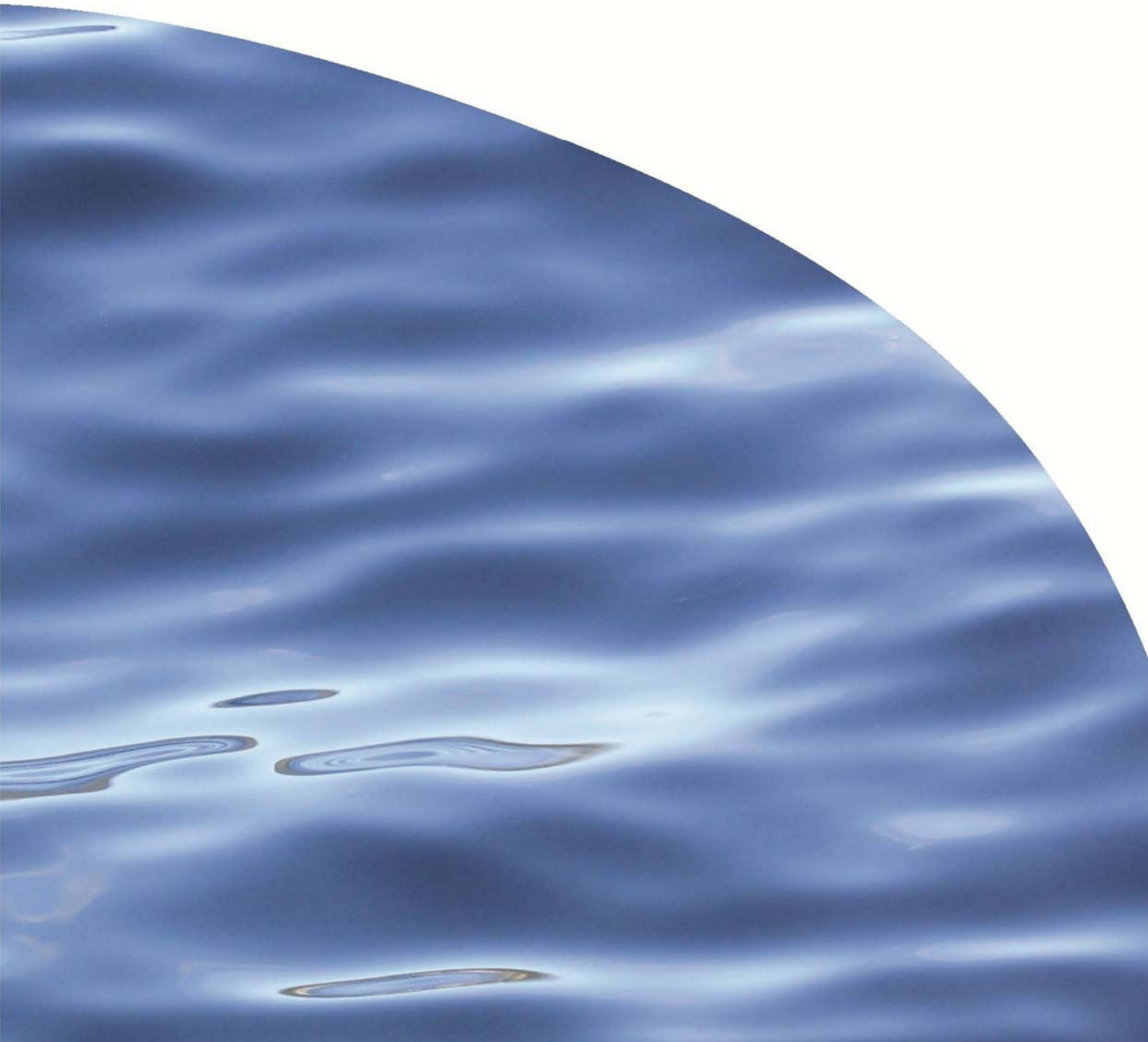
CAWTHRON
The power of science®

REPORT NO. 2980

DEVELOPMENT OF A MULTI-TROPHIC LEVEL METABARCODING TOOL FOR BENTHIC MONITORING OF AQUACULTURE FARMS



DEVELOPMENT OF A MULTI-TROPHIC LEVEL METABARCODING TOOL FOR BENTHIC MONITORING OF AQUACULTURE FARMS

XAVIER POCHON, NIGEL KEELEY, SUSIE WOOD

Prepared for Seafood Innovation Limited Ltd, New Zealand King Salmon Ltd, Ngāi Tahu Seafood, the Ministry for Primary Industries, Waikato Regional Council, and the Marlborough District Council

CAWTHRON INSTITUTE
98 Halifax Street East, Nelson 7010 | Private Bag 2, Nelson 7042 | New Zealand
Ph. +64 3 548 2319 | Fax. +64 3 546 9464
www.cawthron.org.nz

REVIEWED BY:
Anastasija Zaiko

APPROVED FOR RELEASE BY:
Chris Cornelisen

ISSUE DATE: 16 January 2017

RECOMMENDED CITATION: Pochon X, Keeley N, Wood S 2017. Development of a new molecular tool for biomonitoring New Zealand's fish farms. Prepared for Seafood Innovation Limited Ltd, New Zealand King Salmon Ltd, Ngāi Tahu Seafood, the Ministry for Primary Industries, Waikato Regional Council, and the Marlborough District Council Cawthron Report No. 2980. 48 p. plus appendices.

EXECUTIVE SUMMARY

The Cawthron Institute was commissioned by a range of private and government agencies and industry partners to develop a molecular-based tool for assessing benthic impacts associated with salmon farming practices in New Zealand. The analysis was undertaken using cutting-edge molecular techniques, with the view that over time these rapidly evolving techniques could be integrated into the current suite of assessment tools routinely used by industry partners and stakeholders. The proposed approach also has the potential to improve monitoring efficiencies (e.g. quicker turn around, lower costs) while being transferable to other enrichment point sources in the marine environment (e.g., river plumes, dredging sites, and other aquaculture operations).

We used a method known as environmental DNA/RNA (eDNA/eRNA) metabarcoding[1]. This enables the cost-effective identification of organisms by personnel without taxonomic expertise through matching short gene fragments to a reference sequence library. We collected macrofaunal and physico-chemical data (used to calculate the traditional Enrichment Stage [ES] index) and molecular data (used for metabarcoding of three taxonomic groups: foraminifera, bacteria and eukaryota) from samples collected at three New Zealand's salmon farms over three years and from two regions (Marlborough Sounds and Stewart Island). Using a range of statistical and analytical approaches, we allocated Eco-Groups with consistent ecological responses to enrichment to over 500 known and unknown taxa across the three taxonomic groups. Using these Eco-Groups we successfully developed individual Metabarcoding Biotic Indices (MBI) for the three taxonomic groups (foraminiferal-MBI, bacterial-MBI, and eukaryotic-MBI) and a multi-trophic MBI combining all of the taxonomic groups or just the Bac and Euk.

Our results reveal that the weakest relationship was between the foraminiferal-MBI and ES, whereas strong ($R^2 > 0.9$, i.e. 90%) relationships were obtained with the multi-trophic MBI, or when just bacterial and eukaryotic data were combined. Furthermore, there was no difference in the relationships when using eDNA or eRNA. These results indicate that foraminifera should be excluded from the multi-trophic MBI. It may be possible to use only eDNA samples for routine monitoring, which will markedly reduce analytical time and costs, however, further testing is required to confirm this. The identification of high-quality molecular Eco-Groups and corresponding sequence GenCodeIDs from this project has established an extremely valuable sequence reference database to which future eDNA samples from fish farms can be automatically compared to rapidly calculate the multi-trophic MBI.

The strong correspondence between the ES and multi-trophic MBI confirms that this tool is ready for implementation. We recommend that the multi-trophic MBI be used in parallel to current ES methods for a phase-in period of four years to further ground-truth the tool and in turn facilitate its incorporation as the primary monitoring tool in the Marlborough Sounds.

[1] Rapid method of biodiversity assessment that combines two technologies: DNA-based identification and high-throughput DNA sequencing. It uses universal Polymerase Chain Reaction (PCR) primers to mass-amplify DNA barcodes from mass collections of organisms or from eDNA.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

1. INTRODUCTION

Assessing the impact of aquaculture on marine environments is a requirement under coastal permits issued under the Resource Management Act 1991 and represents a significant ongoing and increasing cost to industry. For example, the New Zealand aquaculture industry spends in the order of $1.2m p.a. on such monitoring. Expansion and development of new farms (e.g., 49 new consents since 2012) require ongoing environmental assessment, and by 2025 all existing (i.e., ca. 700) farms will be subjected to environmental monitoring programmes (Aquaculture New Zealand magazine, October 2013).

