methane, and aromatic compound degraders that are abundant as oil weathers.

Using these primers, we measured some statistically significant differences in the numbers of bacterial cells carrying these marker genes at different stations. Although the differences in abundance at different stations are statistically significant, the magnitude of the differences is subtle. The small magnitude of differences is not surprising because no source of natural hydrocarbon leakages has been detected in the study area. We could not determine the significance, if any, of these small differences in the abundance of bacterial cells to ecological processes in the GAB with the geochemical data currently available. As these differences are idiosyncratic and vary between the different genes, and even between different primers for the same gene, there were no consistent spatial differences in the abundances of genes related to hydrocarbon degradation. It is thus unlikely that they are related to the presence of a specific source of hydrocarbons, and they may instead be related to some other aspect of environmental variation.

This study was conducted using three different sets of primers: universal primers whose design was obtained from the literature (Kloos et al., 2006, Holmes et al., 1995, Sei et al., 1999), regionally-specific primers we designed based on our previous work with Illumina based high throughput sequencing, and primers designed for clades that were made up of sequences we obtained from the literature, as well as sequences derived from our previous Illumina based assays. The universal primers typically detected the most cells, as would be expected since they target a larger amplicon and were used to select the region targeted for sequencing. However, for *pmoA*, these broad primers lacked the sensitivity to detect subtle changes in the abundance of cells. The regionally specific primers could detect nearly as many cells as the universal primers, despite being designed for a single clade, but detected more statistically significant changes in abundance (Figure 6.4A-C). Their use shows promise to develop a monitoring tool with a greater dynamic range (i.e. applicable at both higher and lower abundances) than the universal primers. Also, because of the shorter amplicon size, these assays are likely to perform better if used in a high throughput scenario.

Cells containing *alkB* were more abundant than cells containing either *c23o* or *pmoA*. Water samples from the Gulf of Mexico collected outside the hydrocarbon plume had a greater relative abundance of *alkB* than *c23o*, however, *pmoA* had equivalent abundance to *alkB* in those samples (Rivers et al., 2013).

As detailed in our previous report (Hook et al., 2016a), amplicon based high throughput sequencing was a highly effective approach to determine the diversity of gene sequences in the GAB. However, we would not recommend using this approach in either routine monitoring or in oil spill response, as it has higher costs per sample, the assays take longer to run, and the data takes longer to analyse. By contrast, the qPCR assays can be run very quickly and cheaply, and can be coupled with a microfluidics device such as the Fluidigm® system (<https://www.fluidigm.com/products/biomark-hd-system>), which can run 96 primer pairs against 96 samples in a single afternoon, to make them high throughput. The regionally specific primers, designed using outputs of our previous sequencing work,

provide both the sensitivity and dynamic range to be utilised in any monitoring efforts that may be needed in the GAB.

6.6 References

- Abbasian, F., Lockington, R., Mallavarapu, M., Naidu, R., 2015. A Comprehensive Review of Aliphatic Hydrocarbon Biodegradation by Bacteria. *Applied Biochemistry and Biotechnology* 176, 670-699.
- Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., Polz, M.F., 2005. PCR-induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and Environmental Microbiology* 71, 8966-8969.
- Ahmed, M., Ross, A., Stalvies, C., Armand, S., Fuentes, D., Gong, S., Sestak, S. 2014. Organic geochemical study of seawater, seabed sediment and headspace gas samples to monitor baseline hydrocarbon levels around BP permits in the Great Australian Bight, Australia. Report to the BP-GAB alliance.
- Atlas, R.M., 1995a. Bioremediation of petroleum pollutants. *International Biodeterioration & Biodegradation*. 35, 317-327.
- Atlas, R.M., 1995b. Petroleum biodegradation and oil spill bioremediation. *Marine pollution bulletin* 31, 178-182.
- Atlas, R.M., Hazen, T.C., 2011. Oil biodegradation and bioremediation: a tale of the two worst spills in U.S. history. *Environmental science & technology* 45, 6709-6715.
- Bargiela, R., Mapelli, F., Rojo, D., Chouaia, B., Tornés, J., Borin, S., et al. 2015. Bacterial population and biodegradation potential in chronically crude oil-contaminated marine sediments are strongly linked to temperature. *Nature Publishing Group*, 5, 11651. <http://doi.org/10.1038/srep11651>.
- Chakraborty, R., Borglin, S.E., Dubinsky, E.A., Andersen, G.L., Hazen, T.C., 2012. Microbial Response to the MC-252 Oil and Corexit 9500 in the Gulf of Mexico. *Frontiers in microbiology* 3, 357.
- Dubinsky, E.A., Conrad, M.E., Chakraborty, R., Bill, M., Borglin, S.E., Hollibaugh, J.T., Mason, O.U., Piceno, Y.M., Reid, F.C., Stringfellow, W.T., Tom, L.M., Hazen, T.C., Andersen, G.L., 2013. Succession of Hydrocarbon-Degrading Bacteria in the Aftermath of the Deepwater Horizon Oil Spill in the Gulf of Mexico. *Environmental science & technology* 47, 10860-10867.
- Engelbrektson, A., Kunin, V., Wrighton, K.C., Zvenigorodsky, N., Chen, F., Ochman, H., Hugenholtz, P., 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *Isme Journal* 4, 642-647.
- Hazen, T.C., Dubinsky, E.A., DeSantis, T.Z., Andersen, G.L., Piceno, Y.M., Singh, N., Jansson, J.K., Probst, A., Borglin, S.E., Fortney, J.L., Stringfellow, W.T., Bill, M., Conrad, M.E., Tom, L.M., Chavarria, K.L., Alusi, T.R., Lamendella, R., Joyner, D.C., Spier, C., Baelum, J., Auer, M., Zemla, M.L., Chakraborty, R., Sonnenthal, E.L., D'Haeseleer, P., Holman, H.Y., Osman, S., Lu, Z., Van Nostrand, J.D., Deng, Y., Zhou, J., Mason, O.U., 2010. Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science* 330, 204-208.
- Head, I.M., Jones, D.M., Roling, W.F., 2006. Marine microorganisms make a meal of oil. *Nature reviews. Microbiology* 4, 173-182.
- Hook, S., van de Kamp, J., Williams, A., Tanner, J., Bodrossy, L. 2016a. Spatial distribution and diversity in hydrocarbon degrading microbes in the Great Australian Bight I: Functional Gene Distribution. Great Australian Bight Research Program, GABRP Research Report Series Number 8, September 2016.
- Hook, S., van de Kamp, J., Williams, A., Tanner, J., Bodrossy, L. 2016b. Spatial distribution and diversity in hydrocarbon degrading microbes in the Great Australian Bight II: Analysis of bacterial and archaeal community structure. Great Australian Bight Research Program, GABRP Research Report Series Number 9, September 2016.

- Kleindienst, S., Grim, S., Sogin, M., Bracco, A., Crespo-Medina, M., Joye, S.B., 2015. Diverse, rare microbial taxa responded to the Deepwater Horizon deep-sea hydrocarbon plume. *The ISME journal*.
- Kloos, K., Munch, J.C., Schloter, M., 2006. A new method for the detection of alkane-monoxygenase homologous genes (alkB) in soils based on PCR-hybridization. *Journal of Microbiological Methods*, 66: 486-496. doi: 10.1016/j.mimet.2006.01.014.
- Knief, C. 2015. Diversity and Habitat Preferences of Cultivated and Uncultivated Aerobic Methanotrophic Bacteria Evaluated Based on pmoA as Molecular Marker. *Frontiers in Microbiology*, 6.
- Lamendella, R., Strutt, S., Borglin, S., Chakraborty, R., Tas, N., Mason, O.U., Hultman, J., Prestat, E., Hazen, T.C., Jansson, J.K., 2014. Assessment of the Deepwater Horizon oil spill impact on Gulf coast microbial communities. *Frontiers in microbiology* 5, 130.
- Lu, Z., Deng, Y., Van Nostrand, J.D., He, Z., Voordeckers, J., Zhou, A., Lee, Y.-J., Mason, O.U., Dubinsky, E.A., Chavarria, K.L., Tom, L.M., Fortney, J.L., Lamendella, R., Jansson, J.K., D'Haeseleer, P., Hazen, T.C., Zhou, J., 2012. Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *ISME Journal* 6, 451-460.
- Mackay, I.M., 2004. Real-time PCR in the microbiology laboratory. *Clinical Microbiology and Infection* 10, 190-212.
- Meyer, S., Moser, R., Neef, A., Stahl, U., Kampfer, P., 1999. Differential detection of key enzymes of polyaromatic-hydrocarbon-degrading bacteria using PCR and gene probes. *Microbiology-UK* 145, 1731-1741.
- O'Donnell, J.L., Kelly, R.P., Lowell, N.C., Port, J.A., 2016. Indexed PCR Primers Induce Template-Specific Bias in Large-Scale DNA Sequencing Studies. *PloS one* 11.
- Rivers, A.R., Sharma, S., Tringe, S.G., Martin, J., Joye, S.B., Moran, M.A., 2013. Transcriptional response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill. *The ISME journal* 7, 2315-2329.
- RSC. 2015. *The Behaviour and Environmental Impacts of Crude Oil Released into Aqueous Environments*. Royal Society of Canada, Ottawa, Ontario, Canada. <<https://rsc-src.ca/en/expert-panels/rsc-reports/behaviour-and-environmental-impacts-crude-oil-released-into-aqueous>>.
- Sei, K., Asano, K., Tateishi, N., Mori, K., Ike, M., Fujita, M., 1999. Design of PCR primers and gene probes for the general detection of bacterial populations capable of degrading aromatic compounds via catechol cleavage pathways. *Journal of Bioscience and Bioengineering*, 88: 542-550. doi: 10.1016/S1389-1723(00)87673-2.
- WANG, W. P., WANG, L. P. & SHAO, Z. Z. 2010. Diversity and Abundance of Oil-Degrading Bacteria and Alkane Hydroxylase (alkB) Genes in the Subtropical Seawater of Xiamen Island. *Microbial Ecology*, 60, 429-439.

7. A COMPARISON OF ECOGENOMIC TOOLS FOR ASSESSING BIODIVERSITY IN THE GREAT AUSTRALIAN BIGHT

Andrew Oxley¹, Jodie van de Kamp², Paul Greenfield³, Sarah Stephenson³, Jason E. Tanner¹.

1. SARDI Aquatic Sciences West Beach, SA 5024

2. CSIRO Oceans and Atmosphere Hobart, TAS 7001

3. CSIRO Oceans and Atmosphere North Ryde, NSW 2113

7.1 Executive summary

To complement standard morphological approaches to documenting benthic biodiversity in the Great Australian Bight (GAB), we employed two molecular metabarcoding approaches, using the nuclear 18S rRNA gene and the mitochondrial gene-encoding cytochrome c oxidase subunit I (COI). While the former has been the traditional target for barcoding studies, there has been a recent shift to the latter, which is suggested to have better species-level resolution, although due to its recent adoption it suffers from having fewer species sequences available for identification. We analyse sediment samples from the 2013 *RV Southern Surveyor* cruise using both genes, and for the 2015 *RV Investigator* cruise using only the 18S gene. For both genes, approximately 1000 operational taxonomic units (OTUs - putative species) were identified, although the majority of these could not be assigned species level taxonomy (86.5% and 95.5%, respectively). Only seven of 303 taxa from the GAB that were collected on the 2015 voyage and individually barcoded were identified in the 65 sediment samples that were successfully analysed for the COI gene. Both gene regions produced similar results in terms of overall patterns in assemblage structure, with a clear depth gradient, although COI produced a tighter clustering of samples at each depth sampled than 18S, possibly as a result of apparent undersampling of the number of taxa present that are amenable to detection using 18S. Despite showing similar patterns in assemblage structure, there were clear differences between the gene regions in what taxa were detected, with 18S showing higher numbers of nematode and platyhelminth OTUs, while COI showed higher numbers of annelids, arthropods and molluscs.

7.2 Introduction

Molecular (metabarcoding/ecogenomic) based analyses are rapidly evolving tools which offer the ability to improve our understanding of the diversity of marine ecosystems (Leray and Knowlton, 2016). Being highly sensitive and cost-effective, these technologies have the capacity to provide detailed information relating to the taxonomic diversity/richness of a sample without the need for traditional sorting and morphological taxonomic identification, which can be both time intensive and costly. On the flip side, these technologies are only as good as the reference databases used to put a taxonomic identity to the gene sequences obtained, which, in turn, are reliant on the provision of species which have been taxonomically verified (as vouchered specimens) by professional marine taxonomists and the deposition of their associated genetic sequences within public repositories. Whilst most next generation sequencing (NGS) surveys to date have implemented the nuclear 18S rRNA gene as the marker of choice, recent consensus has shifted towards the adoption of alternative markers like the mitochondrial gene-encoding cytochrome c oxidase subunit I (COI) (Bucklin et al. 2011), which has been reported, in some cases, to provide higher species-level resolution and more accurate estimates of diversity (Leray and Knowlton, 2016; Tang et al. 2012). However, despite international efforts like the “Barcode of Life” project (Costa and Carvalho 2010), there is still considerable under-representation of many species (particularly from marine benthic environments; Leray and Knowlton 2015) in the public sequence repositories, compromising the utility of COI diversity surveys. In line with this, both the 18S rRNA and the COI genes were used to profile the assemblage associated with benthic sediments from the Great Australian Bight (GAB), thus providing a more complete picture of species diversity from these systems.

7.2.1 Objectives

We investigate the diversity of life associated with benthic sediments from the GAB using an ecogenomics-based approach. High throughput sequencing (NGS) assays targeting hypervariable region 4 of the 18S rRNA gene (18S-V4) and the 5' region of the COI gene (COI-5P) (as a metabarcoding assay) were used to capture the likely array of species occupying these niches at a range of different depths. This study represents one of the first to implement a metabarcoding approach for evaluating the benthic diversity of deep sea sediments, and highlights the sensitivity of the method over traditional approaches.

7.3 Methods

7.3.1 Sample collection and DNA extraction

Benthic sediment samples were collected at depth stratified stations (200, 400, 1000, 1500, 2000 & 2800 m) along 5 north-south transects within the GAB (Figure 6.1) using a 6-core multicorer from KC (Denmark) incorporated into an instrumented coring platform (ICP) that could be controlled from the vessel and allowed reliable sample collection (Sherlock et al. 2014). A total of 75 samples were collected from 25 stations in 2013 aboard the *RV Southern Surveyor* cruise SS2013_C02, 3-22 April 2013 (over the depth range 200-2000m), and 24 samples in 2015 from the *RV Investigator*, which included samples from an additional depth (2800 m), and several repeat stations from 2013 (Appendix 7.1). Three cores from each deployment were subsampled for ecogenomics using 30 mm diameter minicores, with the top 2 cm from each minicore extruded into a DNA free tube and frozen at -20 °C prior to DNA extraction at CSIRO. A fourth core from each deployment was sampled for sediment grain size and carbon content (see Williams et al. 2017).

Ten grams of sediment were used for each DNA extraction. DNA was extracted using the PowerMax® Soil DNA Isolation Kit (Mo Bio Laboratories Inc, USA), modified as follows: 10 min incubation at 70 °C after adding the lysis solution (C1), and extending the incubation times to 30 min. Once DNA was eluted, the sample was concentrated to dry pellets in a “speed vac” vacuum concentrator, and then washed in

100% ethanol to remove excess salts. The quality and quantity of all DNA was checked using a NanoDrop™ 8000 Spectrophotometer (Thermo Scientific™). DNA was aliquoted into multiple plates, vacuum dried and stored at -20 °C prior to use in downstream assays.

7.3.2 Deep-sequencing of benthic sediment samples

18S rRNA gene amplification and sequencing

The V4 hypervariable region has several advantages for metabarcoding surveys of eukaryotic diversity (as summarised by de Vargas et al. 2015) including it's universally conserved length and simple secondary structure, the inclusion of both stable and highly variable regions, nucleotide positions allowing discrimination of taxa, and it is extensively represented in public reference databases allowing taxonomic assignment for known lineages. In addition, this metabarcoding approach has been adopted by both local Australian (e.g. Marine Microbes Project) and global (e.g. TARA Oceans) biodiversity initiatives.

The universal eukaryotic primers TAREukF1 (CCAGCASCYGC GGTAATTCC) and TAREukREV3 (ACTTTCGTTCTTGATYRATGA) developed by Amaral-Zettler et al. (2009) were used to amplify the V4 hypervariable region of the 18S rRNA gene from a total of 99 samples (75 collected in 2013 and 24 collected in 2015 as detailed previously). The size of the rDNA PCR products ranged between 350-380bp. The products 5' and 3' were labelled with unique 10-nucleotide tags during the PCR amplification, with a total of 54 unique tags for the total sample list including the positive controls, split across two separate libraries. Amplification reactions were performed in 50 µl reactions including 2 µl of template DNA using the AmpliTaq Gold® 360 Master Mix (Applied Biosystems™) standard PCR protocol. The PCR's were performed on an Eppendorf Mastercycler® (Eppendorf™) thermocycler. The following thermocycler conditions were applied to the reaction; 95 °C for 3 min (1 cycle), followed by 95 °C for 30 s, annealing at 58 °C for 30 s, extension at 72 °C for 60 s (35 cycles); then 72 °C for 7 min (1 cycle) and a final hold at 4 °C. Each PCR included 3 positive controls and one negative water control. The PCR success and integrity was examined by running a MultiNA digital gel (Shimadzu™, Japan).