Seabed impacts resulting from fish farm activities in the Marlborough Sounds, New Zealand, are currently determined by measuring chemical properties of sediment and changes in macrofaunal diversity. These parameters are incorporated into the Enrichment Stage (ES) index (Keeley et al. 2012a, 2012b, 2013; MPI 2015), which provides regulators and producers with an integrated, weight-of-evidence-based measure of environmental impact (Box 1). The biological information included in the ES index incorporates a variety of statistics that are commonly used to describe benthic macrofaunal assemblages (e.g., richness, diversity). The overall impact status of a site can be projected on a discrete scale ranging from 1 (natural/pristine) to 7 (extremely enriched azoic conditions; Table 1). An ES 5 represents an important stage on the enrichment scale, as it corresponds to the peak in macrofaunal abundance (and sometimes biomass, Keeley et al. 2013), characterised by the proliferation of opportunistic taxa (Keeley et al. 2012a, 2012b). An ES 5 also represents the upper regulatory limit for fish farming compliance in the Marlborough Sounds.

Although very effective, the macrofaunal assessment component of the ES approach is labour intensive, expensive and slow. It requires a high-level of taxonomic expertise, which is a skill of shrinking capability worldwide (Jones 2008). Furthermore, it is exclusively based on the ‘visible’ portion of the benthic ecosystem, omitting important microbiota (e.g., bacteria and protists) which are known to also display community shifts in responses to benthic enrichment associated with feed-added and non-feed aquaculture (McCaig et al. 1999; Bissett et al. 2006; Pawlowski et al. 2014a). The increasing need from international aquaculture markets to demonstrate environmentally sustainable production systems (Kobayashi et al. 2015) implies there is a demand for the development of accurate, rapid and cost-effective environmental diagnostics monitoring tools.

Box 1: Enrichment Stage methodology: Managing for effects of aquaculture on the seabed

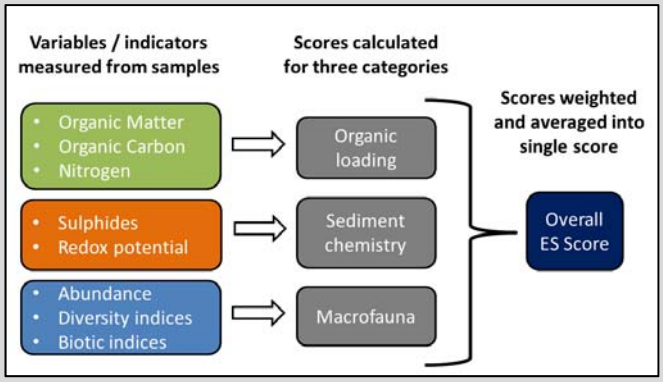
The issue: Aquaculture farms cause impacts to the seabed as a result of deposition of organic material onto the underlying seabed. In the case of finfish farms, feed and faecal deposition results in high levels of organic enrichment, which in turn results in the depletion of oxygen within sediments, changes in sediment chemistry (e.g. increase in sulphides), and changes in the number and types of organisms living in the sediments. Mussel and oyster farms lead to similar but much less intense effects on the seabed through the deposition of faecal material and increased sedimentation rates.

Managing for effects: Effects on the seabed must be managed to minimise the ecological footprint of aquaculture and also ensure farms remain healthy and productive. Managing farm consents requires indicators that are linked to the activity and are robust enough to assess compliance according to environmental quality standards. Furthermore, the responses of indicators need to be sensitive to changes in the activity (e.g. feed levels, stocking densities) over time and spatially in order to facilitate adaptive management responses that keep effects within a consented limit (standard).

Enrichment Stage (ES) methodology: There are numerous single measures and indicators that can be calculated for describing effects of aquaculture; however, they all respond differently to aquaculture activities, thereby complicating their application within a compliance and adaptive management framework. The need for a single, robust measure for managing aquaculture consents led to the development of the ES methodology, which integrates multiple measures of the biological, chemical and physical changes that occur within the depositional footprint of aquaculture farms. The ES methodology is grounded in the well-known concept of ecological succession in stressed environments and organic enrichment gradients used in developing benthic health indices around the world. The end result is an 'overall ES score' that captures the full range of possible effects in a single measure, from pristine natural conditions (ES = 1) to extremely enriched and impacted conditions (ES = 7).

How is ES calculated? A number of variables are first measured and then indicators that contribute to (and complement) the ES score are estimated. Important contributors to ES scores include measures of organism abundance and the diversity of communities living on and within the sediments. These organisms are referred to broadly as macrofauna, and infauna for those living within the sediments.

Measurements of variables and indicators for seabed samples are used to calculate scores on an equivalent ES scale (1 to 7) for three categories: organic enrichment, sediment chemistry, and macrofauna. These are then given weightings and combined to calculate an overall ES score for a given location.



How is ES applied? The ES methodology has been used successfully to monitor and manage salmon farms in the Marlborough Sounds, Big Glory Bay in Stewart Island, in Akaroa Harbour, and for mussel farms in Tasman Bay and offshore Opotiki. Its application requires a number of steps, including estimating the depositional footprint of a farm and locating sites for monitoring that represent a gradient from the farm to the outer zone of likely effects. The overall ES score can be assessed against standards for these locations (e.g. ES < 4 beyond a set distance from a farm). In some cases, standards may also be set for individual indicators or parameters (e.g. levels of trace metals and sulphides). Individual variables and indictors that contribute to the overall ES score have value in their own right, and play an important role in interpreting results, making the ES methodology a 'weight of evidence' approach.

More information can be found in the Ministry for Primary Industries Aquaculture Guidance series online (http://www.mpi.govt.nz/).

Table 1. Narrative criteria describing seven enrichment stages (ES) used by Keeley et al. 2012a, 2012b. Table was modified from Macleod & Forbes (2004) and Pearson & Rosenberg (1978).

ES	General description	Environmental characteristics
1	Natural/pristine conditions	Environmental variables comparable to unpolluted/ un-enriched pristine reference site.
2	Minor enrichment: Low level enrichment. Can occur naturally or from other diffuse anthropogenic sources.	Species richness usually greater than for reference conditions. Zone of 'enhancement' – minor increases in abundance possible. Mainly compositional change. Sediment chemistry unaffected or with only very minor effects.
3	Moderate enrichment: Clearly enriched and impacted. Significant community change has occurred.	Notable abundance increase, richness and diversity usually lower than reference. Opportunistic species (i.e. capitellids) begin to dominate.
4	Major enrichment 1: Transitional stage between moderate effects and peak macrofauna abundance. Major community change.	Diversity further reduced, abundances usually quite high, but clearly sub-peak. Opportunistic species begin to dominate, but other taxa may still persist. Major sediment chemistry changes.
5	Major enrichment 2: Highly enriched. State of peak macrofauna abundance.	Very high numbers of one of two opportunistic species (i.e. capitellids, nematodes). Species richness very low. Major sediment chemistry changes. Bacterial mat (*Beggiatoa*) usually evident. Hydrogen sulphide gas (H_2S) or out-gassing on disturbance.
6	Major enrichment 3: Transitional stage between peak and azoic.	Transitional stage between peak and azoic. Richness and diversity very low. Abundances of opportunistic species severely reduced from peak, but not azoic. Total abundance low but can be comparable to reference. Percentage total organic matter (TOM) can be very high.
7	Severe enrichment: Azoic/abiotic; sediments no longer capable of supporting macrofauna.	None, or only trace numbers of macrofauna remain. Some samples with no taxa. Spontaneous H_2S out-gassing. *Beggiatoa* usually present but can be suppressed. TOM can be very high.

Recent breakthroughs in high-throughput sequencing (HTS) technologies allow for species diversity to be estimated rapidly from small amounts (2–10 g) of sediment using a technique known as environmental DNA (eDNA) metabarcoding (Baird & Hajibabaei 2012; Taberlet et al. 2012). Metabarcoding enables the identification of organisms by personnel without taxonomic expertise through matching short gene fragments (from HTS data) to a reference sequence library. Standardised protocols can be developed and the results are defendable and auditable (Ji et al. 2013; Valentini et al. 2009). These qualities make metabarcoding a cost-effective, reliable and rapid option to meet the increasing need for large-scale environmental assessments. Other advantages of using molecular-based tools include a considerable reduction in the amount and scale of equipment required for collecting

material in the field (Pochon et al. 2015a, 2015b, 2015c; Pawlowski et al. 2016b), and the easy maintenance of large genomics archives due to the small size of DNA/RNA extracted products. These products can be stored indefinitely and are then available in the future for use with constantly evolving molecular technologies.

Over the last four years the Cawthron Institute's researchers have gained an in-depth knowledge of eDNA/eRNA metabarcoding for characterising biodiversity change of microscopic unicellular organisms called foraminifera (or 'forams') along benthic enrichment gradients associated with New Zealand's finfish farming (Pochon et al. 2015a) and oil/gas production (Laroche et al. 2016). Foraminifera were found to be relatively sensitive to high (> ES 5) enrichment stages with the overall presence/abundance of bioindicators decreasing in proximity to fish farms (Pochon et al. 2015a, 2015b). This highlighted the need to explore additional biological indicators for monitoring environments with high organic enrichment. Dowle et al. (2015) showed that the use of metabarcoding that targeted bacterial communities along benthic enrichment gradients was very effective in detecting bacterial community shifts, particularly in cases where ES was greater than 5. Additionally, Pochon et al. (2016) recently demonstrated that bacterial metabarcoding can effectively complement the traditional ES method in its ability to detect potential stressor footprints associated with mussel farming in the Firth of Thames (New Zealand).

The Cawthron Institute was tasked by a range of private and government agencies and industry partners to develop a multi-trophic Metabarcoding Biotic Index (mt-MBI). Contributing organisations included: Seafood Innovation Ltd, New Zealand King Salmon Ltd, Marlborough District Council, Waikato Regional Council, The Ministry for Primary Industry (Aquaculture Unit), and Ngāi Tahu Seafood.

The overarching goal of this project was to develop and test the robustness of the mt-MBI method as an alternative to traditional macrofaunal benthic monitoring in New Zealand. To achieve this objective we analysed 105 sediment samples collected over three years from a range of high- and low-flow salmon farms in New Zealand, and developed an mt-MBI, incorporating multiple bioindicator organisms from three taxonomic groups: foraminifera (For), bacteria (Bac), and general eukaryotes (Euk). The project had three research aims: (1) to investigate the ability of the method to reduce analytical costs through data multiplexing while still accurately capturing the biological diversity present in sediment samples across the three groups; (2) to assess whether the spatio-temporal community composition observed within each taxonomic group varied similarly along enrichment gradients of fish farms; and (3) to use a multi-variate weight-of-evidence-based approach (Keeley et al. 2012a, 2012b) to identify key molecular Eco-Group taxa that consistently correlated with traditional ES, and use these to validate the mt-MBI using both eDNA and eRNA datasets.

2. METHODS

2.1. Samples

Using a Van Veen Grab sampler (surface area 0.1 m^2), a total of 105 sediment samples were collected from three Chinook salmon (*Onchorhyncus tshawytscha*) farms in New Zealand (Figure 1), corresponding to 35 stations in triplicate. Farms were positioned between 25 m and 35 m above the seabed and experienced different water flow regimes, but each encompasses a wide cross-section of annual feed inputs and associated impact (Keeley et al. 2012a, 2012b). Sampling stations were positioned directly alongside the cages, along an enrichment gradient radiating outward in a semi-linear trajectory from the cages (0–200 m), and at control sites (Table 2, Figure 1). Samples (n = 90) from two farms in the Marlborough Sounds (Otanerau [OTA] and Te Pangu [TEP]) were collected at the same stations (n = 15 stations per farm) over a period of 3 consecutive years (2012–2014), always in early November. The OTA farm had a mean current velocity of < 7 $cm \cdot s^{-1}$ at 20 m water depth (approximately mid-water) and is hereafter referred to as a 'low-flow' farm. The TEP farm had a mean velocity of > 15 $cm \cdot s^{-1}$ and is therefore described as a 'high-flow' farm. Both TEP and OTA are considered to be representative of the other high- and low-flow farms in New Zealand's main salmon farming region (Keeley et al. 2012a, 2012b). Samples (n = 15) from the third farm (Big Glory Bay [BGB], Stewart Island) were collected in November 2014. The BGB farm is considered 'low-flow'.

Molecular samples were collected in parallel with physico-chemical and taxonomic surveys of macrofauna to enable a comparison of the metabarcoding tool with the traditional ES index. Specifically, each sediment grab was sampled for sediment grain size, total organic matter, total free sulphide, redox potential, and macrofaunal count data (MCD), and DNA/RNA isolation, as described in Pochon et al. (2015a, 2015b). The isolation of sediment material destined for molecular analyses was performed using gloves and sterilised individual spatulas, and the material directly transferred from an intact sediment core into an appropriately pre-labelled sterilised tube, representing either *(*a) a 5-10 mL cryotube for direct cryopreservation into liquid nitrogen or (b) a 10-50 mL tube containing an appropriate amount of DNA/RNA isolation buffer (e.g. RNALater™ or MoBio LifeGuard™). The material (2–5 g) was always collected from the top 1–2 cm of an intact sediment layer, in order to best capture the overall biodiversity of micro- and macro-organisms occurring at the sampled location.

Additionally, to account for potential spatial patchiness of micro- and macro-organisms in the sediment, samples were collected from three distinct sediment grabs per station (Table 2). Isolated samples were transported to the laboratory and immediately stored at -80°C until further processing.

Table 2. Summary of samples, sampling events and datasets utilised in this study, all of which were analysed for DNA and RNA for taxa groups foraminifera (For), eukaryota (Euk) and bacteria (Bac), in conjunction with all other environmental variables, including macrofauna count data (MCD). Blue cells highlight the 'Eco-Group assignment dataset' (comprising three subsets: MS Y1, MS Y2, SI Y2) and yellow cells highlight the 'validation dataset' (comprising MS Y3). Regions: Marlborough Sounds = MS; Stewart Island = SI. Farms names: Otanerau = OTA, Te Pangu = TEP, Big Glory Bay = BGB. Ctl = control locations.

Region	Farm-Station	Distance from cage (m)	Water Flow	Sample ID# Year 1 (Y1, 2012)	Sample ID# Year 2 (Y2, 2013)	Sample ID# Year 3 (Y3, 2014)
Queen Charlotte Sound, Marlborough Sounds (MS)	OTA Ctl3	4,000	Low	S10, S11, S12	S43, S44, S45	S73, S74, S75
	OTA Ctl4	625	Low	S13, S14, S15	S40, S41, S42	S70, S71, S72
	OTA 150	150	Low	S7, S8, S9	S37, S38, S39	S67, S68, S69
	OTA 50	50	Low	S4, S5, S6	S34, S35, S36	S64, S65, S66
	OTA Cage	0	Low	S1, S2, S3	S31, S32, S33	S61, S62, S63
	TEP Ctl1	1,870	High	S28, S29, S30	S58, S59, S60	S88, S89, S90
	TEP Ctl2	1,200	High	S25, S26, S27	S55, S56, S57	S85, S86, S87
	TEP 200	200	High	S22, S23, S24	S52, S53, S54	S82, S83, S84
	TEP 60	60	High	S19, S20, S21	S49, S50, S51	S79, S80, S81
	TEP Cage	0	High	S16, S17, S18	S46, S47, S48	S76, S77, S78
Big Glory Bay, Stewart Island (SI)	BGB Ctl1	400	Low	N/A	S103, S104, S105	N/A
	BGB Ctl2	300	Low	N/A	S100, S101, S102	N/A
	BGB 150	150	Low	N/A	S97, S98, S99	N/A
	BGB 50	50	Low	N/A	S94, S95, S96	N/A
	BGB Cage	0	Low	N/A	S91, S92, S93	N/A

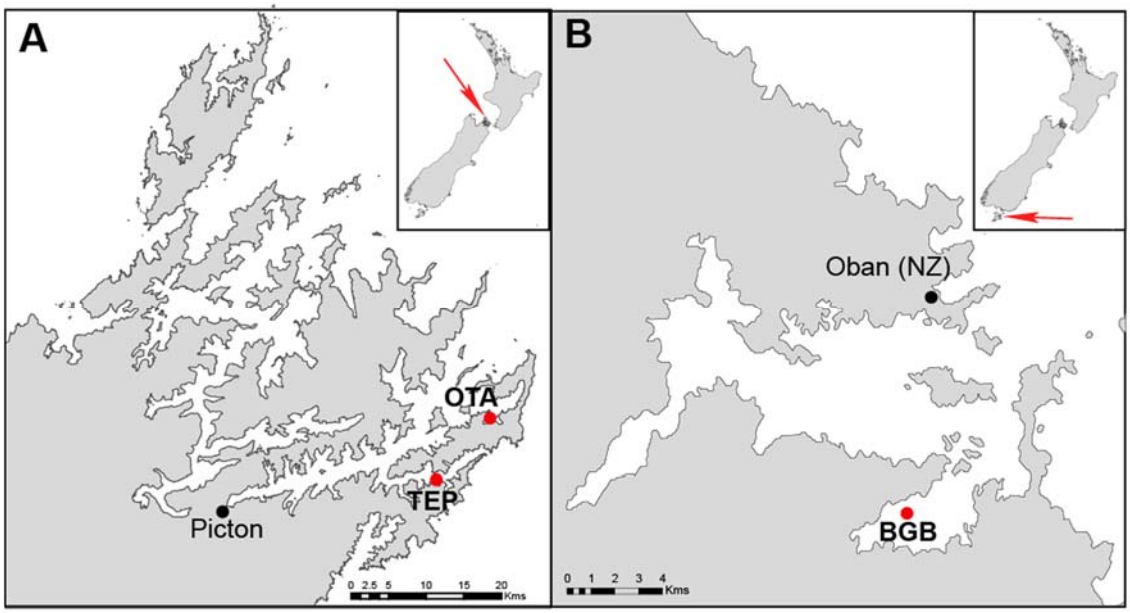


Figure 1. Site maps indicating: (A) study sites (salmon farm locations) within the Marlborough Sounds. Farm names: Otanerau = OTA, Te Pangu = TEP; (B) the salmon farm location within Big Glory Bay (BGB), Stewart Island.