Subsequent to amplification, PCR products were purified using an AMPure XP® PCR purification system (Agencourt Biosciences, Beverly, MA, USA). Amplification and purification success was interrogated on a MultiNA digital gel (Shimadzu™, Japan). The amplicon concentrations were measured using a NanoDrop™ 8000 Spectrophotometer (Thermo Scientific™, Waltham, MA USA). Nucleic acid concentrations for each sample were measured in preparation for pooling the final two libraries. In preparation for sequencing, the labelled products were mixed in equimolar concentrations, with a final purification performed using AMPure XP®. The two pooled libraries of 54 samples were prepared with a 2x 250bp Illumina Tru-seq library preparation kit and were sequenced across a single Illumina Miseq lane. In addition to the sediment samples, one internal reference sample containing mussel (*Mytillus*) DNA was processed in three sample replicates as a positive control for each library, with a total of six positive *Mytillus* controls run. The samples were split across two amplicon libraries for Miseq preparation. The Illumina Miseq sequencing was performed by the Ramaciotti centre, UNSW, and raw fastq files were obtained through the Illumina basespace program. Five samples were considered to have failed (< 10 reads per sample) and were removed from the dataset from the outset.

COI-5P amplification and sequencing

To elucidate the taxonomic composition of the benthic sediment DNA extracts, an Illumina deep-sequencing approach previously established for multiplexing hundreds of samples for evaluating bacterial community diversity (Camarinha-Silva et al. 2014; Chaves-Moreno et al. 2015) was adapted here for use with the universal COI-5P gene region primers mCOLintF and jgHCO2198, which were

designed for targeting a broad range of marine taxa including, among others, Arthropoda, Mollusca, Cnidaria, Annelida, Chordata, Echinodermata, Bryozoa and Sipuncula (Leray et al., 2013). This procedure implements a multi-step PCR-based strategy, whereby sample-specific barcodes and Illumina specific indices, adaptors and multiplex sequencing primer regions are integrated over consecutive cycles of PCR following pre-enrichment of the target region (in this case COI-5P). Reactions were performed according to methods detailed by Camarinha-Silva et al. (2014) for generating Illumina MiSeq ready libraries, where each PCR contained 2.5 µl of 10x TaKaRa Hot Start buffer (Clontech), 25 µM of each dNTP, 1 µM of each primer, 1 U of TaKaRa Taq Hot Start DNA polymerase (Clontech), ~50-100 ng of DNA template, made up to 25 µl with PCR grade H₂O (Qiagen). All COI-5P reactions were subjected to a “touchdown” PCR profile as recommended by Leray et al. (2013), which consisted of an initial denaturation at 94 °C for 3 min followed by 6 cycles comprising denaturation for 30 s at 94 °C, annealing for 30 s at 54 °C (-1 °C per cycle) and extension for 30 s at 72 °C followed by 29 cycles at 48 °C annealing temperature and a final elongation step at 72 °C for 5 min. As a negative (no template) control, PCR grade water was run alongside the other reactions, with the resultant PCR products visualised by agarose gel electrophoresis.

A total of 75 benthic sediment DNA extracts obtained from the initial survey in 2013 were evaluated using the COI-5P assays, of which 65 samples produced libraries which could be sequenced. Libraries were subsequently purified using Agencourt AMPure XP beads (Beckman Coulter), quantified using the Quant-iT™ Picogreen® dsDNA kit (Life Technologies) and pooled in equimolar ratios before being sequenced on the MiSeq platform (Illumina, San Diego, CA) using 250 nt paired-end sequencing chemistry through the Australian Genome Research Facility (AGRF). Amplicons generated from DNA from a single fish species (Jack Mackerel, *Trachurus declivis*) were sequenced alongside the samples as a control.

7.3.3 Bioinformatics sequence data processing

18S-V4

The raw Illumina data was processed using an in-house custom pipeline, GHAP, based on the USearch tool set (Edgar 2013). The GHAP hybrid pipeline takes files of reads and produces tables of classified OTUs and their associated read counts across all samples. The raw MiSeq Illumina data was de-multiplexed, de-replicated and then clustered at 97% similarity to generate OTUs. The median sequence from each OTU was then classified by BLASTing it against a curated set of 18S reference sequences derived from the SILVA v128 SSU reference set. The pipeline then mapped the starting reads back onto these OTU representative sequences to get accurate read counts for each OTU/sample pairing, and then generated OTU tables in both text and .biom (v1) formats, complete with taxonomic classifications and species assignments. OTU summaries were also produced at various taxonomic levels.

COI-5P

The 65 COI-5P sequenced samples from 2013 returned ~2.1 million raw paired-end reads respectively. The sequence reads from these samples were processed using the same pipeline as was used for 18S (described above). The classification of the resultant OTUs was done by BLASTing the median sequence for each OTU against a reference set of COI sequences. The initial set of COI sequences was downloaded from GenBank on February 6th, 2017, and included sequences from the Barcode of Life Data collection (BOLD, Ratnasingham and Hebert 2007). This set of sequences was initially dereplicated by removing identical sequences, and further reduced clustering at 99% identity. This two-step dereplication reduced the size of the reference set from 1,453,855 to 446,560 sequences, allowing the OTU matching phase of the pipeline to run in a few hours rather than several days. The dereplicated sequences were supplemented by a set of 304 sequences derived from samples collected in the GAB (see chapter 3).

For both gene regions, the taxonomic lineage assigned to each OTU was based on that of its closest match in the reference set. The most specific taxonomic rank was assigned using cut-offs based on the identity value returned from the match. OTU sequences that showed >98.35% identity to their closest reference sequence were assigned to the same Species, between 89.64-98.34% to the same Genus, 87.10-89.63% to the same Family, 85.01-87.09% to the same Order, 79.01-85% to the same Class and <79% to the same Phylum. These taxonomic cut-off values are the same ones used in chapter 3.

7.3.4 Statistical analyses

A data matrix comprising the sequence counts of 6,393/18,280 OTUs (18S-V4 and COI-5P respectively) across the 94/65 benthic sediment samples from different depths was first used to generate rarefaction curves. Rarefaction analysis was conducted to evaluate sequencing depth and subsequent coverage for each gene target. The resultant data matrices were filtered to only include those OTUs that contributed >0.01 % to the total 1,126,113/2,053,400 reads generated. The filtered data matrices comprising the percent standardised fourth-root transformed abundances of 1106/1033 OTUs were used to explore for patterns between sample types (i.e. depths), where samples were ordinated using principal coordinate analysis (PCO) after constructing a sample-similarity matrix using the Bray-Curtis algorithm (Bray and Curtis, 1957). Superimposing the relative abundance of a select few OTUs onto the ordination plot as bubbles was used to explore patchiness within/across depths. Differences in the profiles between depth categories and transects for each gene target were evaluated using Permutational Multivariate Analysis of Variance (PERMANOVA), with type III (partial) sums of squares, p-values generated using 9999 permutations of residuals under a reduced model (Anderson, 2001) and $\alpha=0.05$, followed by pairwise tests where appropriate. Differences in OTU dominance across the 94/65 samples were explored using k-dominance plots, by plotting cumulative ranked abundances against OTU rank. Environmental variables were correlated to community composition data using distance-based linear modelling (DISTLM) with distance-based redundancy analysis (dbRDA), and Pearson correlations with environmental variables were also overlaid on the principal co-ordinates plots. For the DISTLM, the overall best model was selected from all environmental procedures using the BEST procedure and the small sample size variation of AIC (AIC_c) with 9999 permutations. All analyses involving environmental variables were conducted at the station level, as environmental data corresponding to individual cores were not collected. Multivariate and rarefaction/dominance analyses were performed in PRIMER (v.7.0.11) PRIMER-E, Plymouth Marine Laboratory, UK, (Clarke et al., 2001).

Comparing COI-5P and 18S datasets

In order to compare the pattern generated from both the 18S and COI-5P assays, datasets were filtered to only include matched samples, leaving a total of 61 samples with sufficient read counts from 2013 in the analysis. Both datasets were further filtered to only include those OTUs that contributed >0.01 % to the total reads in each dataset. Accordingly, the 18S dataset comprising 990 OTUs was compared to the COI-5P dataset comprising 1026 OTUs from the same 61 samples. First, the OTUs were aggregated to their respective phyla and counted so that the number of OTUs from each phylum could be directly compared. Secondly, a new PCO ordination plot was generated for each dataset based on just the matched samples. The pattern of multivariate similarity in each data set was formally compared by a Mantel-like test using the Spearman Rank correlation method (RELATE routine in PRIMER, with 9999 permutations).

7.4 Results and Discussion

Deep-sequencing of both the 18S-V4 rRNA and the 5' region of the cytochrome c oxidase subunit (COI-5P) genes was performed for evaluating the eukaryotic assemblage in a total of 75 sediment cores collected over a range of depths (from ~200-2000 m) within the GAB, with an additional 24 cores only sequenced for 18S. As with all sampling methods (Kingsford and Battershill, 1998) biases may be introduced into the data depending upon the gene marker, target region, or primers used in the molecular assays (Coward et al., 2015; Elbrecht and Leese, 2015; Leray and Knowlton, 2016). For example, it has been reported that certain regions of the 18S rRNA gene may preferentially amplify Protists, whilst the COI gene may under-represent these organisms yet select for Arthropods (for a review see Leray and Knowlton, 2016). Thus the use of two different targets provides a broader coverage of taxa than might be obtained from a single target. This is especially pertinent given the recent adoption of COI (over the 18S gene) as a universal marker, and the likely under-representation of COI sequences within the public data repositories used for inferring taxonomy, especially for unique or largely underexplored ecosystems like the GAB, although 18S sequences for species at these depths are also sparse.

7.4.1 18S rRNA gene based estimates of benthic diversity

Metabarcoding of the V4 region of the 18S rRNA gene was undertaken for all 94 samples resulting in a total of 1,126,113 sequences obtained post QC and merging of paired reads. Sequencing depth exhibited a broad range across the 94 libraries (546 – 40,700 reads, average = $11,980 \pm 7,293$). Clustering of sequences at $\geq 97\%$ similarity produced 6,393 OTUs with a range of 144 – 834 OTUs per sample. Further filtering of the dataset to include only those OTUs with an abundance of $> 0.01\%$ and those samples with > 600 reads per sample reduced the dataset to 1,005,222 reads across 93 samples (2,262 – 36,755 reads, average = $10,890 \pm 6,624$ per sample) distributed across 1,106 OTUs (Richness of 97 – 470 OTUs per sample). Rarefaction analysis demonstrated that samples were largely under-sampled (Figure 7.1). While the most abundant eukaryotes are probably represented in the dataset for most samples, it is likely we fail to identify less abundant members of the community, thus this would not be a “true” representation of the diversity in all samples. The poor coverage demonstrated in these results does not reflect a general failure of using the 18S-V4 metabarcoding approach to assess diversity, but is rather due to technical issues with the specific batch of primers and/or sample processing.

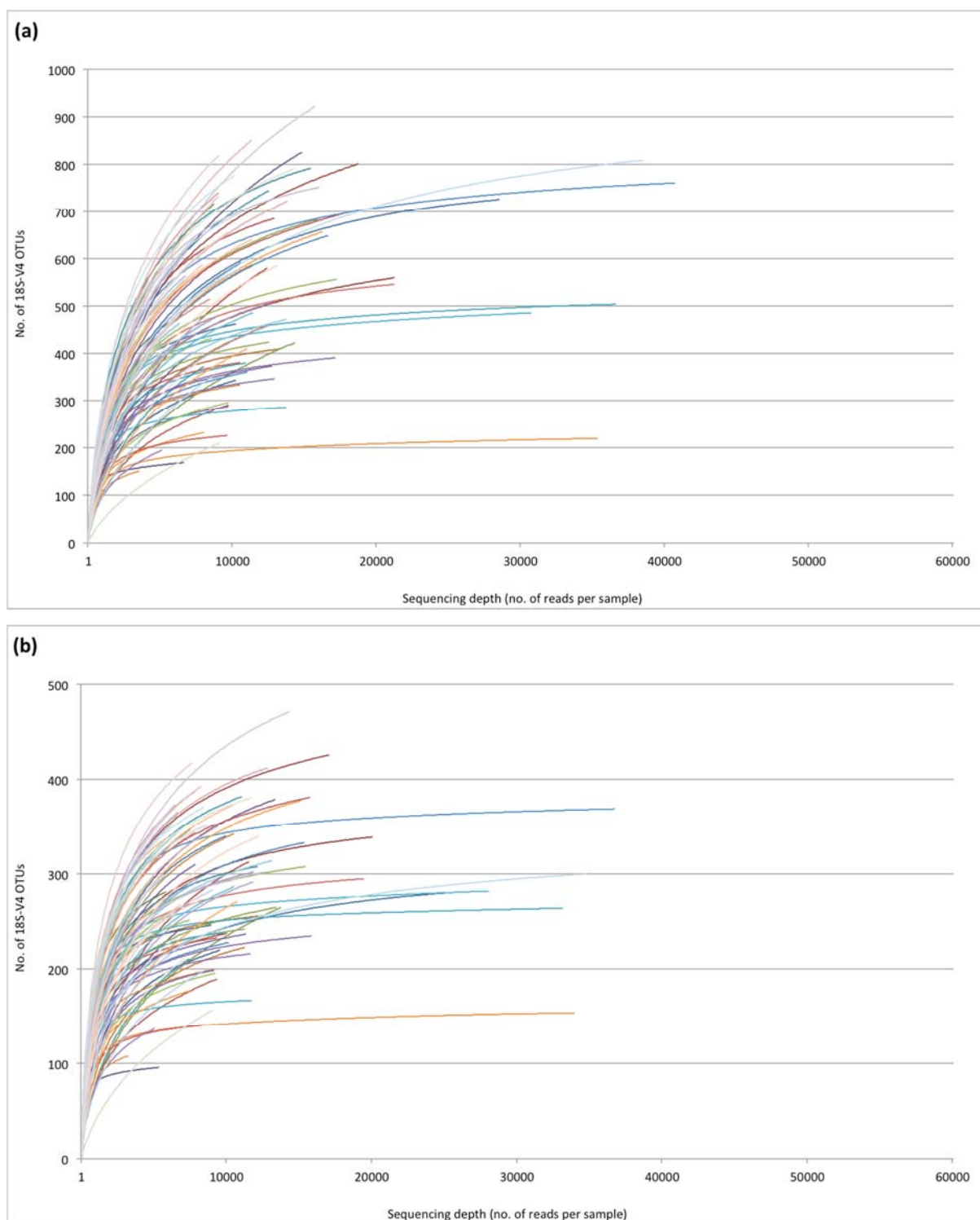


Figure 7.1. Rarefaction curves depicting the number of observed 18S-V4 OTUs obtained from 94 benthic surface sediment samples. Charts plot the number of observed OTUs against the number of reads obtained from (a) pre-filtered and (b) filtered ($\geq 0.01\%$ abundance) data.

Taxonomic assignments of the resultant OTUs were inferred from homology with reference sequences sourced from the SILVA v128 SSU sequence set (Figure 7.2), with classifications coming from the NCBI Taxonomy (<https://www.ncbi.nlm.nih.gov/taxonomy>). Of the 1,106 OTUs identified, ~84.4% could be assigned to some taxonomic level within the Kingdom Eukaryota, with the largest contribution belonging to the phyla Nematoda (15.8%, 175 OTUs), Dinoflagellata (12.7%, 141 OTUs), Ciliophora (8.2%, 91 OTUs), Annelida (4.9%, 54 OTUs), Platyhelminthes (4.8%, 53 OTUs), Sagenista (4.2%, 46

OTUs), Apicomplexa (3.8%, 42 OTUs), Perkinsozoa (3.8%, 42 OTUs) and Arthropoda (3.3%, 37 OTUs). Those that could not be classified at the phylum level, unclassified Eukaryota and unclassified Fungi, accounted for 4.4% (49 OTUs) and 0.8% (9 OTUs), respectively. A further 15.6% of OTUs (173 OTUs) could not be assigned at the Kingdom level, i.e. Eukaryota; 38.4% of OTUs were assigned at the genus level and a further 13.9% at the species level.