2.2. Environmental DNA and RNA extractions, cDNA synthesis

Total RNA and DNA content of ca. 2 g of sediment was extracted from each thawed sediment sample using the PowerSoil™ Total RNA Isolation Kit and DNA Elution Accessory Kit (MoBio, USA) in RNase-free conditions and according to the manufacturer instructions. The quality and purity of isolated RNA and DNA were checked on 1.5% agarose gels and using a Nanophotometer (Implen, Munich, Germany). Trace DNA molecules carried over in RNA extracts were eliminated by two sequential DNase treatments as described in Langlet et al. (2013). The treatment efficiency was verified by running a 50 cycle PCR using DNA-specific foraminiferal forward primer s14F3 [5′-ACGCA(AC)GTGTGAAACTTG] and reverse primer s15.3 [5′-CCTATCACATAATCATGAAAG], following Pawlowski et al. (2014b). Extracted RNA was reverse transcribed using the SuperScript® III reverse transcriptase (Life Technologies). The various extract products (RNA, cDNA and DNA) were aliquoted and stored frozen (-80 or -20°C) until further analysis.

2.3. PCR amplifications and high-throughput sequencing

A two-step tailed PCR amplicon procedure (Sreemanta & Honghua 2002) was applied to all environmental DNA (n = 105 eDNA) and cDNA (n = 105 eRNA) samples to generate Illumina MiSeq™ libraries using specific probes for capturing total bacterial, total eukaryotic, and total foraminiferal assemblages. Three distinct Polymerase Chain Reaction (PCR) analyses were performed on each eDNA/eRNA sample to

successively amplify short, ca. 200-400 base-pair (bp) fragments of the nuclear 16S rRNA bacterial gene, the nuclear 18S rRNA eukaryotic gene (V4 region), and the 37f hypervariable region of the foraminiferal 18S rRNA genes using lineage-specific primers and thermo-cycling conditions following our previous studies (Dowle et al. 2015; Pochon et al. 2015a, 2015b; Zaiko et al. 2016). The universal primers (Appendix 1) were modified to include Illumina™ overhang adaptors as described in Kozich et al. (2013). PCR amplifications were undertaken on an Eppendorf Mastercycler (Eppendorf, Hamburg, Germany) in a total volume of 50 µL using AmpliTaq Gold 360 PCR Master Mix (Life Technologies), 1 µL of each primers and 1 µL of template eDNA/eRNA.

All amplicons were diluted to a concentration of 3 ng·µL^{-1} and sent to New Zealand Genomics Limited (NZGL), University of Auckland, for library preparation. Sequencing adapters and sample-specific indices were added to each amplicon via a second round of PCR using the Nextera™ Index kit (Illumina™). Up to 192 PCR amplicons were selected and pooled together at equimolar concentration for each HTS run. Paired-end sequences (2 x 250) were generated on a MiSeq instrument using the TruSeq™ SBS kit (Illumina™). Sequence data were automatically demultiplexed using MiSeq Reporter (v2), and forward and reverse reads assigned to samples.

2.4. Bioinformatics analysis of high-throughput sequencing data

Bioinformatic analysis of metabarcoding data was performed using USEARCH and VSEARCH tools (Edgar 2010; Rognes 2015). All sequence reads were assessed for quality, and any read that contained a base where the reported Phred quality score dropped below 30 was discarded. Forward and reverse paired-end sequences were assembled independently for each sample using USEARCH. Reads were truncated prior to merging from the first base where the Q score dropped below three. Merged reads that were less than 200 bp in length for bacteria and eukaryotes and less than 100 bp for foraminifera, were discarded. The data were then filtered with VSEARCH, discarding all reads that had more than one error per assembled read (Edgar & Flyvbjerg 2015). The retained sequences in each run were then demultiplexed and the bacterial, eukaryotic, and foraminiferal datasets were concatenated into individual files. Chimeras were identified and removed from the datasets using the UCHIME algorithm (Edgar et al. 2011) in dataset mode.

Molecular ‘species’ defined here as Operational Taxonomic Units (OTUs) were generated using VSEARCH by clustering sequence reads at the 97% similarity threshold. Singletons (i.e., unique sequence reads) found fewer than ten times per entire datasets were removed. To account for differential sequencing depth among samples and target genes, the number of reads per target (rpt) was rarefied (randomly down-sampled to 10,000 rpt). In the cases where the sample/gene contained less than 10,000 sequences, all were retained in the output file. The remaining OTUs were

taxonomically assigned using three distinct sequence reference databases as follows. For bacteria, the Ribosomal Database Project (RDP) taxonomic database Version 9 (Cole et al. 2014) using the RDP classifier (Wang et al. 2007) implemented in QIIME (Caporaso et al. 2010) was used with minimum identity value set at 97%. Sequences of unknown, archaeal, or eukaryotic origin were removed. For general eukaryotes, the Protist Ribosomal 2 (PR2) database (Guillou et al. 2012) was used for taxonomic assignments of OTUs as described above. For foraminifera, an in-house database was generated including all sequences of foraminifera available in GenBank (n = 6378), Sanger sequences obtained from morphologically identified single foraminiferal cells collected from a range of New Zealand fish farm environments (n = 251; Pochon et al. 2015a), as well as previously identified foraminiferal OTUs (n = 937) from three studies (Pochon et al. 2015a, 2015b; Laroche et al. 2016). After careful curating and removal of singleton sequences, the foraminiferal database consisted of 1,131 unique sequences which were used as described above for the assignment of foraminiferal OTUs.

2.5. Multivariate analysis of molecular, physico-chemical, and morphological data

Sequence output files from bioinformatics analyses were filtered using a script written in R (R Development Core Team 2014) to remove OTUs that did not occur in both DNA and RNA datasets (i.e., at least once in one of the triplicate DNA and corresponding RNA samples collected at each site). OTUs which contained fewer than 10 reads were removed. Rarefaction curves of OTU abundance were calculated for each sample/gene using the Vegan package (Oksanen et al. 2014) in R. Global molecular diversity of For, Bac, and Euk was visualised using the plug-in DataBurst implemented in Excel (Microsoft Office version 2013 or later).

Environmental DNA and RNA datasets were explored for each taxonomic group in a variety of forms (raw and normalised) and under different transformation (fourth root and presence-absence formats). In order to investigate whether For, Bac, and Euk community assemblages shifted similarly along enrichment gradients of fish farms in both space and time, we focused on the eDNA Eco-Group (EG) assignment datasets (i.e., MS Y1, MS Y2; SI Y2; Table 2). Bray-Curtis dissimilarity matrices were computed using each fourth-root transformed eDNA read abundance (For, Bac, and Euk) dataset and using square-root transformed MCD, and the results were visualised using a two-dimensional representation of the non-metric multi-dimensional scaling (nMDS) plot in PRIMER 7 (PRIMER-E Ltd, UK). Additionally, the corresponding Pearson correlation vectors were generated to indicate the most abundant taxa and how they correlate with the selected sampling stations in space and time, using the following thresholds: (a) For, Euk, and MCD datasets correlations > 0.75, (b) Bac dataset correlations > 0.90.

For the generation of ES, a range of diversity indices (e.g., Pielou Index (J), Shannon-Weiner Index (H), richness (S) were calculated in PRIMER 7 (Clarke and Gorley 2015) using the DIVERSE function. Statistical comparisons between regression lines was conducted in R using ANCOVA (type III) with the model structures: Model1 (for slope) = aov(Y~ES*DNA_RNA, data), Model2 (for intercept when slope is not significant) = aov(Y~ES+DNA_RNA, data), and the following R packages: 'car', 'compute.es', 'effects', 'ggplot2', 'multcomp', 'pastecs', and 'WRS2'.

2.6. Quantile regression splines

The changes in taxa distributions as a function of ES values were examined using the method of Anderson (2008) and adapted by Keeley et al. (2012a, 2012b). The sequence abundances of potential bioindicator OTUs were plotted as a function of ES and quantile regression spline models were constructed for the 95th percentile (Koenker et al. 1994; Koenker 2005). This percentile fraction represents the value below which 95% of the sequences are expected (also called the $\tau = 0.95$ quantile). In accordance with Anderson (2008), all models were fitted using the function rq() combined with the function bs() implemented in the R package. The top 200 most prevalent OTUs from each of the For and Euk datasets, and top 250 for Bac, were plotted against overall ES, and 95th percentile quantile regression splines were constructed and fitted for each, with a vertical line indicating the point along the x-axis at which maximum abundance was predicted. This process was repeated for each data subsets in the EG assignment dataset (i.e. MS Y1, MS Y2, SI Y2; Table 2).

2.7. Eco-Group and GenCodeID assignment

The quantile regression splines described above provide a quantitative basis for assigning EG to OTU-based taxa akin to those described by Borja et al. (2000) for macrofauna, and have proven to be useful for developing biotic indices, e.g. AZTI Marine Biotic Index (AMBI; Borja et al. 2000), BENThic IndeX (BENTIX; Simboura 2003), MEDiterranean OCCidental (MEDOCC; Pinedo et al. 2012), and Enrichment Stage (ES; Keeley et al. 2012a). Allocation of EG from OTU-based quantile regression plots required assessment of the appropriateness of the fits of the splines and for the level of agreement between the datasets in terms of where the peaks indicated the maximum number of reads along the enrichment scale. A filtering step was then undertaken to eliminate contradictory (and therefore possibly erroneous) information and OTUs that were abundant only in one location or time point (and therefore not consistently present). Each plot (i.e., OTU) was categorised according to the level of agreement (or quality criterion) as in Table 3.

Table 3. Quality score of Eco-Group allocations based on regression splines. This ranking of spline quality allows further test combinations to optimise the index precision. ES = enrichment stage index.

Quality level	Description
0	'Excellent' ES specific bioindicator. Strongly indicates one stage on the enrichment gradient—no positive values at other end of spectrum.
1	'Very good'; close agreement between peaks (span < 1.5 ES), some positive values at other ES levels.
2	'Good'; reasonable agreement (two lines agree closely, or presents possibly useful distribution—no double peaks).
3	'Indifferent'; occurs throughout most ES levels, but might be useful as EG III which is the category for species that are considered indifferent to enrichment.
4	'Poor' / unusable; lines differ by > 3 ES, double peaks or insufficient data points with number of reads > 0.

Of the < 250 most abundant OTUs per dataset, visual inspection of the regression splines were made across space and time to identify and classify individual OTUs into these agreement categories (0-Excellent, 1-very good, 2-good, 3-indifferent, 4-poor). For these, the average position of the maximums of the peaks (along the enrichment scale on the x-axis) was recorded. This position was then related to likely EG, as follows: an OTU that only occurred in unenriched (ES1 or ES2) conditions was considered to be EG I or II accordingly. Similarly, if a species consistently occurred only in highly enriched conditions (> ES5), then it was considered to be EG V (or IV depending on the exact position of the maxima). In situations where the peaks straddled two possible groups, an alternate EG was also recorded (III). A summary of this EG allocation is shown in Table 4.

Table 4. Summary of Eco-Group (EG) categories used to calculate the Metabarcoding Biotic Index (MBI). ES = enrichment stage index.

EG	Definition
I	Clearly unenriched ES1–2.2. Peak far left. Little sign of decreasing below ES2.
II	Low level enrichment ES2.2–3. Peak around ES 2.5.
III	Ubiquitous / tolerant / transitory. Two types here distinguished by 'quality': Q3 = indifferent (spread across gradient), and Q0-2: peak evident around ES3—i.e. defines ES3, rather than being across all groups. Clearly increases from ES2–3.
IV	Organisms that are found in transitory conditions between moderate and highly enriched. Peak occurs before ES5–ES5+ values are lower.
V	Opportunistic taxa dominant in highly enriched conditions—ES5. Peaks at or > ES5. Ideally maxima occurs at far right (ES6).

For each taxonomic group independently (i.e., For, Bac, and Euk datasets) and for each of the < 250 most abundant regression splines per dataset, the number of EG

categories and corresponding quality scores were recorded (see results). The validity of the approach was then checked by using the EG classifications to calculate an mt-MBI value for each investigated sample, as described below.

Finally, a unique genetic barcode, referred to as GenCodeID, was assigned to each representative OTU that was assessed to have a good quality EG assignment (assignment quality 0, 1 or 2). This incorporates four aspects: the first two letters of the taxa group name (i.e. 'fo' for foraminifera, 'ba' for bacteria, and 'eu' for eukaryotes), a unique number from a continuous series from an ordered list of good quality OTUs for that taxa group (e.g. '0001' for the first in the list), the assessed assignment quality for that OTU, and the assigned Eco-Group; e.g. 'eu-0003-1-IV' = eukaryote dataset, third OTU in database, Quality Level 1, EG IV. The isolation of all good quality barcoded sequences from each taxonomic group represent very valuable reference databases against which future environmental samples can be directly screened for rapid and automatic EG identification and mt-MBI calculation.

2.8. Multi-trophic index development

The mt-MBI proposed here is based on the established AMBI structure described in Borja et al. (2000), modified and optimised for HTS-derived For, Bac and Euk read abundance data for both RNA and DNA. To do this, a model was set up to calculate AMBI values based on square-root transformed read abundance data, such that the traditional weightings of each of the EGs (i.e. 0 x %EGI + 1.5 x %EGII + 3 x %EGIII + 4.5 x %EGIV + 6 x %EGV) could be altered to optimise for scatter and linearity when regressed against overall ES. After testing many permutations, the optimal index structure for the Metabarcoding Biotic Index (hereafter referred to as 'MBI') was as follows:

MBI = ((0 x %EGI) + (1.5 x %EGII) + (3 x %EGIII) + (5 x %EGIV) + (12 x %EGV))/100

where EG = Eco-Group.

The effect of the inclusion of OTU allocations with varying levels of agreement and of the 'indifferent' group was also compared by deriving index values that used different combinations of the groups (e.g., Q124 = quality criteria 1,2 and 4, Q12 = quality criteria 1 and 2, etc). The optimum combination was generally, Q012 and as such MBI calculations that are presented here use all Quality 0, 1 and 2 EG assignments.

The MBI was first calculated for each sample and for each taxonomic group separately, resulting in three distinct indices: foraminiferal-MBI (f-MBI), bacterial-MBI (b-MBI), and eukaryotic-MBI (e-MBI). MBI values were calculated based on the established EG and the methods outlined above and the sample sets detailed in

Table 2, first using the EG assignment data (MS Y1, MS Y2, SI Y2), and then on the separate validation dataset (MS Y3).

Finally, to calculate the mt-MBI required integration of the different taxonomic groups. Because each taxonomic dataset has different properties in terms of numbers of sequence reads and the relative distributions within, the three datasets (For, Bac, and Euk) needed to be standardised and / or normalised before combining. After testing several standardisation methods, it was established that the most severe transformation, presence-absence, was necessary to maximise consistency between the resulting MBI values for the different taxonomic groups. It has previously been demonstrated that presence-absence of frequently occurring macrofauna based on metabarcoding data can provide reliable AMBI values (Aylagas et al. 2014). This facilitated combining of the different taxonomic datasets to be used in conjunction with combined EG datasets to produce an mt-MBI.

3. RESULTS

3.1. Multiplexing of molecular samples and global biodiversity

From the 105 sediment samples investigated, a total of 630 PCR amplicons (105 samples x 3 genes x 2 molecules DNA/RNA) were produced and sequenced in multiplex over four MiSeq Illumina™ sequencing runs, resulting in a total of 19,859,791 high-quality sequences (Table 5). Following bioinformatics analysis the number of OTUs (i.e. molecular species) for each taxonomic dataset were: 6,850 bacterial OTUs, 3,281 eukaryotic OTUs and 1,182 foraminiferal OTUs (Table 5). Rarefaction curves analyses showed that all investigated samples reached saturation (Appendix 2), except for 21 foraminiferal samples (S47, S50 [DNA dataset], S4, S25-S26, S28, S37-S41, S47, S49, S59-S60, S79 [RNA dataset]; Appendix 2A) and 6 bacterial samples (S38, S40, S41, S43-S45 [RNA dataset]; Appendix 2B). Despite the limited sequencing depth, these samples were maintained for downstream analyses to assist in identifying bioindicator taxa.

Table 5. Number of sequences and Operational Taxonomic Units (OTUs; 'molecular species') obtained for each studied taxonomic group before and after bioinformatics filtering. For = foraminifera, Bac = bacteria, Euk = eukaryotes.

Sequence Datasets	Number of raw reads	Average number of clean reads per sample (+/- standard deviation)	Number of high-quality reads	Number of OTUs (DNA/RNA filtered)
Entire dataset	28,363,494	94,570 +/- 41,696	19,859,791	N/A
For	7,972,476	35,289 +/- 27,417	7,410,743	1,182
Bac	8,758,155	27,686 +/- 14,019	5,814,261	6,850
Euk	11,632,863	31,594 +/- 13,243	6,634,787	3,281

3.1.1. Foraminiferal biodiversity

The 1,182 foraminiferal OTUs clustered in 3 classes, 19 orders, 68 families, 91 genera and 116 species. Figure 2 shows the overall foraminiferal biodiversity from all sites combined. The majority of OTUs corresponded to Monothalamea, a class of foraminifera with unilocular skeletons (72%), followed by Globothalamea (multilocular foraminifera; 26%), Tubothalamea (tubular foraminifera; 1%), and other (1%). Within Monothalamea, the biodiversity was split evenly between Astrorhizida (36%) and Allogromiida (34%), while Globothalamea was dominated by Rotaliida (21%). The five most dominant genera were represented by *Vellaria*, *Psammophaga*, *Rosalina*, *Planoglabratella*, and *Bathysiphon* (Figure 2).