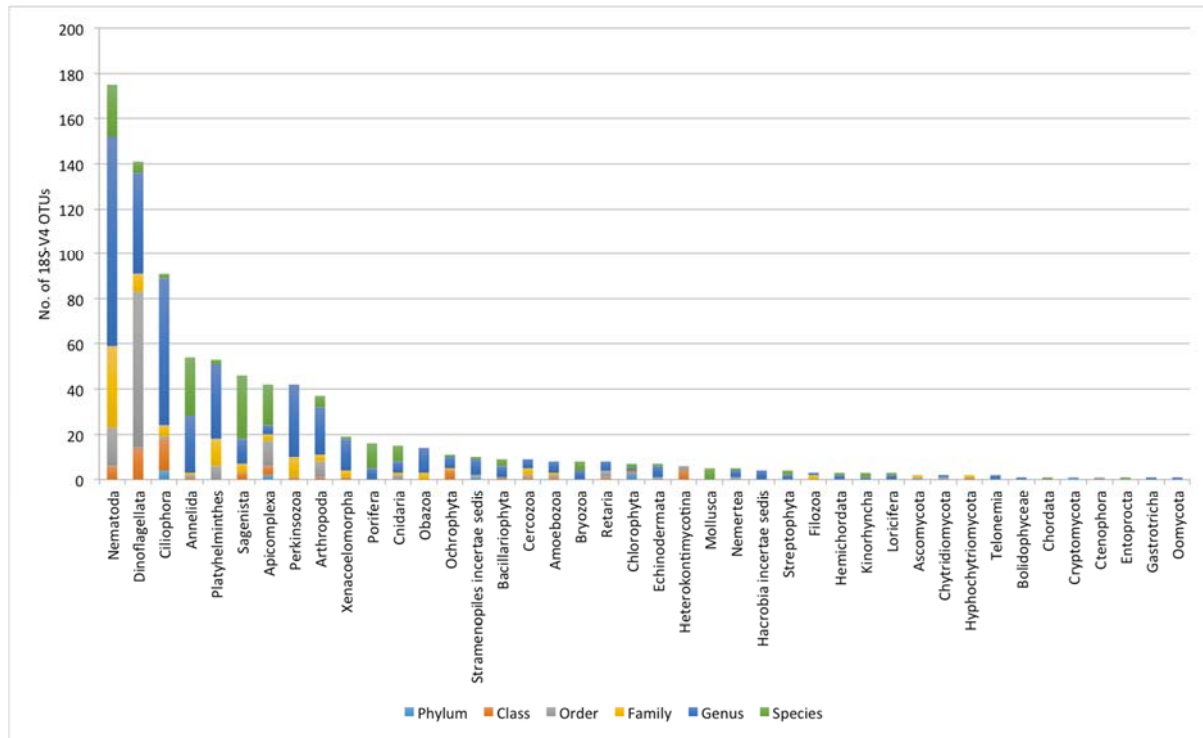


Figure 7.2. Numbers of 18S-V4 OTUs assigned at different taxonomic levels for major phyla following interrogation against a curated SILVA 18S database using the NCBI taxonomy.

The most abundant phyla overall were the Dinoflagellata (21.9%), Annelida (16.9%), Nematoda (11.1%), Ciliophora (5.9%), Platyhelminths (3.7%) and Apicomplexa (3%) (Figure 7.3). Other phyla contributing between 1-3% to the total community included, Porifera, Cnidaria, Bryozoa, Arthropoda, protists Sagenista and Perkinsozoa, Hemichordata, Xenacoelomorpha and Chlorophyta (green algae). Analysis of the relative abundances of the phyla across depth categories (from 200-3000 m) revealed differential patterns of these phyla. In the shallow 200 m depths, Dinoflagellata (26.3%), Porifera (10.9%), Annelida (10.4%), Bryozoa (9.9%) and Nematoda (8.3%) dominated. At 400 m, the Annelida were the most abundant phyla (34.2%) along with the Dinoflagellata (26.6%) and Nematoda (7.9%). From 1000 m and deeper, the Dinoflagellata (15.4-24.0%), Annelida (11.7-22.1%) and Nematoda (7.9-19.9%) dominated. From 400 m and deeper, the Porifera and Bryozoa were only minor representatives of the community (< 0.6%) or absent. Other depth specific features include the increased numbers of Nematoda at 1500 and 2000 m (17.2-19.9%) compared to other depths (7.8-8.3%); the abundance of Ciliophora at 2800 m (12.3%) compared to all other depths (1.9-6.4%); and Mollusca at 400 and 1000 m (2.6-4%) compared to all other depths (0.02—0.09%). The unclassified taxa were represented at all depths (2.6-16.2%) with the largest proportion being found at 1000 m.

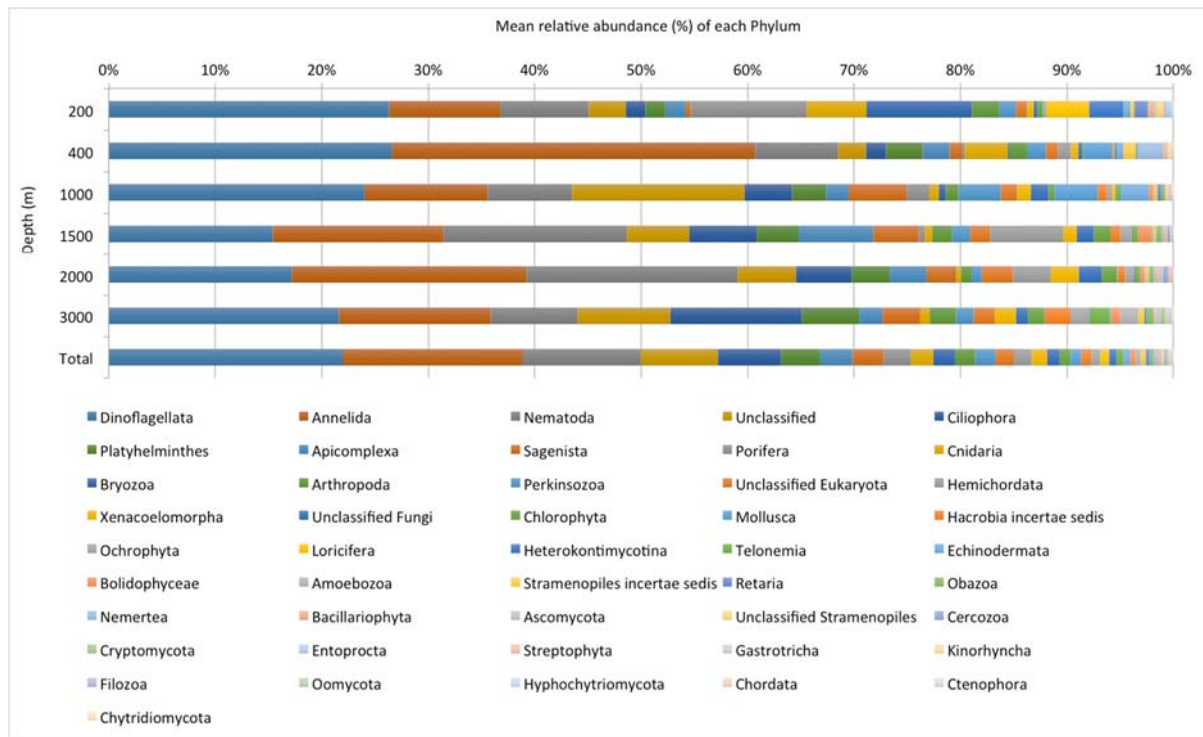


Figure 7.3. Mean relative abundances of OTUs at the phylum level across depths within the GAB based on analysis of the 18S-V4 gene region.

Beta diversity was visualised with principle coordinate analysis (PCO) illustrating depth related changes in eukaryote community structure as determined by the 18S-V4 rRNA gene (Figure 7.4). PERMANOVA analysis demonstrated a small, but significant interaction between transect and depth (pseudo- $F_{19,92} = 1.6023$; $p = 0.0001$). Pairwise testing demonstrated that differences in community composition between transects was minimal and restricted to the middle depths of 1000 m (pseudo- $t = 1.2156$ - 1.5643 ; $p < 0.05$) and 1500 m (pseudo- $t = 1.2993$ - 1.4554 ; $p < 0.05$) on Transects 1 and 2. While a depth effect on community structure was detected (pseudo- $F_{5,92} = 10.137$; $p = 0.0001$) as demonstrated in Figure 7.4 (PCO), the depth that these changes were statistically evident ($pMC < 0.05$) differed with transect. There is a general trend towards changes in community structure between the shallower depths (200/400 m) and the deeper sites (1000 – 2800 m).

The depth trend varies with phylum (Figure 7.5), and was clearest in the composition of Dinoflagellata, Nematoda, Platyhelminthes and Arthropoda. While for the Annelida the two shallower depths were clustered, there seemed to be no discernible pattern in depths greater than 1000 m. Similarly, there were no visible trends in the distributions of Cnidaria and Mollusca.

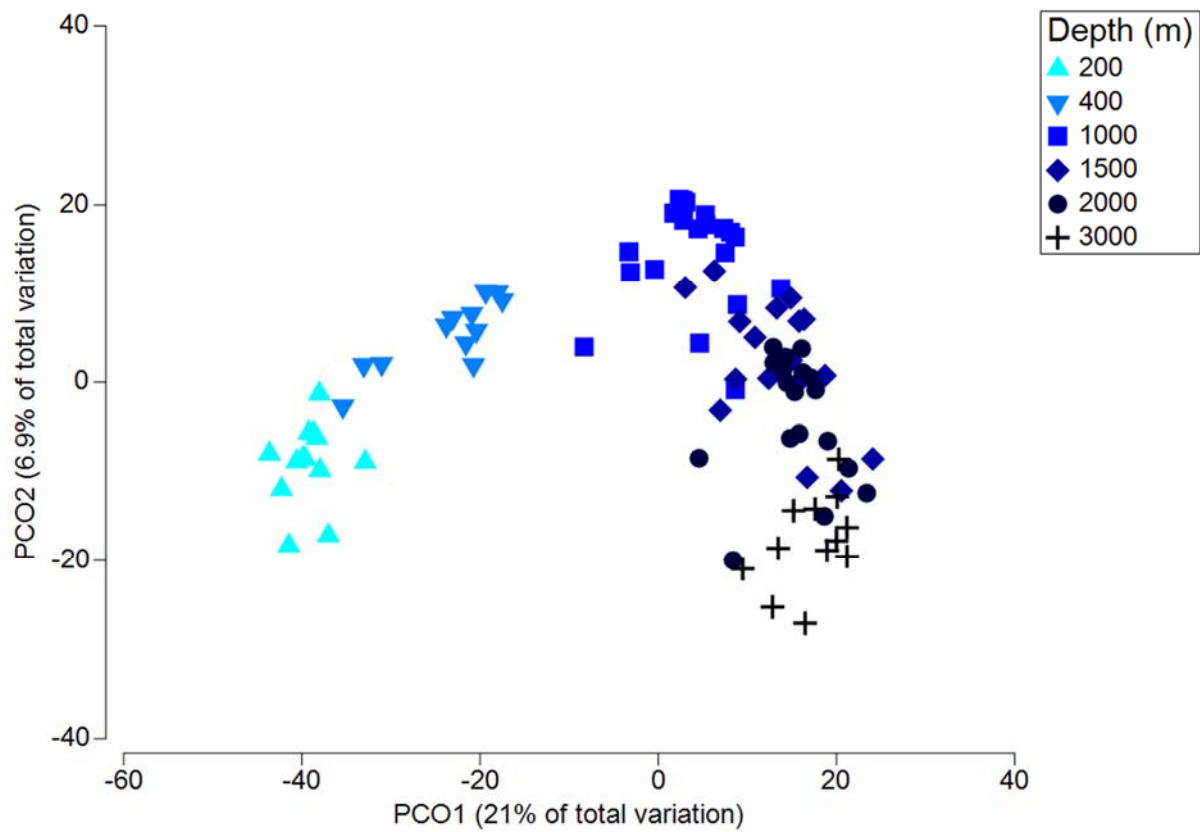


Figure 7.4. Ordination (PCO) plot of the global 18S-V4 OTU profiles from 93 benthic surface sediment samples collected over a range of depths from the GAB.

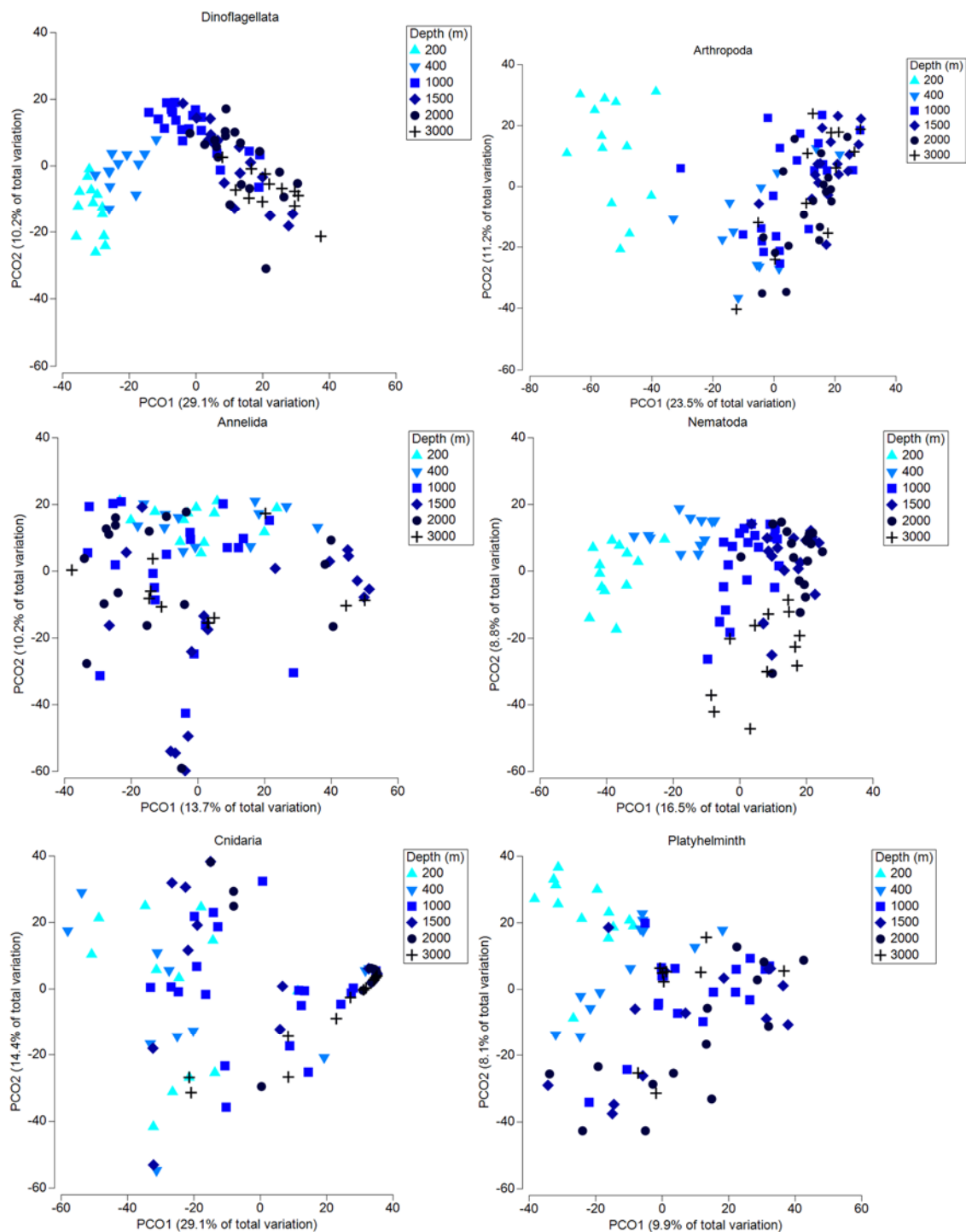


Figure 7.5. Ordination (PCO) plots of the relative abundances of 18S-V4 OTUs for major phyla from 93 benthic sediment samples collected over a range of depths in the GAB.

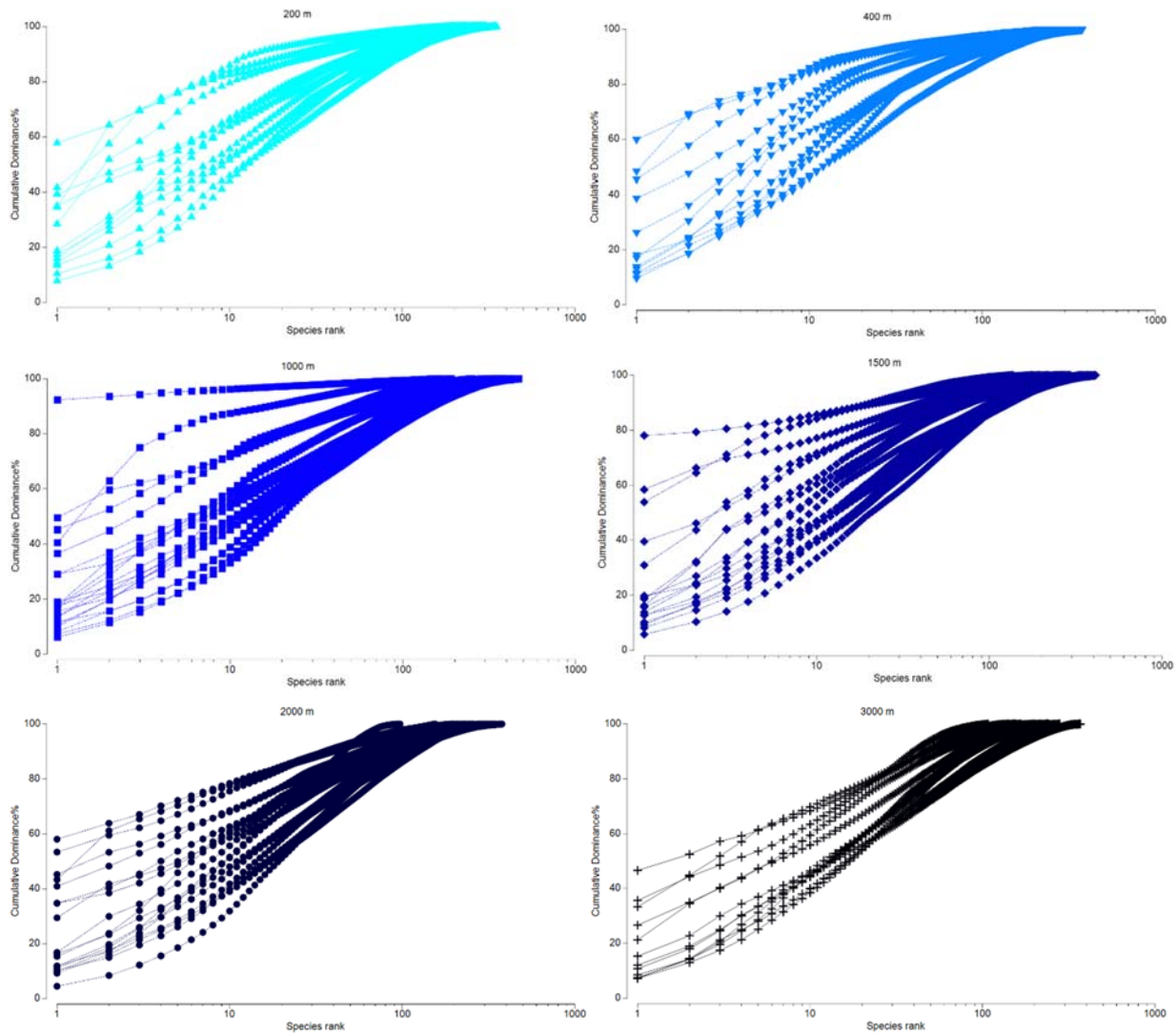


Figure 7.6. K-dominance plots representing the cumulative dominance of 18S-V4 OTUs across depths in the GAB.

Cumulative dominance plots demonstrated that at each depth communities varied from having one or more dominant taxa to many taxa with more uniform abundances (Figure 7.6), highlighting the patchiness of the benthic eukaryotic assemblages of the GAB, which is particularly evident for certain macrofauna as discussed previously (e.g. molluscs and sponges). The patchiness of eukaryotic assemblages as determined by 18S-V4 OTU distributions is further demonstrated in Figure 7.7. While samples cluster by depth, characteristic assemblages are less clearly visualized. There are only four OTUs that span the full range of depths. These are classified as Dinoflagellata (OTU_2 an uncultured *Karlodinium* sp.; and OTU_5 an uncultured *Amphidinium* sp.) and Nematoda (OTU_21 *Sabatieria pulchra*; and OTU_40 an unclassified Nematode of the class *Enoplea*). The uncultured *Amphidinium* sp OTU_5 was present in all samples at all depths.

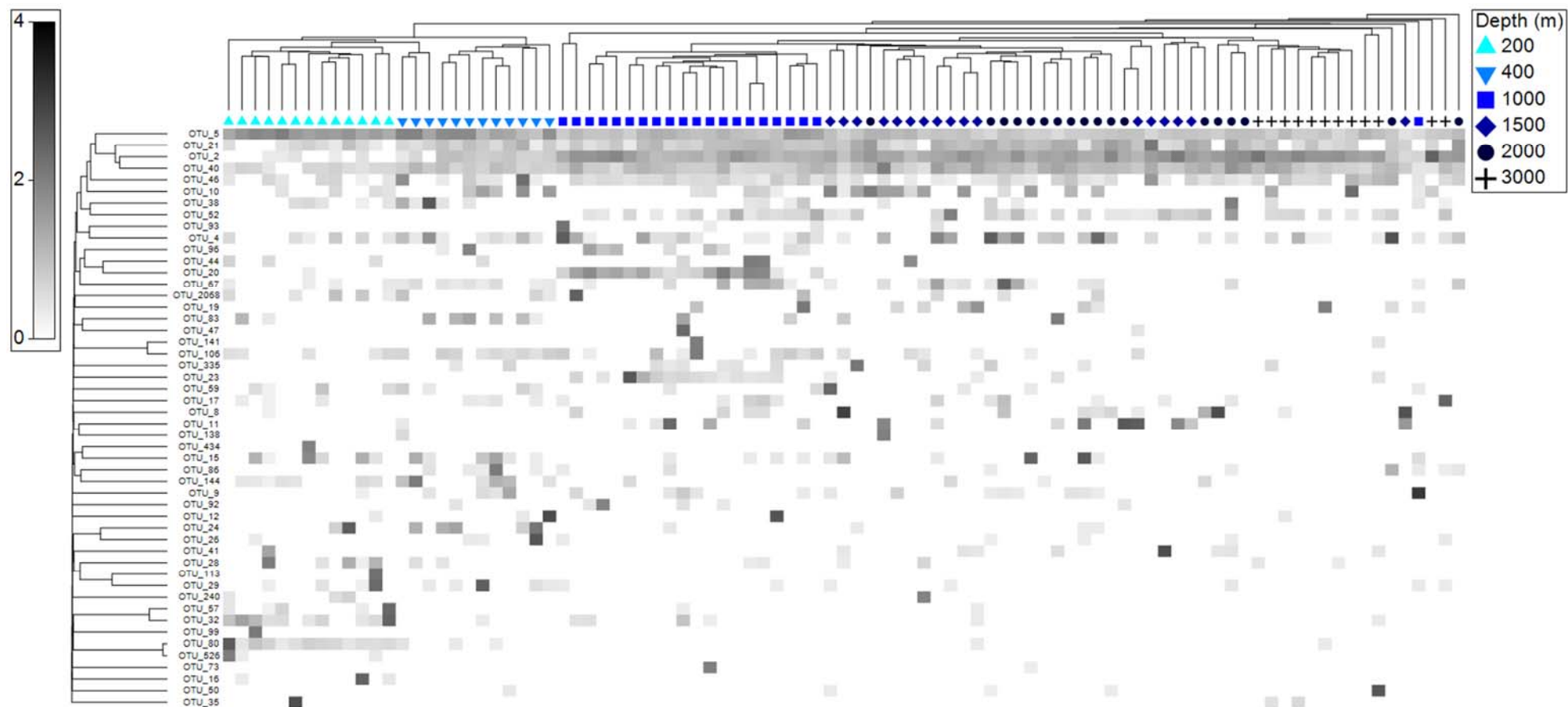


Figure 7.7. Shade plot representing the distribution of the 50 most important 18S-V4 OTUs. Stations have been grouped using Bray-Curtis similarities on the fourth root transformed data, by hierarchical agglomerative clustering with group average linking and SIMPROF (Type 1) tests. OTUs have been grouped using the Index of Association by hierarchical agglomerative clustering with group average linking and SIMPROF (Type 3) tests. White space denotes absence of an OTU at that Station; depth of the grey scale is a linear proportion of the fourth-root transformation of abundance data.

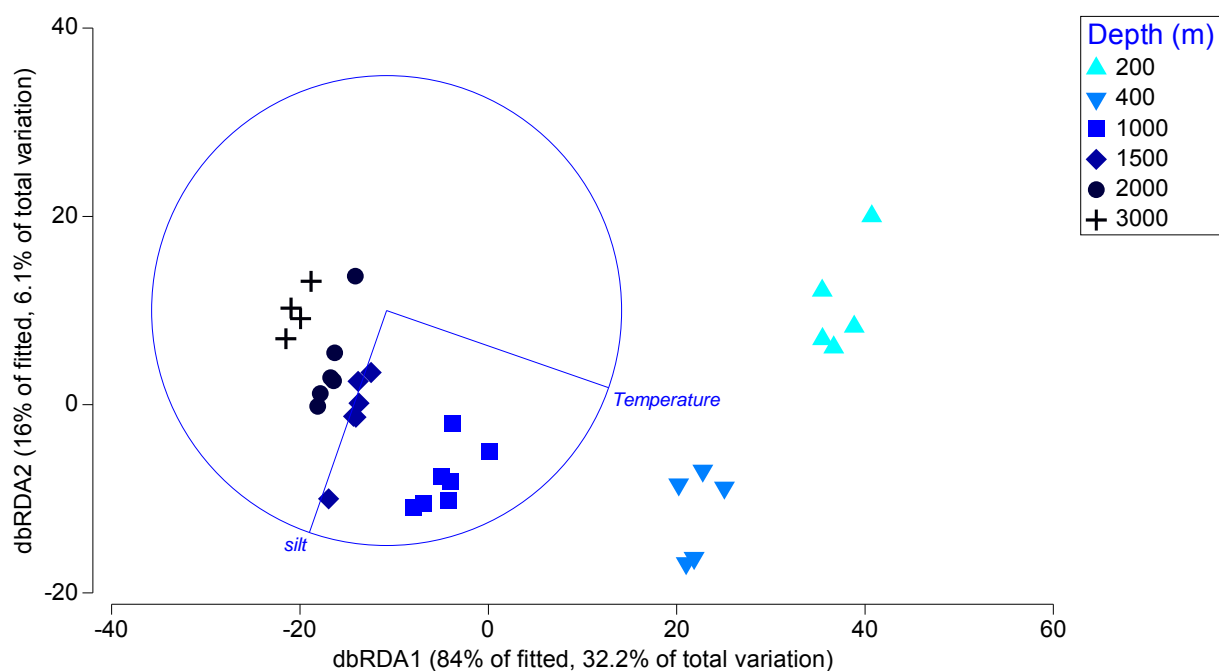


Figure 7.8. Relationships between 18S community composition and environmental parameters as determined by distance-based linear modelling and visualised with distance-based redundancy analysis. Only variables contributing to the final selected model are included. Temperature and depth were strongly collinear ($R > 0.9$), and thus temperature was retained while depth was removed from the analysis.

The relationship between 18S community composition and 10 environmental parameters (consisting of sediment grain size and composition parameters, depth and bottom temperature – see Williams et al. 2017) was determined via distance-based linear modelling (DistLM). Marginal tests indicated that significant ($p < 0.05$) amounts of variation in assemblage structure could be explained by all variables when taken individually except for the proportion of very coarse sand. The strongest predictors were temperature and proportion of clay, explaining 31.9% and 26.4% of variation respectively (pseudo- $F > 11.1$; $p < 0.001$). The BEST selection procedure was used to determine which combination of variables explains the most variation in community structure. The most parsimonious model only consisted of silt and temperature (Figure 7.8) but was only supported weakly ($R^2 = 0.38$).

7.4.2 COI estimates of benthic diversity

In parallel to the 18S rRNA gene assay, a metabarcoding approach for evaluating both the macro- and micro-faunal diversity of the GAB sediment cores was developed using universal marine metazoan COI gene primers. From a total of 75 samples collected in 2013 over a range of depths (from ~200-2000 m), 65 yielded libraries which could be sequenced on the Illumina platform (see Appendix 7.1), producing a total of ~2.1 million sequence reads ($\sim 31,590 \pm 10,427$ (SD) reads/sample); the remaining samples failing amplification due to the lack of DNA comprising COI-5P priming sites and/or the characteristically low, poor-quality DNA yields typical of benthic sediments (Forschner et al., 2008). Following clustering of the reads into OTUs (at an identity of 97%) and further filtering of the data to include only those OTUs occurring at an abundance of $>0.01\%$, a total of approximately 50-350 OTUs were detected per sample. Unlike the situation with 18S, rarefaction analysis shows that most samples reached an asymptote, with the number of OTUs detected not increasing further with increased sampling depth, revealing that the

depth of sequencing (i.e. 10,000 – 50,000 reads/sample) was sufficient to capture the expected diversity in these samples (Figure 7.9). This also contrasts with the traditional infaunal and epifaunal surveys conducted at the same stations, which indicated that substantially more taxa were present than were sampled, albeit based on among samples analyses rather than a within samples analysis as done here (Williams et al. 2017).

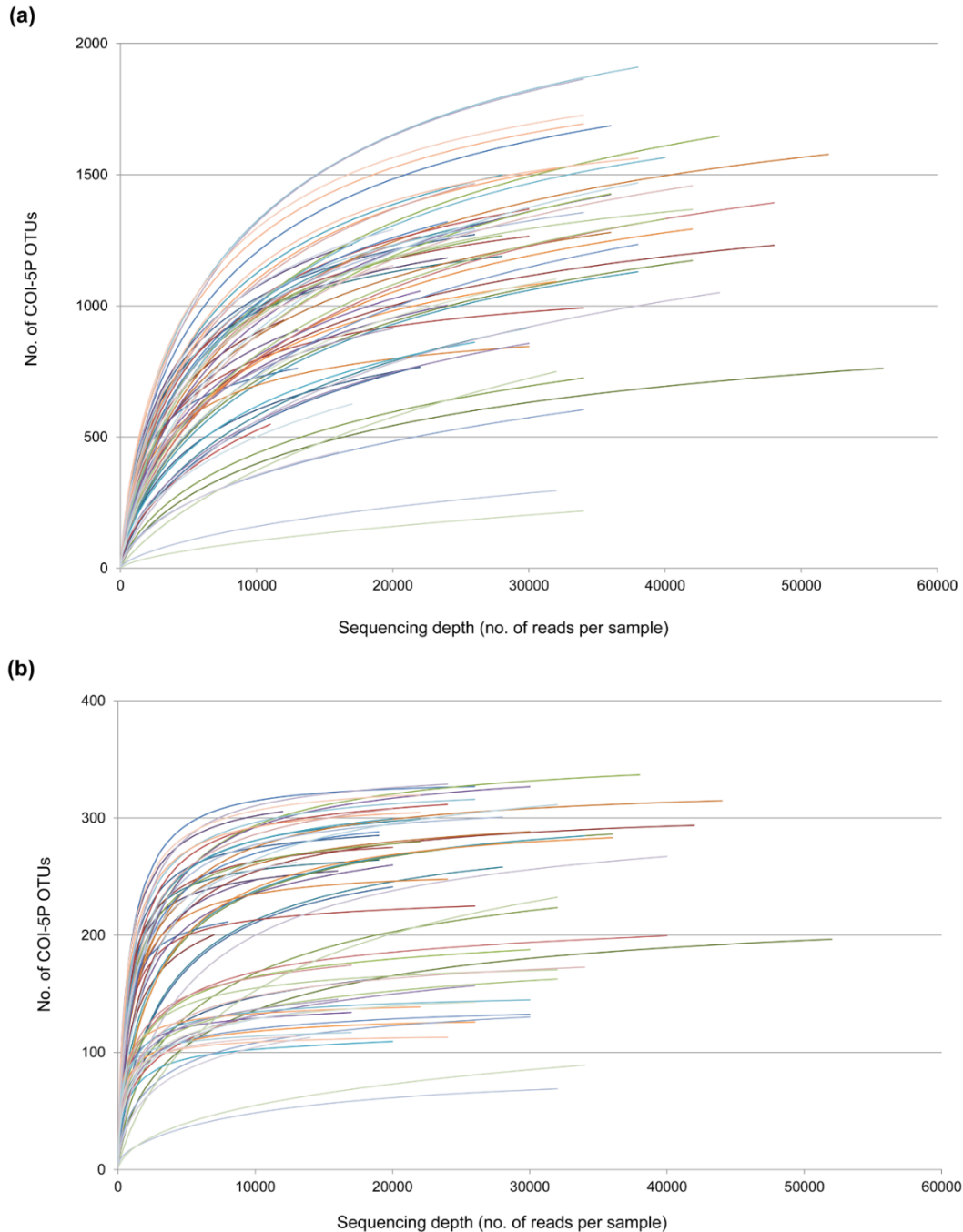


Figure 7.9. Rarefaction curves depicting the number of observed COI-5P OTUs obtained from the 65 benthic sediment samples obtained in 2013. Charts plot the number of observed OTUs against the number of reads obtained from (a) pre-filtered and (b) filtered ($\geq 0.01\%$ abundance) data.

Taxonomic assignments of the resultant OTUs were inferred from homology with reference to COI gene sequences downloaded from GenBank, including sequences from the BOLD systems database

(Ratnasingham and Hebert 2007), and supplemented by the set of sequences obtained in chapter 3. The closest species was reported at a cut-off of $\geq 98.35\%$ identity (in accordance with the procedures used for the COI barcoding of the individual GAB specimens, see Chapter 3). Those OTUs with an identity of $\leq 79\%$ were only able to be assigned at the phylum level. A total of 1,033 OTUs (or distinct taxa) were identified from the 65 samples, of which $\sim 65\%$ could be assigned to some taxonomic level, with the largest contribution belonging to Arthropoda (23.6%, 244 OTUs), Annelida (12.5%, 129 OTUs) and Nematoda (7.5%, 77 OTUs) (Figure 7.10). Other groups included Cnidaria, Bryozoa, Porifera, Mollusca, Echinoderms, Chordata, algae/diatoms (Chlorophyta, Phaeophyceae, Bacillariophyta), various marine worms (Nemertea, Platyhelminthes, Onychophora, Myzostomida, Entoprocta, Xenacoelomorpha), rotifers and Fungi (Basidiomycota). Findings from other benthic systems using similar molecular (metabarcoding) approaches have reported comparable levels of diversity (i.e. between ~ 1000 -2500 OTUs), with Arthropods, Annelids and/or Nematodes being the primary constituents (Leray and Knowlton, 2016; Leray and Knowlton, 2015; Cowart et al., 2015).