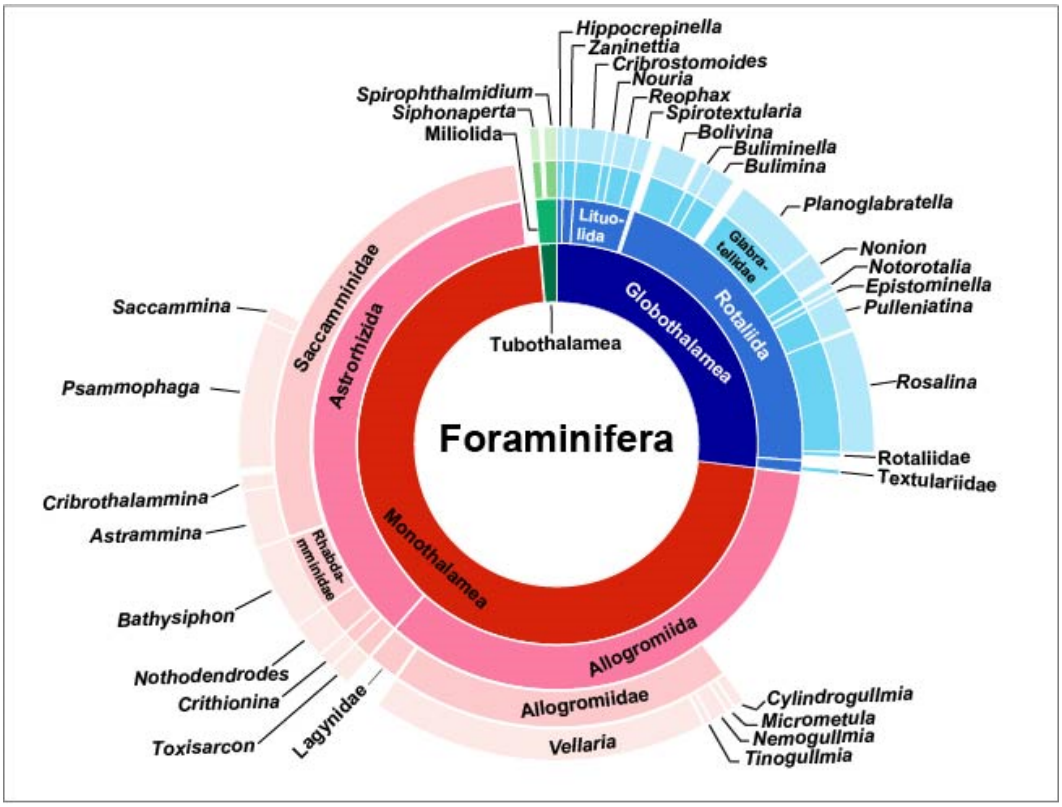


Figure 2. Global biodiversity of foraminiferal Operational Taxonomic Units from all sites combined. The chart shows the abundance of sequences from highest to lowest taxonomic levels (phylum = inner circle to genera = outer circle).

3.1.2. Bacterial biodiversity

The 6,850 bacterial OTUs clustered in 38 phyla, 78 classes, 133 orders, 254 families and 685 genera. The majority of OTUs corresponded to the phyla Proteobacteria (74%), followed by Bacteroidetes (7%), Acidobacteria (5%), Firmicutes (3%), Planctomycetes (2%), and other (9%; Figure 3). The phylum Proteobacteria was dominated by Gammaproteobacteria (38%) and Deltaproteobacteria (23%). The four most dominant genera were *Sulfurovum*, *Thioprofundum*, *Vibrio*, and *Desulfosarcina*.

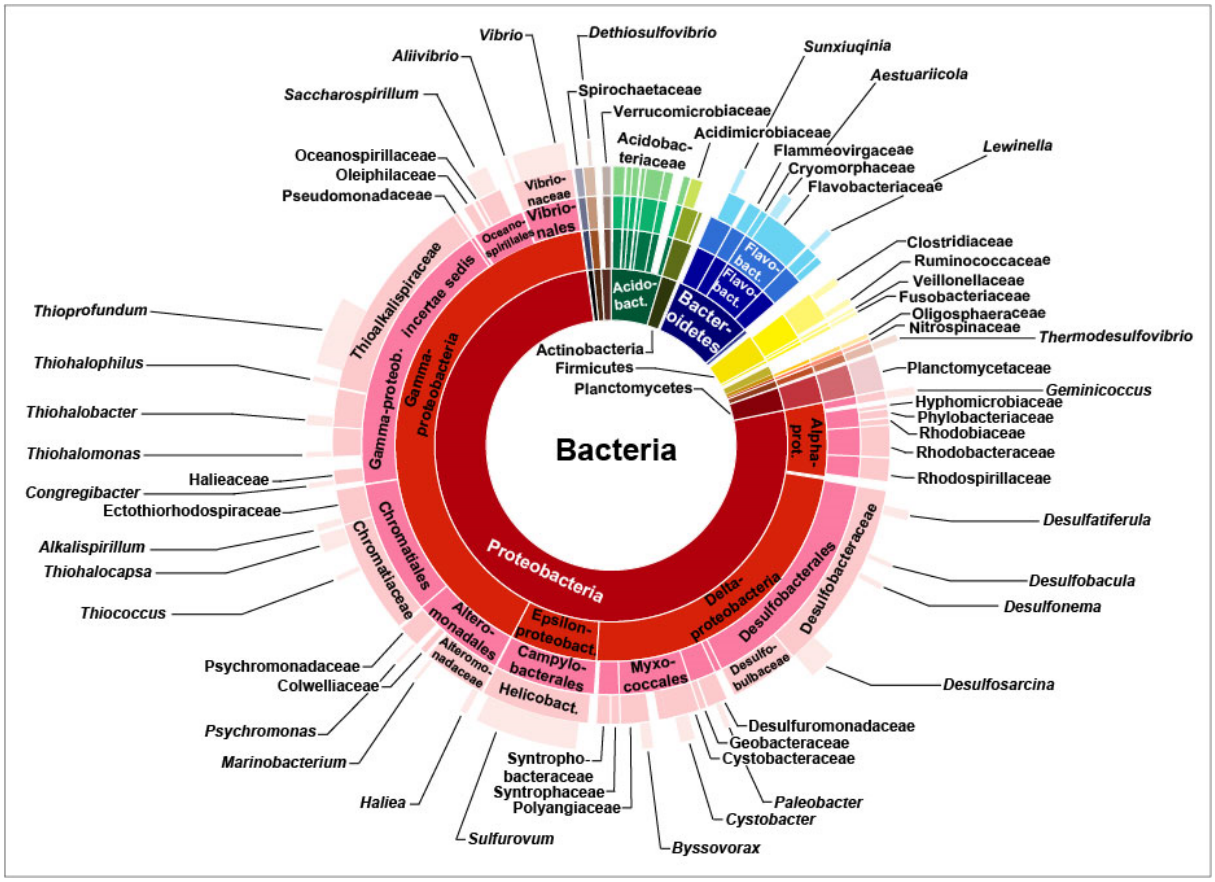


Figure 3. Global biodiversity of bacterial Operational Taxonomic Units from all sites combined. The chart shows the abundance of sequences from highest to lowest taxonomic levels (phylum = inner circle to genera = outer circle).

3.1.3. Eukaryotic biodiversity

The 3,281 eukaryotic OTUs clustered in 6 supergroups, 33 phyla, 106 classes, 176 orders, 332 families, 744 genera, and 885 species. The majority of OTUs corresponded to the phylum Metazoa (75%), followed by Ciliophora (14%), Cercozoa (3%), Apicomplexa (3%), Dinophyta (2%), and other (3%; Figure 4). The four most dominant genera were *Capitella*, *Sabatieria*, *Microstomum*, and *Tracheloraphis*.