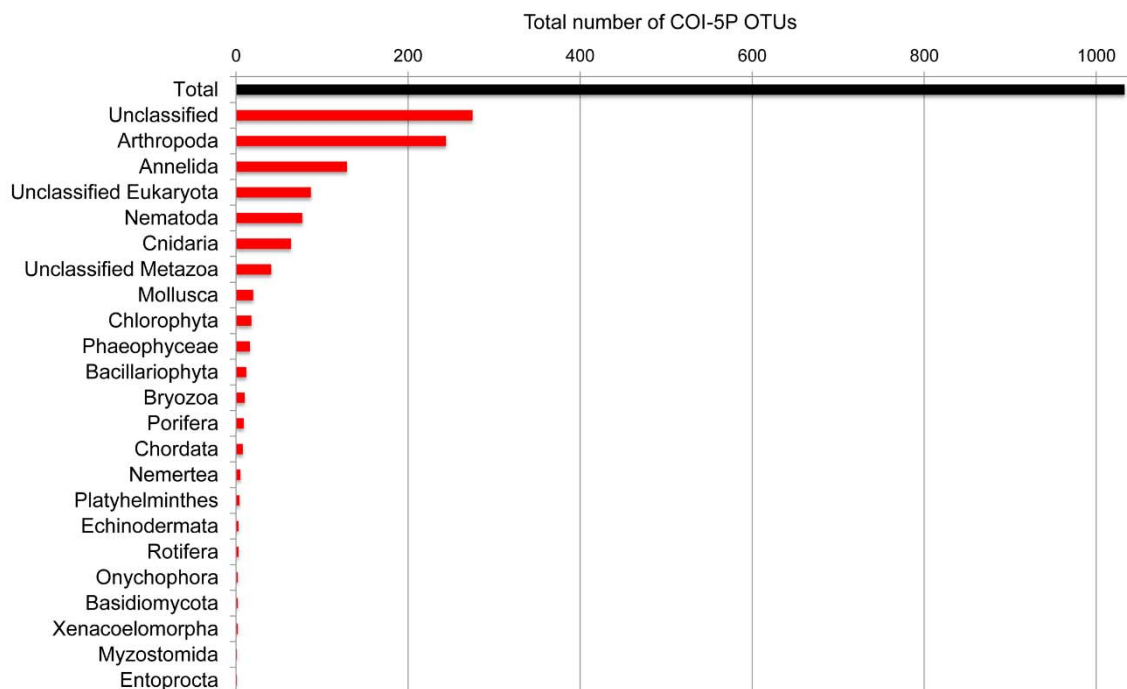


Figure 7.10. Total numbers of COI-5P OTUs assigned to major phyla based on interrogation of sequences against the BOLD systems database.

Overall, while species-specific assignments could be obtained for some OTUs (e.g. groups like the Mollusca, Arthropoda and Porifera, where identities of 99.6-100% were observed for certain species including *Dentallida* sp., *Ebalia nux*, and *Styocordyla chupachups*), the taxonomic resolution was generally quite low, with the majority ($\geq 60\%$) only assigned to the phylum level (Figure 7.11). In conjunction with the substantial proportion ($\sim 35\%$) of sequences which could not be assigned to any specific taxonomic group (Figure 7.10), this finding likely reflects the under-representation of sequences from vouchered specimens from these environments in the public repositories. Indeed, earlier studies have also reported the presence of large numbers (10-36%) of unclassified OTUs (Leray and Knowlton, 2015). Whilst this may, in part, reflect the uniqueness of the assemblages within the GAB, sequences from only 7 of the individual barcoded specimens obtained in Chapter 3 matched with COI-5P OTU

sequences obtained from the metabarcoding analysis conducted here: 1 Mollusc, 1 Arthropod, 2 Echinoderms and 3 Cnidaria (Table 7.1). This result was not entirely unexpected given the mismatch between the organisms that were successfully barcoded in Chapter 3 (primarily epifauna), and those targeted by the metabarcoding approach (primarily infauna). Any DNA detected from these epifaunal taxa is likely to be exogenous, and would indicate the presence of the animal nearby rather than actually in the sample. This result suggests that metabarcoding of sediment samples is not a good tool for assessing the epifaunal assemblage present in deep sea environments such as those present in the GAB. The patchiness of these environments would also be a contributing factor, and the small size and cryptic nature of many benthic species which may not be sampled as part of traditional infaunal surveys and which are likely to comprise the highest levels of diversity (Leray and Knowlton, 2015), means that even if infauna were successfully barcoded, they would be likely to make up a relatively small proportion of the OTUs detected by metabarcoding. Many of the infaunal taxa attempted for Chapter 3 also could not be successfully barcoded, and thus the GAB-specific barcodes are primarily for epifaunal taxa. Other contributing factors may include the introduction of artificial diversity through the formation of PCR artefacts or sequencing errors (Carugati et al., 2015; Creer et al. 2010), or the presence of extraneous DNA from organisms in the above water column which naturally accumulates within the deposits and which may be co-amplified in these PCR-based metabarcoding assays (Dell'Anno and Danovaro, 2005; Monrad et al., 2016). However, given that the COI-5P dataset was first filtered to retain only those OTUs with an abundance of >0.01% (thereby reducing the likelihood of the inclusion of such artefacts), the presence of extraneous DNA in these samples is more likely, particularly given the reasonable contribution here of OTUs from photosynthetic organisms (namely chlorophyta and phaeophyceae) (Figure 7.11) which are likely to have been transported from shallower waters. In part, this may be resolved in future studies by targeting the active communities by the sequencing of libraries prepared from RNA rather than DNA extracts (Guardiola et al., 2016; Monrad et al., 2016).

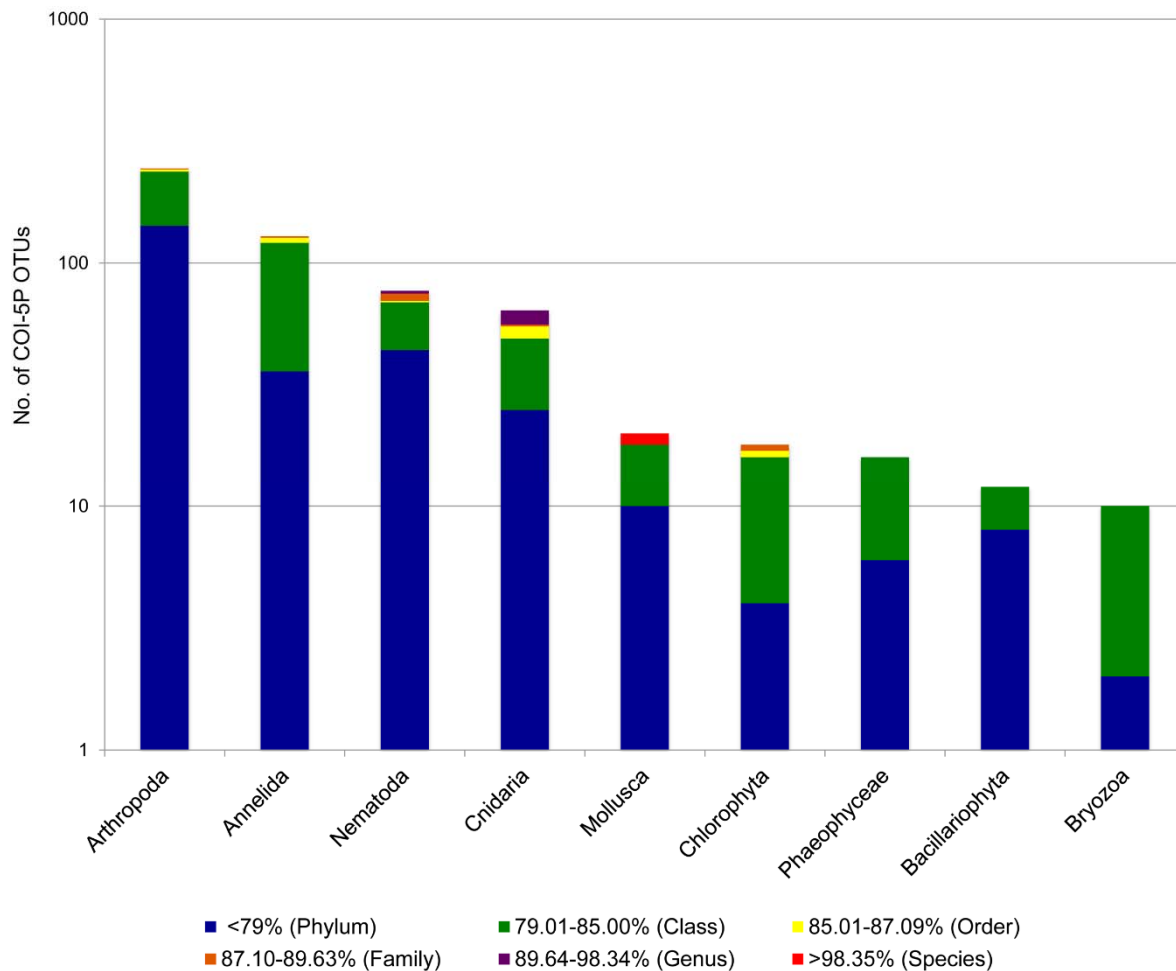


Figure 7.11. Numbers of COI-5P OTUs assigned at different taxonomic levels for major phyla following interrogation of the GenBank COI collection. The % sequence identity cut-off values used at each taxonomic level are indicated in the key.

The spatial distribution of samples that contained taxa barcoded in Chapter 3 is largely concordant with the distribution of those taxa in the beam trawls, as determined by Williams et al. (2017), with some interesting differences. The barcoded sample of Order Dentallida was collected at T5 1000 m, and members of this order were broadly distributed over all depths and transects. However, DNA of this taxon was only found in a single sediment sample from T1 1000 m. The *Ophiocten* samples collected were all *O. australis*, and were found on all transects at the 1500 and/or 2000 m stations, while DNA was only found in the two westernmost transects, although at the same depths. *Phormosoma* (identified as *P. cf bursarium*) was only collected at T4 1500 m, although unidentified members of the same family (Phormosomatidae) were collected at T1 1000 m, while the DNA sample came from T3 1000 m. *Collosendeis gigas* was not identified through morphological taxonomy, although a number of congeners were, with a broad distribution from 1000 – 3000 m, while the DNA samples were all from 200 m, although the sequence match in this case was only reliable to class level. Class Pycnogonida was present at 200 m, although rare, and was substantially more abundant at 1500-2000 m. The *Corallimorphus profundus* sequence matches are only reliable to phylum level, and cnidaria were broadly distributed across transects and depths, so nothing can be said about the match between beam trawl catches and sediment DNA.

Table 7.1. List of COI-5P OTUs which match the COI sequences obtained from individual GAB barcoded specimens. The total numbers of sequence reads are indicated and are given for each depth category.

COI-5P OTU no.	Total no. reads	Organism (closest hit)	Accession no.	% sequence identity	Total sequence reads per depth category (m)				
					200	400	1000	1500	2000
OTU_9	22252	Mollusca; Scaphopoda; Unclassified Dentaliida sp.	ABTC132340	100	0	0	22252	0	0
OTU_1755	113	Echinodermata; Ophiuroidea; Ophiurida; Ophiuridae; Ophiocten sp.	276_N112-01_Ophi2	100	0	0	0	7	106
OTU_2542	72	Echinodermata; Echinoidea; Echinothuroidea; Phomosomatidae; Phormosoma bursarium	159_N101-01_echi	100	0	0	72	0	0
OTU_2717	59	Arthropoda; Chelicerata; Pycnogonida; Pantopoda; Colossendeidae; Colossendeis gigas	207_N128-01_Arth2	79.7	59	0	0	0	0
OTU_13143	7	Cnidaria; Unclassified Cnidaria sp. (Corallimorphus profundus)	330_N133-01	77.1	0	0	0	7	0
OTU_10600	6	Cnidaria; Unclassified Cnidaria sp. (Corallimorphus profundus)	330_N133-01	78.7	6	0	0	0	0
OTU_17850	5	Cnidaria; Unclassified Cnidaria sp. (Corallimorphus profundus)	330_N133-01	76.9	0	0	0	5	0

Analysis of the relative abundances of the phyla across depth categories (from 200-2000 m) revealed the predominance of Annelids (24.1-43.5%), Arthropods (11-26.5%) and Unclassified taxa (23.7-30.3%) (i.e. those with an identity of < 79% with representative sequences in the BOLD systems database) at all depths (Figure 7.12). Notable depth specific features included higher abundances of Bryozoa (6.5%), Porifera (1.8%), Echinoderms (1.2%), Platyhelminthes (0.8%) and Nemertea (0.8%) in shallower waters (particularly at 200m), while greater depths comprised higher abundances of Nematodes (4.6-5.9%), Chordata (0.1-0.9%), Unclassified Eukaryota (4.9-6.6%) and Unclassified Metazoa (2.2-2.6%). In addition, Entoprocta were only detected at 200 m, while Basidiomycota, Xenacoelomorpha and Myzostomida were only detected at greater depths from 1000-2000 m, indicating the likely strong depth-specific niche associations of these taxa, a feature reported previously for some species (Rouse et al., 2016; Xu et al., 2014). The dominance of Bryozoa and Porifera in shallow waters is in accordance with catches in a beam trawl survey conducted at the same stations in 2015, but that survey found higher abundances (biomass and density) of echinoderms in deeper waters, while tunicates and fish (chordata) predominantly occurred in shallow waters (Williams et al., 2017).

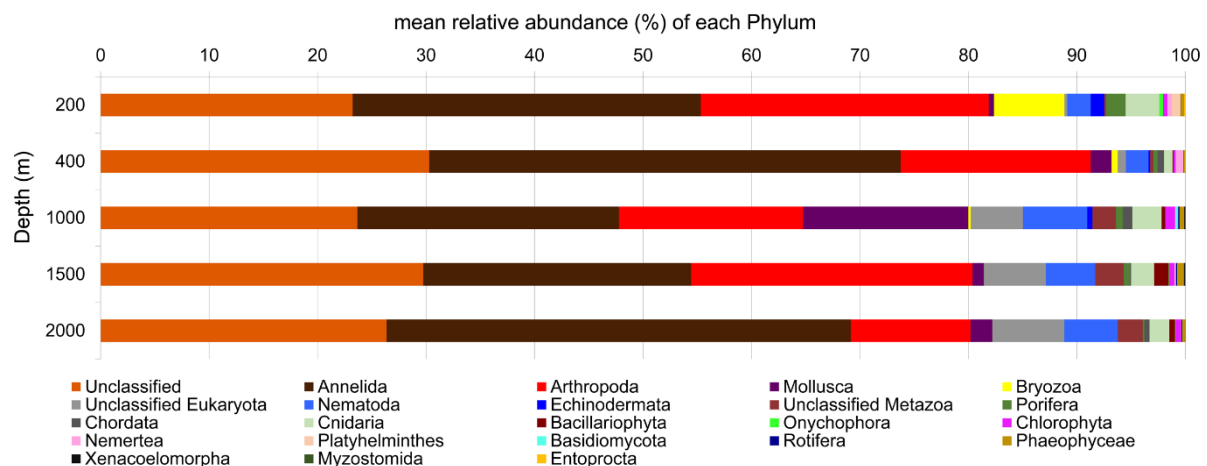


Figure 7.12. Mean relative abundances of OTUs at the phylum level across depths within the GAB based on analysis of the COI-5P gene region.

Ordination revealed that the samples were clearly grouped by depth (Figure 7.13). The PERMANOVA indicated a significant depth by transect interaction ($F_{16,38}=1.97$, $p<0.001$). However, pairwise tests indicated that the only significant differences among transects were between T2 and each of T3 and T5 at 400 m depth, while the majority of depth pairs differed on each transect except that the 1000, 1500 and 2000 m sites were the same on some. The depth trend varies with phylum, and among the five most abundant, was clearest in the Arthropoda and the Cnidaria (Figure 7.14), while the two shallower depths were intermingled for the Nematoda and Mollusca, and the three deeper were intermingled for the Annelida and Mollusca.

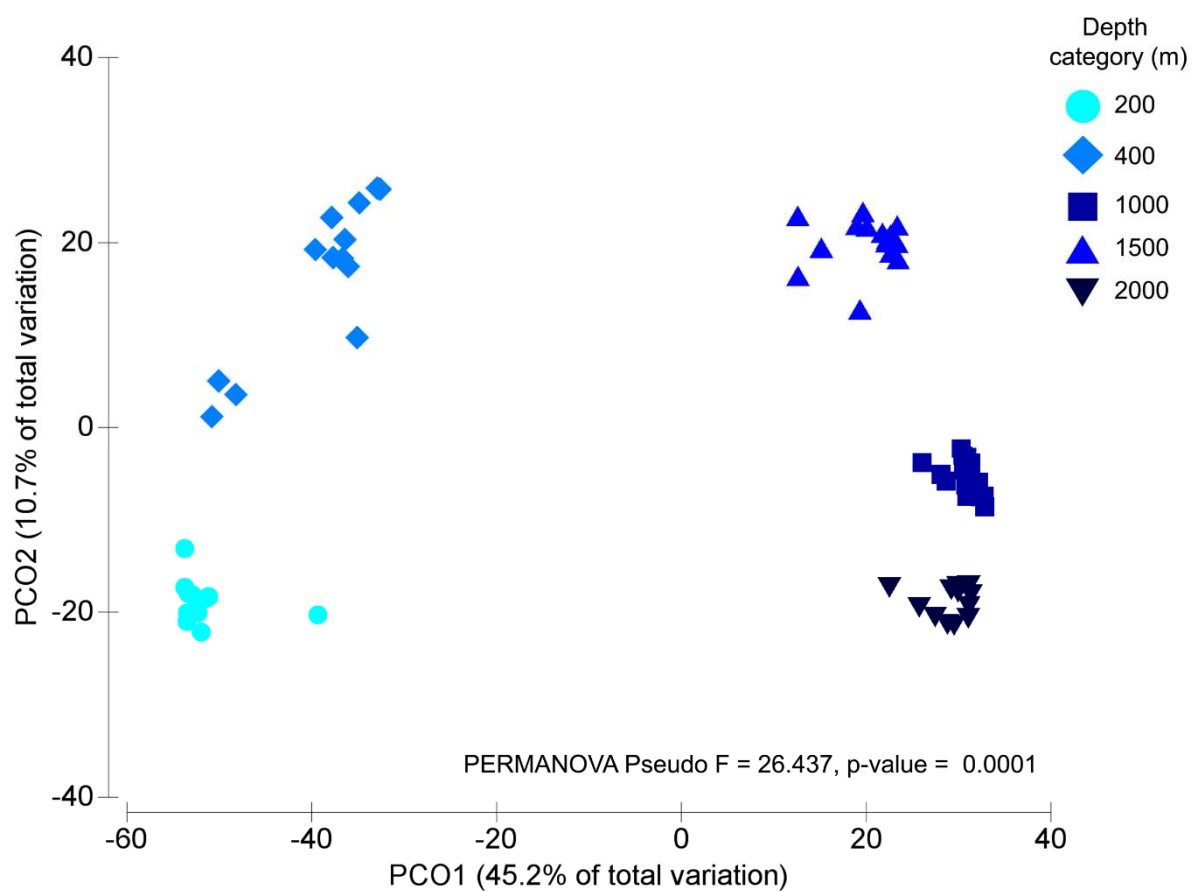


Figure 7.13. Ordination (PCO) plot of the global COI-5P OTU profiles from 65 benthic sediment samples collected over a range of depths from the GAB.

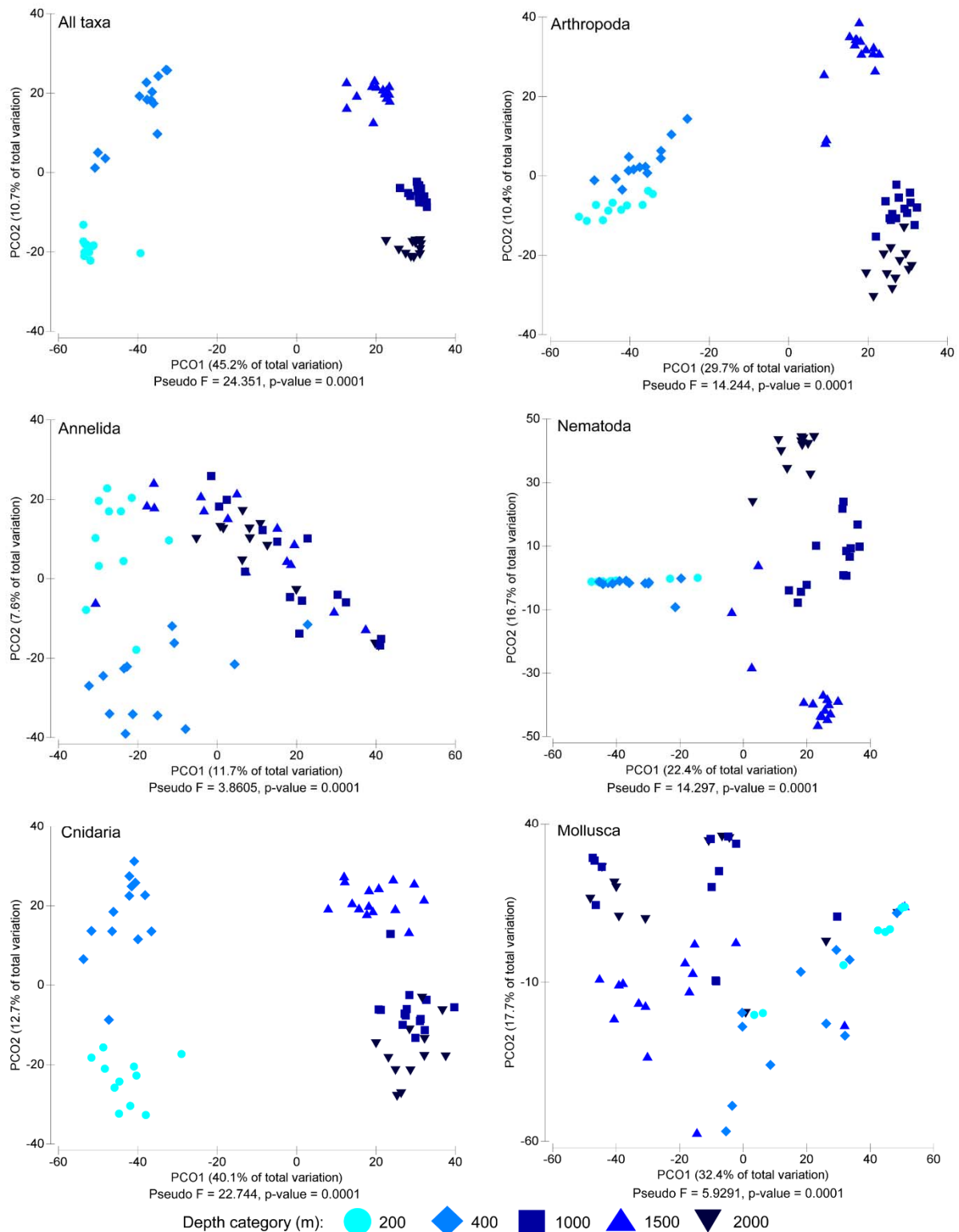


Figure 7.14. Ordination (PCO) plots of the relative abundances of COI-5P OTUs for all taxa and major phyla from 65 benthic sediment samples collected over a range of depths in the GAB.

Further analysis of the samples by cumulative dominance of the OTUs revealed that any one depth could vary from having one or more dominant taxa to many taxa with more uniform abundances (Figure 7.15). This highlights the patchiness of the faunal assemblages of the GAB benthos, which is particularly evident for certain macrofauna (e.g. molluscs, decapod crustacea and sponges) rather

than microfauna (e.g. diatoms, fungi and rotifers – see Figure 7.16), and may reflect limitations of the metabarcoding assay where DNA is extracted from only a small sample of sediment. Given this and the generally marked heterogeneity of benthic assemblages at all scales, particularly between sediment depths (Guardiola et al., 2016), future studies should be directed to bulk extractions from larger numbers of replicate samples.

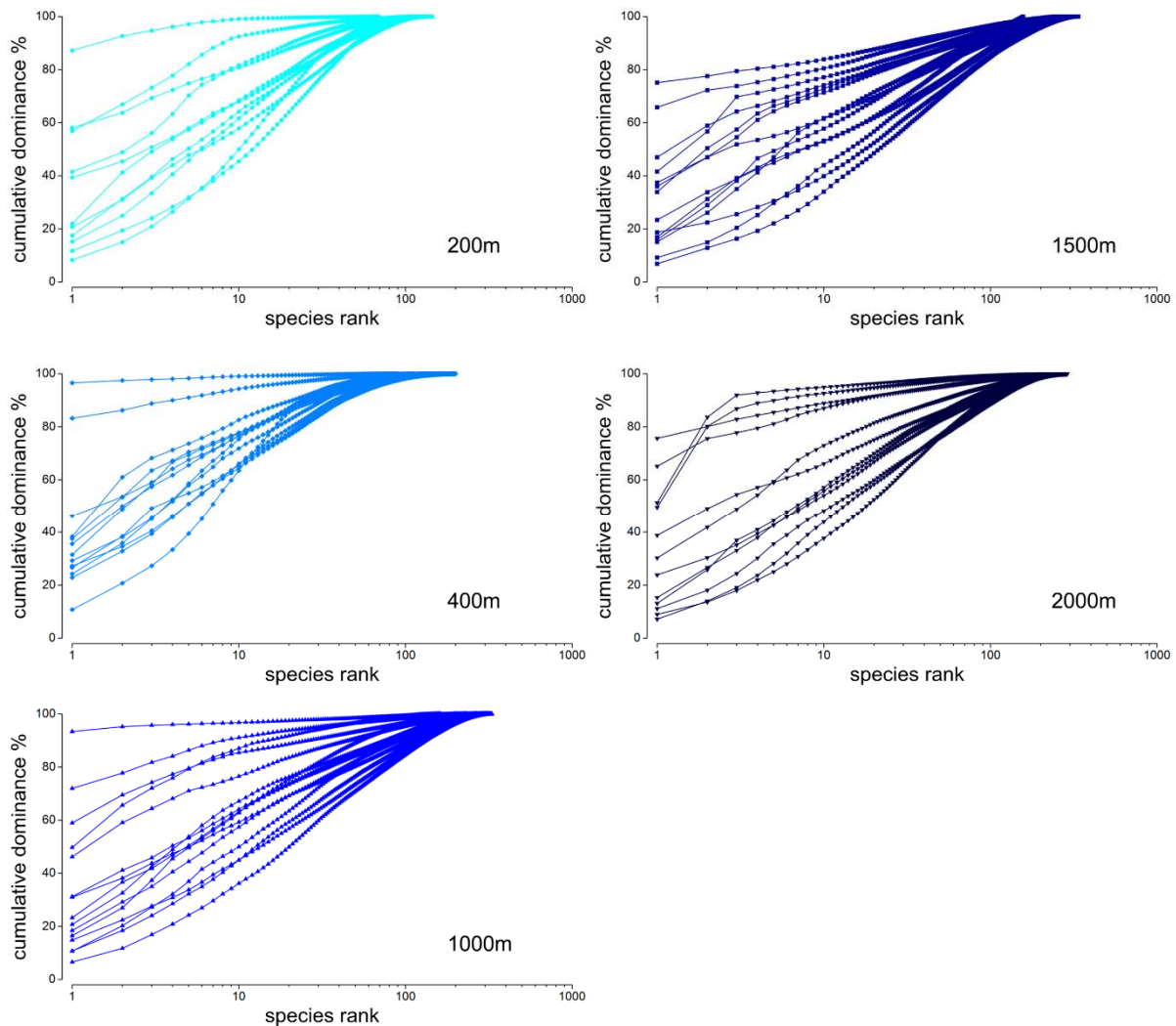


Figure 7.15. K-dominance plots representing the cumulative dominance of COI-5P OTUs across depths in the GAB.

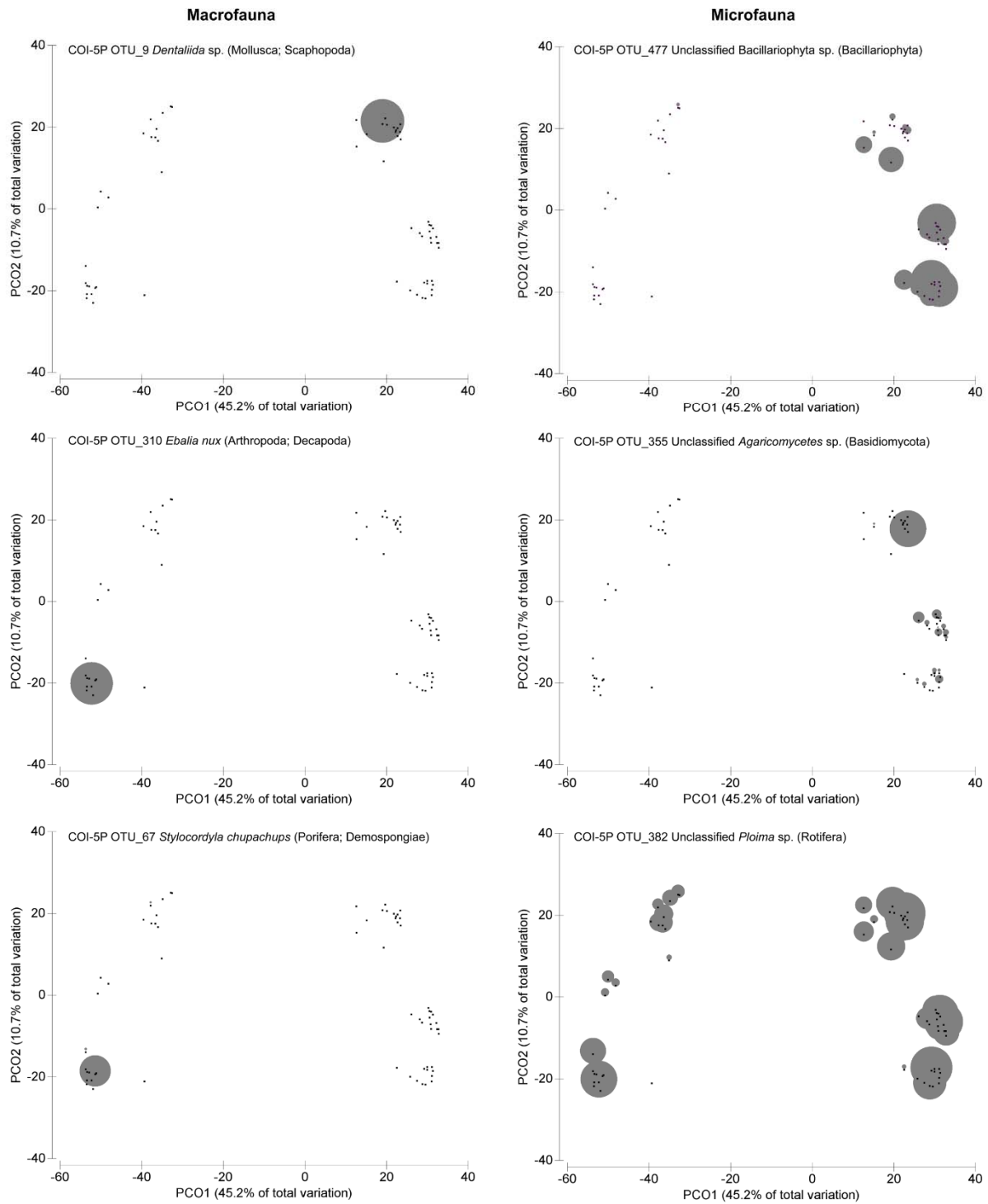


Figure 7.16. Bubble plots depicting the relative abundances of example macro- and microfaunal COI-5P OTUs across all 65 benthic sediment samples collected in 2013 from a range of depths within the GAB. Black dots represent individual samples and grey bubbles the relative abundance.