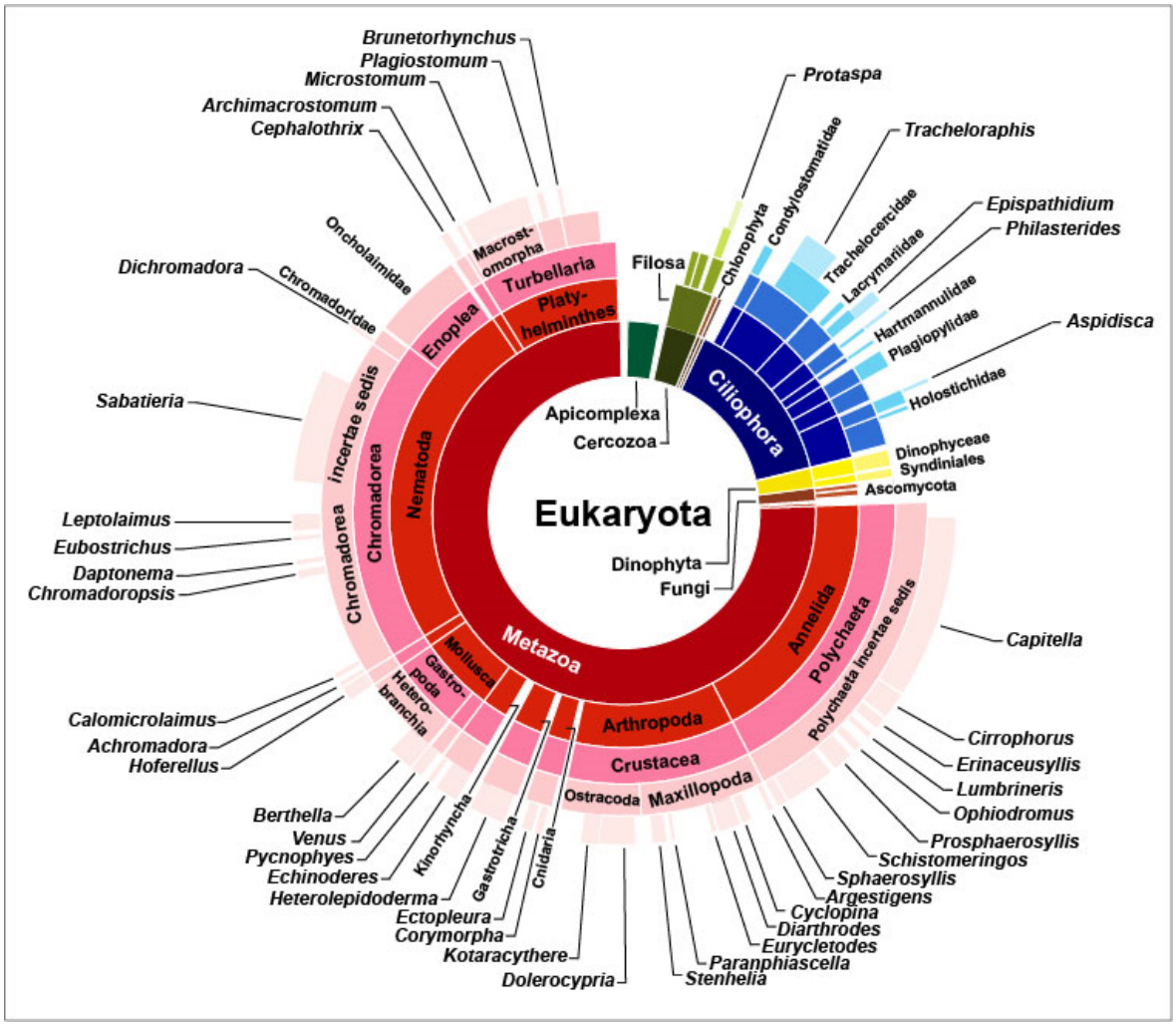


Figure 4. Global biodiversity of eukaryotic Operational Taxonomic Units from all sites combined. The chart shows the abundance of sequences from highest to lowest taxonomic levels (phylum = inner circle to genera = outer circle).

3.2. Datasets comparison - community composition shifts along enrichment gradients of salmon farms

The validity of constructing an mt-MBI was tested initially by displaying how each taxonomic community responds individually to the proximity to farms, and therefore to organic enrichment levels. Figure 5 shows that the multivariate analysis of the fourth-root transformed For, Bac, and Euk read abundance datasets and square-root transformed MCD produced comparable arrangements of the samples in two-dimensional space. In each plot there was a clear right to left shift of samples in response to proximity to farm (distance, m), and therefore organic enrichment levels. Enrichment increases from natural / unenriched samples on the right of the plots to highly enriched near-farm (0 m) samples on the left (Figure 5A-D). As such, the

composition of the eDNA-based OTUs within each taxonomic group changed in a predictable manner in accordance with enrichment.

Farm-related (and current flow) differences (expressed predominantly on the y-axis of the MDS plot) were, more evident in the MCD, For and Euk datasets, than in the Bac dataset (Figure 5A-D). Notably, in the two Marlborough Sounds (MS) farms, one high-flow (TEP) and the other low-flow (OTA) shared very similar bacterial communities across enrichment gradients. In general the Stewart Island (SI, BGB) metabarcoding datasets (For, Bac, Euk) from the separate geographic region, did not cluster with MS datasets, despite responding to enrichment in a similar manner. This contrasted with the MCD data where the SI data grouped with the 'low-flow' farms, regardless of the large distance separating the two regions. The differences between surveys (Y1 and Y2) were small relative to the differences between farms and between stations.

The correlation vectors associated with the MDS plots (Figure 5E-H) indicated that most datasets had dominant taxa that correlated strongly with both decreasing and increasing enrichment as well as with increasing flow. For example in the For dataset (Figure 5E), some taxa correlated most strongly with enrichment levels (corresponding to the x-axis of the MSD plot) from un-enriched (e.g., *Buliminella tenuata* [barcodes fo-0069-1-II], *Cribrothalamina alba*, and an un-identified Sacamminidae [fo-122-4]), to highly-enriched environments (e.g., *Vellaria pellucidus* [fo-0005-2-IV]). In contrast, other foraminiferal taxa correlated more with flow and/or regional differences (e.g., *Planoglabratella opercularis* and *V. pellucidus* [fo-124-4]). Similar patterns were observed in the Euk metabarcoding (Figure 5G) and MCD (Figure 5H) datasets, although the latter one did not seem to be affected by decreasing flow. In contrast, bacteria communities (Figure 5F) showed very distinct delineations in response to enrichment, either strongly decreasing or increasing, and with very little compositional changes in response to flow. For example, bacterial communities associated with enriched sediments were largely dominated by anaerobic, sulphate-reducing Desulfobacterales and microaerophilic Campylobacterales. In contrast, un-enriched sediments were characterised by a larger diversity of bacterial assemblages representing at least thirteen different orders (data not shown), and were dominated by Gammaproteobacteria.