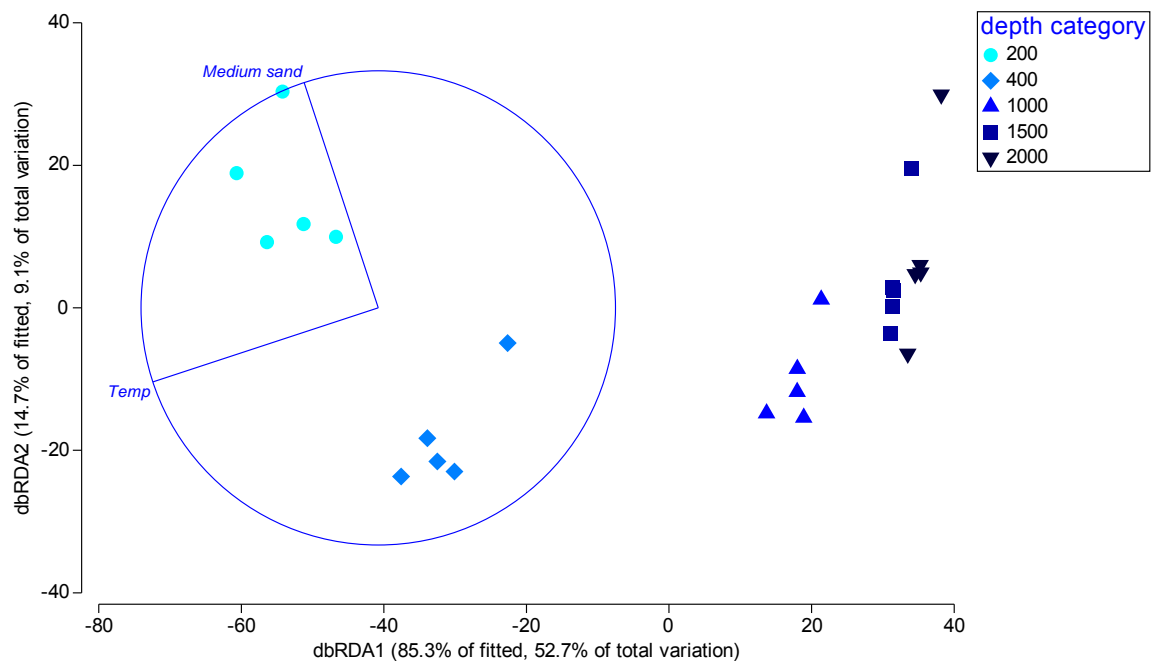


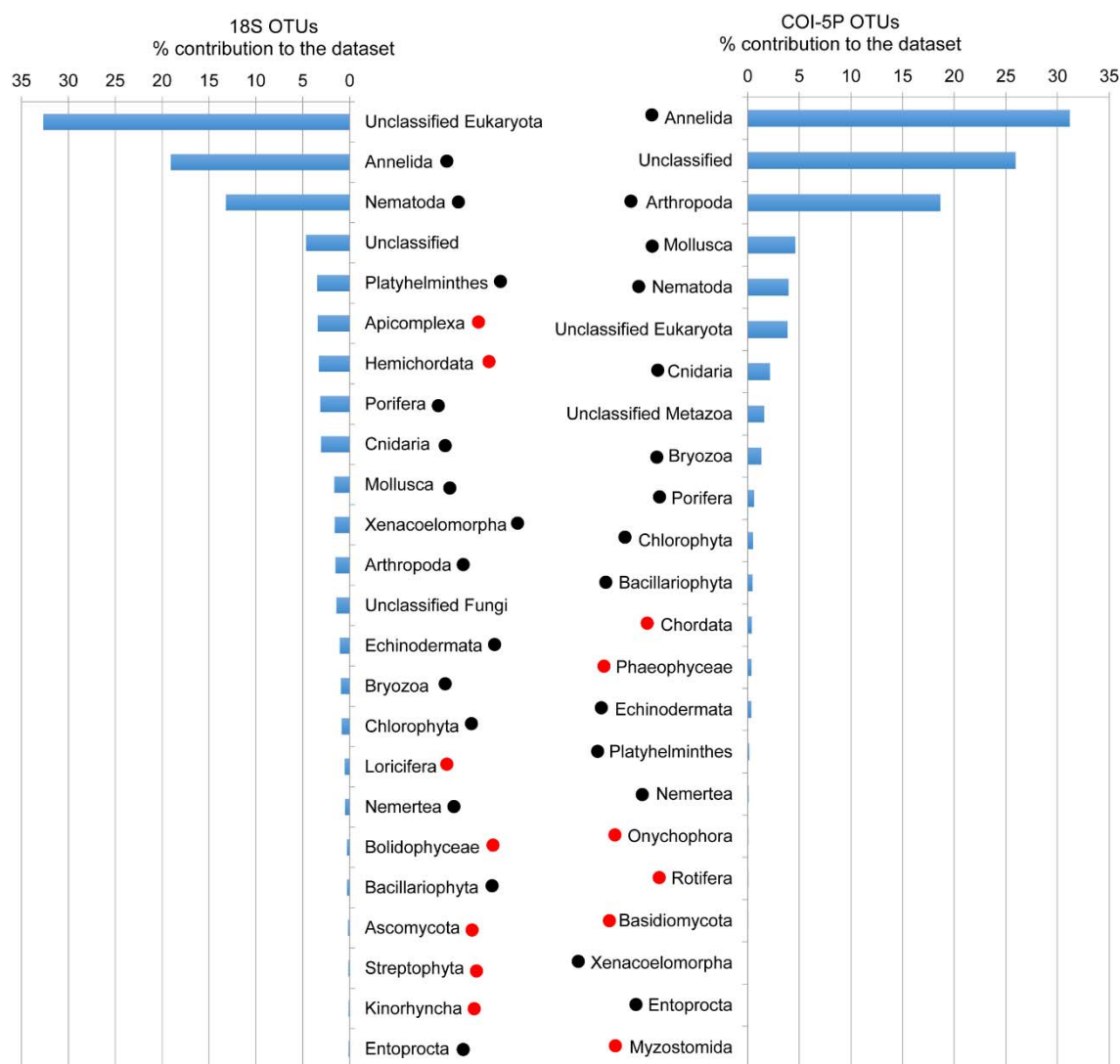
Figure 7.17. Relationships between COI community composition and environmental parameters as determined by distance-based linear modelling and visualised with distance-based redundancy analysis. Only variables contributing to the final selected model are included. Temperature and depth were strongly collinear ($R > 0.9$), and thus temperature was retained while depth was removed from the analysis.

The relationship between COI community composition and the environmental covariates came up with a BEST model that only included bottom temperature and medium sand (Figure 7.17, $R^2 = 0.62$). Marginal tests indicated that significant ($p < 0.05$) amounts of variation in assemblage structure could be explained by all variables when taken individually except for the proportion of very coarse sand, gravel and total carbon. The strongest individual predictors were temperature and proportion of very fine sand, explaining 52.6% and 43.0% of variation, respectively (pseudo- $F > 17.3$; $p < 0.001$).

7.4.3 Comparisons between COI-5P and 18S rRNA markers for estimating benthic diversity

Direct comparisons between the COI-5P and 18S rRNA markers were made from a total of 61 matched samples collected in 2013 (see Appendix 7.1). As a first step, the rarefaction curves were compared to elucidate the depth of coverage of the two approaches (Figure 7.1 and Figure 7.9). Unlike that observed for COI-5P, a low sequencing depth and lack of asymptote for most samples was apparent for the 18S gene assay, indicating that these NGS libraries were largely under-sampled. Despite this shortcoming, and in contrast to earlier reports that the 18S gene may under-estimate diversity in comparison to the COI gene (Leray and Knowlton, 2016; Tang et al. 2012), this assay was still able to recover a comparable number of OTUs to that obtained for the COI-5P assay (i.e. 990 vs 1026), indicating that for these samples both approaches are likely to provide equivalent estimates of richness. As anticipated based on the relatively recent adoption of COI for barcoding (Bucklin et al., 2011; Costa and Carvalho 2010), there was much greater ability to identify OTUs to species in the 18S data set (13.9%) compared to the COI dataset (~0.5%). Overall, 84.4% of 18S OTUs could be assigned at least phylum level taxonomy, while 65% of the COI-5P OTUs could be identified to this level. In comparing the phyla detected using the two approaches, several common and unique taxa were apparent, with 14 phyla detected by both assays and 7 being unique to 18S and 6 to COI-5P (Figure

7.18). However, for those phyla detected by both approaches the numbers of OTUs assigned to each were largely disproportionate between the two gene targets (e.g. Arthropods were underrepresented by 18S and Platyhelminthes by COI-5P), and likely reflects the inherent biases of the respective gene primers (Leray and Knowlton, 2016). Nevertheless, comparison of the depth-specific patterns across samples generated by both datasets (by comparing the similarity matrices for the 18S and COI-5P assays by rank correlation) revealed that both gene targets produced patterns that were moderately to well correlated (Figure 7.19), although comparisons of individual phyla indicate weaker relationships (Figure 7.20). Thus, whilst either approach is likely to provide reasonable global patterns of diversity, the broadest coverage of the benthic assemblages within the GAB is likely to be obtained by the implementation of both the 18S rRNA and COI-5P gene targets. If resources only allow one approach to be pursued, then the focus should be COI, as this is now the major focus of barcoding efforts on individual organisms, and over time the COI database is likely to become more comprehensive than that for 18S. However, given that the main expense is in sample collection, it would make sense to continue using both gene regions when possible, to both take advantage of the historical emphasis on 18S, and to ensure that as wide a range of taxonomic groups as possible are captured. In both cases, further concerted efforts to barcode the GAB's unique species will be required if good taxonomic resolution below the phylum level is desired. One obvious difference between the two gene regions was the tighter clustering of samples in the COI-5P analysis compared to the 18S. This pattern may be related to the apparent under-sampling of taxa by the 18S assay (Figure 7.1), which would result in a noisier data set and thus less ability to discriminate between different groups of samples.



(21 classified and 3 unclassified taxa)

(20 classified and 3 unclassified taxa)

Figure 7.18. Comparison of the total numbers of 18S and COI-5P OTUs assigned to major phyla from 61 matched benthic sediment samples obtained in 2013 from the GAB. Common and unique taxa which could be classified to the phylum level are marked for each gene target with black and red dots respectively.

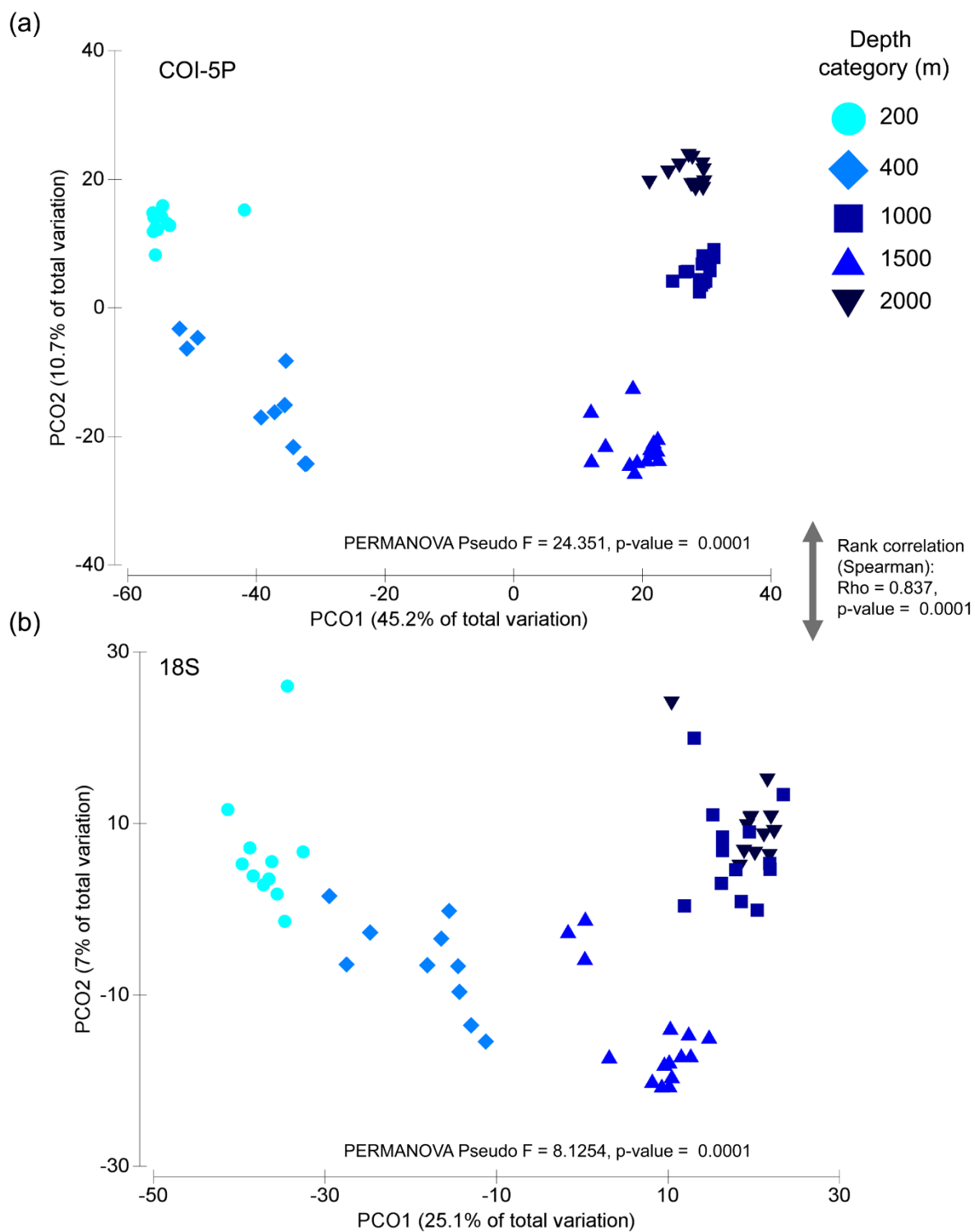


Figure 7.19. Ordination (PCO) plots comparing the (a) COI-5P and (b) 18S global OTU profiles obtained from the sequencing of 61 matched samples collected over a range of depths from the GAB in 2013.

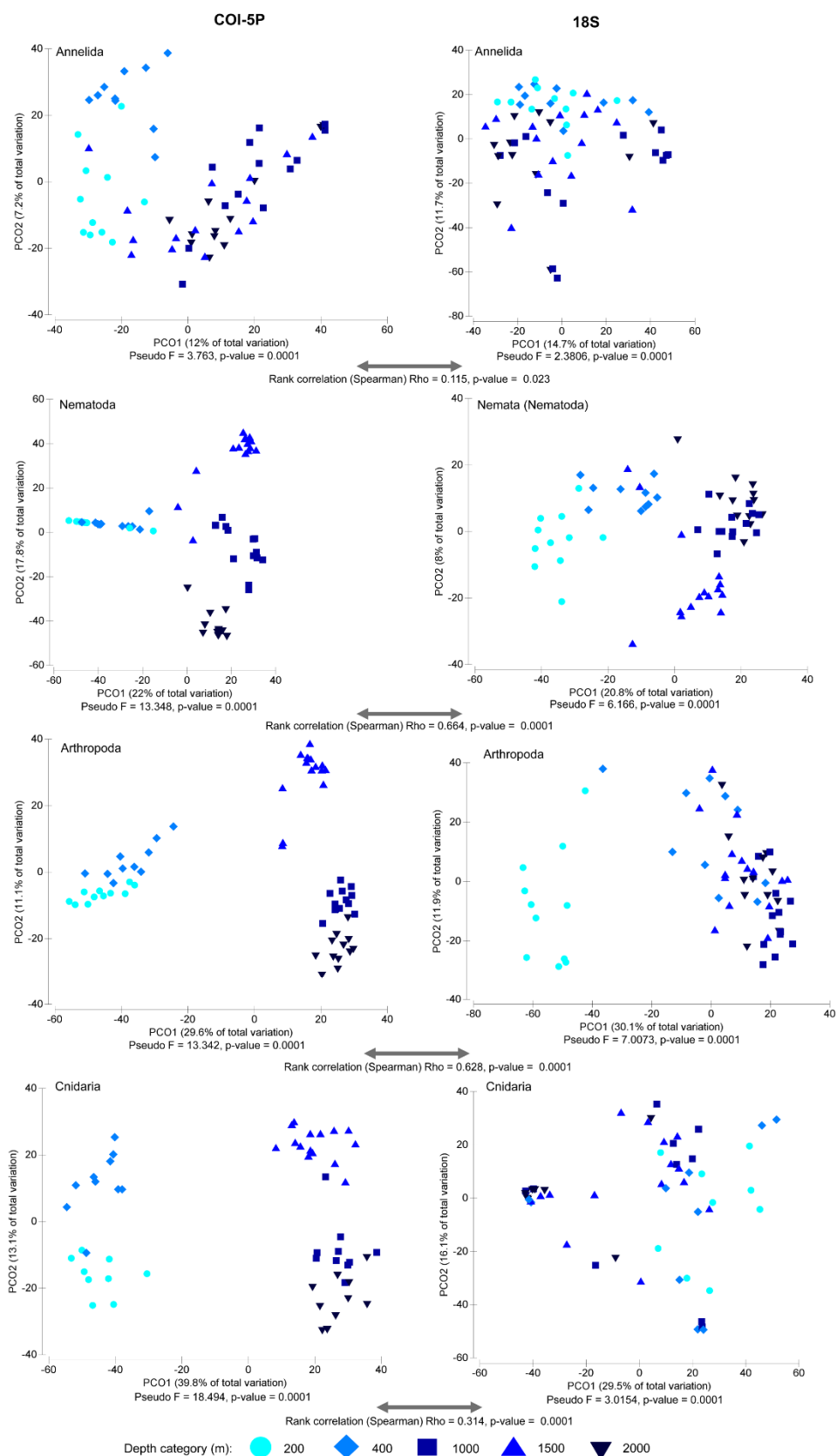


Figure 7.20. Ordination (PCO) plots comparing the relative abundances of COI-5P and 18S OTUs for four major phyla across depths from 61 matched GAB benthic sediment samples collected in 2013.

Appendix 7.1. List of benthic sediment samples collected over a range of depths in 2013 and 2015 from the GAB and the corresponding gene target/s sequenced.

Sample IDs and the corresponding site details are provided alongside the numbers of sequence reads (as a measure of sequencing depth) obtained from the deep-sequencing of NGS libraries prepared from the 18S rRNA gene and/or the 5' region of the COI gene (COI-5P). *A total of 1,154,891 sequence reads were obtained from the sequencing of 99 samples (75 from 2013 and 24 from 2015) for the 18S rRNA gene assay; and a total of 2,053,400 reads from 65 samples (obtained in 2013 only) for the COI-5P assay. †Samples were omitted from further analysis due to the absence or small numbers of reads obtained following sequencing. A total of 61 matched samples (highlighted in grey) were used in downstream comparisons between the 18S and COI-5P datasets.

Sample no.	Year	Transect	Depth (m)	Depth class (m)	Gene target sequenced		No. sequence reads*	
					18S	COI-5P	18S	COI-5P
5078	2013	5	2004	2000	+	-	16206	-
5079	2013	5	2004	2000	+	+	2745	56311
5080	2013	5	2004	2000	+	+	21202	24257
5156	2013	5	1524	1500	+	+	10296	28568
5157	2013	5	1524	1500	+	+	6865	1743
5158	2013	5	1524	1500	+	+	5190	27445
5284	2013	4	1494	1500	+	+	11405	49288
5285	2013	4	1494	1500	+	+	11078	42049
5286	2013	4	1494	1500	+	+	8895	17350
5358	2013	4	1924	2000	+	+	8483	31198
5359	2013	4	1924	2000	+	+	21196	37854
5360	2013	4	1924	2000	+	+	5413	23526
5490	2013	3	2006	2000	+	+	6545	30143
5491	2013	3	2006	2000	+	+	12716	34577
5492	2013	3	2006	2000	+	+	6343	13275
5571	2013	3	1479	1500	+	+	7043	29521
5572	2013	3	1479	1500	+	+	5110	52832
5573	2013	3	1479	1500	+	-	9085	-
5748	2013	1	2014	2000	+	-	5773	-
5749	2013	1	2014	2000	+	-	6650	-
5750	2013	1	2014	2000	+	+	4956	13414
5869	2013	2	2010	2000	+	+	13369	34027
5870	2013	2	2010	2000	+	+	12504	29485
5871	2013	2	2010	2000	+	+	13717	24566
5926	2013	2	1482	1500	+	+	11321	38318
5927	2013	2	1482	1500	+	+	10276	31780
5928	2013	2	1482	1500	+	+	6777	36342
6006 [†]	2013	1	1544	1500	+	+	4	30943
6007	2013	1	1544	1500	+	+	13818	45819
6008	2013	1	1544	1500	+	+	6618	36819
6091	2013	1	1015	1000	+	+	7459	29399
6092	2013	1	1015	1000	+	+	18779	42910
6093	2013	1	1015	1000	+	+	7815	24789
6177	2013	1	414	400	+	+	8769	11417

6178	2013	1	414	400	+	-	14830	-
6179 [†]	2013	1	414	400	+	+	0	37972
6255	2013	1	204	200	+	+	15417	22821
6256	2013	1	204	200	+	-	12207	-
6257	2013	1	204	200	+	+	28536	26163
6329	2013	2	188	200	+	+	7906	33767
6330	2013	2	188	200	+	+	8042	38114
6331	2013	2	188	200	+	-	11360	-
6409	2013	2	477	400	+	+	12446	48509
6410	2013	2	477	400	+	+	14314	40269
6411	2013	2	477	400	+	+	7242	31330
6487	2013	3	203	200	+	+	546	40130
6488	2013	3	203	200	+	+	16045	33498
6498	2013	3	203	200	+	+	8715	34534
6571	2013	3	385	400	+	+	9097	23365
6572	2013	3	385	400	+	+	10559	43904
6576 [†]	2013	3	385	400	+	+	6	21971
6634	2013	2	996	1000	+	+	12545	38020
6635	2013	2	996	1000	+	+	11374	34157
6636	2013	2	996	1000	+	+	10655	35883
6767	2013	3	1020	1000	+	+	9099	27949
6768	2013	3	1020	1000	+	+	9104	33611
6769	2013	3	1020	1000	+	+	15722	35770
6858	2013	4	211	200	+	+	12940	27877
6859	2013	4	211	200	+	+	15957	39718
6860	2013	4	211	200	+	+	38514	33645
6930 [†]	2013	4	440	400	+	+	0	42086
6931	2013	4	440	400	+	-	8886	-
6932	2013	4	440	400	+	+	8036	35336
7003	2013	4	1015	1000	+	+	13130	44261
7004	2013	4	1015	1000	+	+	7370	39176
7005	2013	4	1015	1000	+	+	9025	34822
7098	2013	5	980	1000	+	+	14349	21023
7099	2013	5	980	1000	+	-	10626	-
7100	2013	5	980	1000	+	-	10119	-
7173	2013	5	235	200	+	-	9725	-
7174 [†]	2013	5	235	200	+	-	7	-
7175	2013	5	235	200	+	+	7321	21778
7250	2013	5	375	400	+	+	6378	32483
7251	2013	5	375	400	+	+	7699	16102
7252	2013	5	375	400	+	+	6928	17978
0433	2015	4	2761	2800	+	-	17192	-
0434	2015	4	2761	2800	+	-	12988	-
0435	2015	4	2761	2800	+	-	30757	-
0607	2015	3	2694	2800	+	-	13740	-
0608	2015	3	2694	2800	+	-	40699	-

0609	2015	3	2694	2800	+	-	9625	-
0644	2015	2	2994	2800	+	-	35340	-
0645	2015	2	2994	2800	+	-	12768	-
0646	2015	2	2994	2800	+	-	10923	-
0661	2015	1	2867	2800	+	-	10537	-
0662	2015	1	2867	2800	+	-	12567	-
0663	2015	1	2867	2800	+	-	3521	-
0856	2015	2	996	1000	+	-	2442	-
0857	2015	2	996	1000	+	-	9725	-
0858	2015	2	996	1000	+	-	16621	-
1285	2015	1	995	1000	+	-	17295	-
1286	2015	1	995	1000	+	-	8503	-
1287	2015	1	995	1000	+	-	17719	-
3499	2015	1	1561	1500	+	-	36652	-
3500	2015	1	1561	1500	+	-	8052	-
3501	2015	1	1561	1500	+	-	9706	-
3515	2015	1	2012	2000	+	-	9794	-
3516	2015	1	2012	2000	+	-	11039	-
3517	2015	1	2012	2000	+	-	10507	-

7.5 References

- Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26: 32-46.
- Bray, J.R., Curtis, J.T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs* 27: 325-349.
- Bucklin, A., Steinke, D., Blanco-Bercial, L. (2011). DNA barcoding of marine metazoa. *Annual Review of Marine Science* 3: 471-508.
- Camarinha-Silva, A., Jáuregui, R., Chaves-Moreno, D., Oxley, A.P., Schaumburg, F., Becker, K., Wos-Oxley, M.L., Pieper, D.H. (2014). Comparing the anterior nares bacterial community of two discrete human populations using Illumina amplicon sequencing. *Environmental Microbiology* 16: 2939-2952.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Wildmann, J., Yatsunenko, T., Zaneveld, J., Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7: 335-336.
- Carugati, L., Corinaldesi, C., Dell'Anno, A., Danovaro, R. (2015). Metagenetic tools for the census of marine meiofaunal biodiversity. *Marine Genomics* 24: 11-20.
- Chaves-Moreno, D., Plumeier, I., Kahl, S., Krismer, B., Peschel, A., Oxley, A.P.A., Jáuregui, R., Pieper, D.H. (2015). The microbial community structure of the cotton rat nose. *Environmental Microbiology Reports* 7: 929-935.
- Clarke, K.R., Warwick, R.M. (2001). Change in marine communities: an approach to statistical analysis and interpretation, 2nd edition. Plymouth: PRIMER-E Limited. 172 pp.
- Costa, F.O., Carvalho, G.R. (2010). New insights into molecular evolution: prospects from the Barcode of Life Initiative (BOL). *Theory in Biosciences* 129: 149-157.
- Cowart, D.A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., Arnaud-Haond, S. (2015). Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. *PLoS One* 10: e0117562.
- Creer, S., Fonseca, V.G., Porazinska, D.L., Giblin-Davis, R.M., Sung, W., Powers, D.M., Packer, M., Carvalho, G.R., Blaxter, M.L., Lamshead, P.J.D., Thomas, W.K. (2010). Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology* 19 (Suppl. 1): 4-20.
- Dell'Anno, A., Danovaro, R. (2005). Extracellular DNA plays a key role in deep-sea ecosystem functioning. *Science* 309: 2179.
- Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26: 2460-2461.
- Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* 10: 996-998.
- Guardiola, M., Wangenstein, O.S., Taberlet, P., Coissac, E., Uriz, M.J., Turon, X. (2016). Spatio-temporal monitoring of deep-sea communities using metabarcoding sediment DNA and RNA. *PeerJ* 4: e2807.
- Elbrecht, V., Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PLoS ONE* 10: e0130324.

- Forschner, S.R., Sheffer, R., Rowley, D.C., Smith, D.C. (2008). Microbial diversity in Cenozoic sediments recovered from the Lomonosov Ridge in the Central Arctic Basin. *Environmental Microbiology* 11: 630-639.
- Kingsford, M., Battershill, C., 1998. Studying temperate marine environments: A handbook for ecologists. Canterbury University Press, Christchurch, NZ.
- Leray, M., Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences USA* 112: 2076-2081.
- Leray, M., Knowlton, N. (2016). Censusing marine eukaryotic diversity in the twenty-first century. *Philosophical Transactions of the Royal Society B* 371: 20150331.
- Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T., Machida, R.J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology* 10:34.
- Monrad, R., Lejzerowicz, F., Darling, K.F., Lecroq-Bennet, B., Pedersen, W., Orlando, L., Pawlowski, J., Mulitza, S., de Vargas, C., Kucera, M. (2016). Plankton-derived environmental DNA extracted from abyssal sediments preserves patterns of plankton macroecology. *Biogeosciences Discussions*
- Ratnasingham, S., Hebert, P.D. (2007). BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes* 7: 355-364.
- Rouse, G.W., Wilson, N.G., Carvajal, J.I., Vrijenhoek, R.C. (2016). New deep-sea species of *Xenoturbella* and the position of *Xenacoelomorpha*. *Nature* 530: 94-97.
- Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G., Fontaneto, D. (2012). The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences USA* 109: 16208-16212.
- Williams, A., Tanner, J.E., Althaus, F., Sorokin, S.J., MacIntosh, H., Green, M., Brodie, P., Loo, M., 2017. Great Australian Bight Benthic Biodiversity Characterisation. Final Report GABRP Project 3.1. Great Australian Bight Research Program, GABRP Research Report Series Number 16, 361pp.
- Xu, W., Pang, K-L., Luo, Z-H. (2014). High fungal diversity and abundance recovered in the deep-sea sediments of the Pacific Ocean. *Microbial Ecology* 68: 688-698.
- Zhang, J., Kobert, K., Flouri, T., Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30: 614-620.

8. DISCUSSION

This project applied a suite of molecular approaches to provide a broader overview of the fauna present in the deep GAB than would have been achieved by conventional methods alone (as detailed in Williams et al. 2017). A particular focus was on the microbial assemblages present, as components of this assemblage have the potential to degrade hydrocarbons. The presence of these organisms in any abundance would indicate two things. First, that there is a natural source of hydrocarbons present in the GAB to provide a food source to allow them to thrive. Second, that the GAB ecosystem is in at least some ways pre-adapted to the presence of hydrocarbons, and thus may show a greater level of resilience in the event of an oil spill than would otherwise be the case. We found a number of microbial taxa in the GAB that are related to known hydrocarbon degrading species elsewhere in the world, and thus may have the potential to degrade hydrocarbons themselves. This finding is reinforced by the finding that genes involved in 3 key pathways of hydrocarbon degradation are abundant in the GAB. However, both the microbes present, and their hydrocarbon degradation genes, differ from those found elsewhere. Thus, while we can speculate that the microbial assemblage in the GAB does have the capacity to deal with hydrocarbons, we cannot assume that they will respond in the same way that microbial assemblages in regions such as the Gulf of Mexico do. If we wish to make predictions about how these microbial assemblages might respond to an oil spill, we will need to undertake experimental work in the laboratory, exposing assemblages to the type of oil likely to be present in the GAB environment. Such experiments would allow the determination of how quickly the microbial assemblage changes in the presence of hydrocarbons, and in particular whether putative hydrocarbon degraders increase in abundance while taxa that do not utilise hydrocarbons decrease. It would also be possible to measure the breakdown rates of the oil and any derivative hydrocarbons, and correlate these with changes in the microbial assemblage. This work would have to be undertaken on freshly collected cores maintained at low temperatures (2-3 °C) reflective of conditions at the bottom of the deep ocean.

As well as the microbial assemblage, we also used molecular techniques to examine the eukaryote assemblages present in the GAB. Traditional methods for examining these assemblages involve the collection of samples using gear such as the multicorer used for our samples, beam trawls, dredges etc, and then a long and detailed process of sorting out individual organisms from the samples and identifying them. Even then, it is generally only possible to include macroscopic species, with the large numbers of microscopic species likely present being missed. Using metabarcoding approaches, it is possible to obtain a much broader overview of the biodiversity present more quickly, and covering all size fractions. However, like all sampling techniques, metabarcoding has its own biases, and so still does not provide a complete picture of all organisms present. We used two different gene regions (18S and COI) for metabarcoding, with both showing broadly similar results in terms of the number of putative taxa present (~1000), and the depth structuring of the assemblage. However, there were some clear differences between the two gene regions with respect to what taxa were detected. One of the drawbacks of this approach is that it is difficult to obtain species level determinations for many of the taxa present in a poorly sampled area such as the GAB, as this first requires specimens to be collected, identified by a specialist taxonomist, and then individually barcoded. While we did this to a certain extent, of the 303 species individually barcoded, only 7 were detected in the metabarcoding analyses. In part, this likely reflects that sample quality for the infaunal specimens for which individual barcoding was attempted was poor, and thus most of the barcodes elucidated were for epifauna, whereas the sediment samples that were metabarcoded would primarily have contained infauna, and most epifauna DNA would have been endogenous. There is thus a lot more effort required to individually barcode species from the GAB before this technique can provide detailed information on the taxa present at the species level, although it does perform much better at higher taxonomic levels.

The molecular approaches developed and employed for both the microbial and eukaryotic assemblages have the potential to provide rapid assessment tools for use in monitoring of any future oil and gas developments in the GAB. In particular, we have developed a series of qPCR assays for microbial taxa involved in hydrocarbon degradation as a low cost, high throughput tool for future monitoring. As they both utilise sediment samples for analysis, they provide better potential for replication than traditional epifaunal surveys, where 1 sample may take the better part of 10-12 hours in depths around 3000 m (compared to 3-6 replicate samples from a multicorer in 5-6 hours). The same level of replication can be obtained from traditional infaunal surveys, which utilise the same type of sediment sample as we used for the molecular surveys, although infaunal analysis requires more extensive follow-up work in the laboratory. A current disadvantage with the molecular approaches is that it is more difficult to appreciate what the ecological consequences of any changes might be, as the results are not necessarily quantitative, and most taxa present can only be identified to higher taxonomic levels. It may thus be necessary to supplement molecular assessments with more detailed traditional assessment if the former indicate that a substantial impact has occurred.

Williams, A., Tanner, J.E., Althaus, F., Sorokin, S.J., MacIntosh, H., Green, M., Brodie, P., Loo, M., 2017. GAB benthic biodiversity characterisation. GABRP research report Number 16, Great Australian Bight Research Program, October 2017. 356 pp.

9. APPENDIX 1: DATA MANAGEMENT

9.1 Raw datasets created

Raw mitochondrial DNA sequences from the COI gene for individual benthic specimens.

Raw mitochondrial DNA sequences from the COI gene from sediment samples.

Raw 18S sequence data from sediment samples.

Raw 16S sequence data from sediment and water samples.

Raw functional gene (*alkB*, *c23o*, *pmoA*) sequence data from sediment and water samples.

9.2 Data processing and derived datasets

Filtered mitochondrial DNA sequences from the COI gene for individual benthic specimens.

Filtered and clustered mitochondrial DNA sequences from the COI gene for sediment samples to remove likely artefacts and separate into distinct operational taxonomic units (OTUs). These OTUs were then matched to gene sequences in the BOLD database at appropriate taxonomic levels.

Filtered and clustered DNA sequences from the 18S gene for sediment samples to remove likely artefacts and separate into distinct operational taxonomic units (OTUs). These OTUs were then matched to gene sequences in the SILVA database at appropriate taxonomic levels.

Filtered and clustered 16S gene sequence data for both bacterial and archaeal marker genes allocated to OTUs, and OTU abundance tables for each sample and gene combination.

Filtered and clustered functional gene data, and OTU abundance tables for each sample and gene combination.

9.3 Data curation and archive

Individual sequence data are located on an eResearch SA repository database and a Molecular Ecology Lab at Flinders University hard drive.

Sediment COI data are located on SARDI Tier 1 network drive (cluscbedfs02\user30\SARDI\Env and Ecol\BP GAB\Benthic\Molecular).

Other data are located on CSIRO SC Data Store – Ruby.

9.4 Data access, use agreements and licensing

Cawthron Institute and University of Geneva. Material transfer agreement for DNA from sediment samples for use in a study on forams.

9.5 Publication of datasets

Individual COI data will be published on GenBank when the associated paper is published.

Other data will be made available on the NCBI Sequence Read Archive on publication of the associated papers.

10. APPENDIX 2: PROJECT PUBLICATIONS

10.1 Papers

Nil to date

10.2 Presentations

L Bodrossy, J Van De Kamp, S Hook, J Tanner and A Williams. Diversity and abundance of microbial hydrocarbon degradation pathways in the offshore environments of the Great Australian Bight. July 2017, Australian Marine Sciences Association Conference, Darwin.

J van de Kamp, S Hook, A Williams, JE Tanner, L Bodrossy. Diverse hydrocarbon degrading microbial community in a pristine ocean basin. August 2016, 16th International Symposium on Microbial Ecology, Montreal, Canada.



THE UNIVERSITY
of ADELAIDE



Flinders
UNIVERSITY