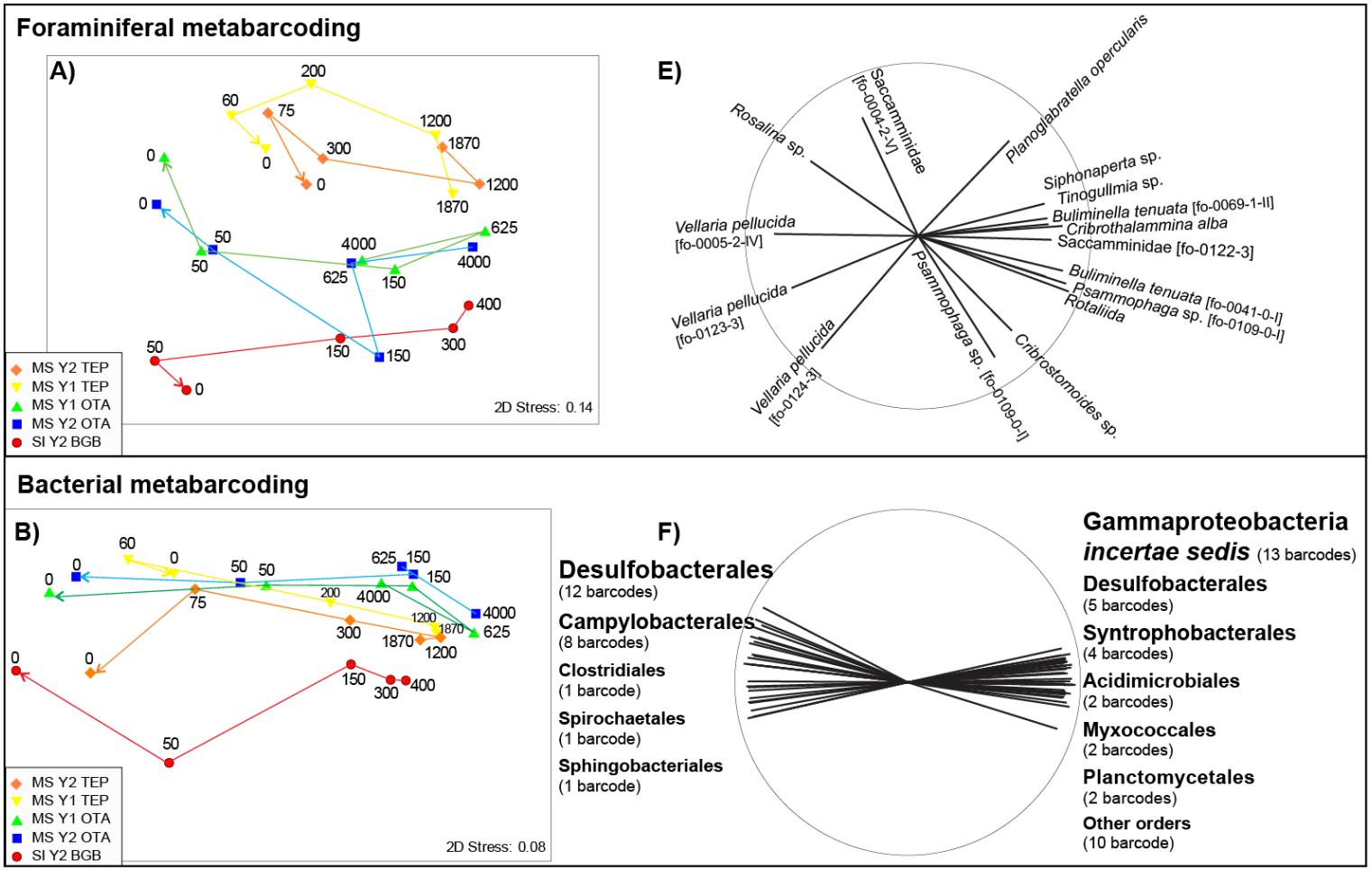


Figure 5. Two-dimensional Multi-Dimensional Scaling plots based on fourth-root transformed environmental DNA read abundance data for: (A) foraminifera (For), (B) bacteria (Bac), (C) eukaryotes (Euk), and (D) square-root transformed macrofauna count data (MCD). Colour lines correspond to the fish farms Te Pangu (TEP, yellow and orange), Otanerau (OTA, green and blue) and Big Glory Bay (BGB, red) investigated during year 1 (Y1) and year 2 (Y2). Numbers correspond to the distance to the fish cages in metres.

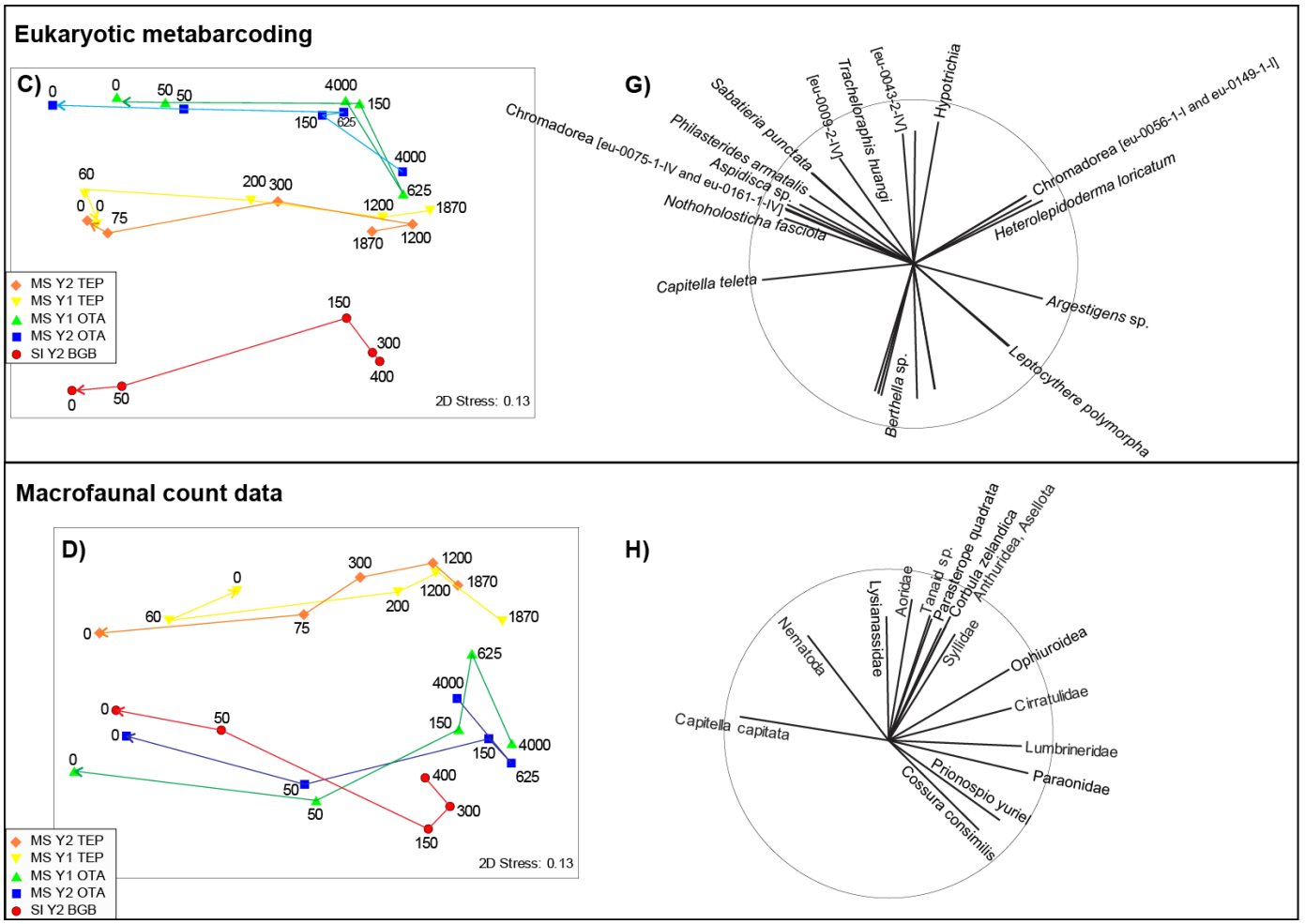


Figure 5, continued. Vector plots (E, F, G, H) indicate taxa (species or OTUs assigned at the lowest taxonomic level) that had a Pearson correlation > 0.75, with the exception of bacteria, which displays those at >0.90. The names in square brackets correspond to specific GenCodeIDs (see Appendix 3); these are only indicated in this figure for distinct OTUs assigned to identical taxa names. Taxa that were not identified as high-quality Eco-Group (EG) received a GenCodeID with indifferent/unusable quality score and without EG status (e.g., fo-0123-3).

Macrofauna Count Data (MCD) in the enriched, near-farm sediments were typically associated with high abundance of nematodes and capitellid polychaetes (Figure 5H). Analogous Euk metabarcoding responses were observed (Figure 5G) as shown by the high abundance of polychaetes (e.g., *Capitella teleta*), and nematodes (un-identified Chromadorea [eu-0075-1-IV and eu-0161-1-IV] and *Sabatieria punctata*). Additionally, ciliates were well represented in enriched samples using Euk metabarcoding, as evidenced by the co-dominance of the marine hypotrichous ciliate *Notoholosticha fasciola* and the Scuticociliatia *Philasterides armatalis*.

The more distant, unimpacted sediments were typified by Ophiuroidea, Cirratulidae, Lumbrineridae and Paraonidae in the MCD analysis (Figure 5H), while metabarcoding samples were dominated by Chromadorea nematodes [eu-0056-1-I and eu-0149-1-I], the copepod *Argestigens* sp., and the microscopic pseudocoelomate gastrotrich *Heterolepidoderma loricatum* (Figure 5G).

The comparison of Spearman rank correlation statistics (Rho) between corresponding pairs of For, Bac, Euk, and MCD resemblance matrices, showed that the weakest correlations were obtained with the foraminiferal data (0.591; Table 6). Good correlations were obtained between MCD-Euk (0.747) and MCD-Bac (0.795), and the strongest correlation was obtained between Bac-Euk (0.817).

Table 6 Rho statistics indicating Spearman rank correlation between Bray-Curtis resemblance matrices presented in the multi-dimensional scaling plots in Figure 5. Number of permutations = 9999, statistical significance <0.001 in all cases. For = foraminifera, Bac = bacteria, Euk = eukaryotes, and MCD = macrofaunal count data.

	For	Euk	Bac
Euk	0.784		
Bac	0.69	0.817	
MCD	0.591	0.747	0.795

3.3. Identification of molecular Eco-Groups and comparison of eDNA/eRNA datasets

The first step of identifying molecular EGs for the design of the MBI involved the systematic visualisation of ecological preferences and spatio-temporal consistency/stability of the 200 most prevalent OTUs from each of the foraminiferal and eukaryotic datasets, and the 250 OTUs for bacteria group. A representative set of regression splines (top 25 most abundant) for each taxonomic group and datasets (DNA vs RNA) are given in Appendix 4(A-C). For each investigated OTU, three individual regression splines were generated (Y1 MS, Y2 MS and Y2 IS); the analysis of how consistent or different the three splines plotted against the enrichment stage allowed classification of EG score allocation. Overall, there was a general agreement between EG assignments for eDNA/eRNA data supporting the validity of the process. For example, some OTUs yielded spatio-temporal splines that were in very close

agreement with Enrichment Stage from both eDNA and eRNA datasets and therefore received a high-quality (Q0-Q2) EG allocation score (e.g., foram OTU 2525 [Allogromiida], Appendix 4A; bacterial OTU 6441 [*Thioprofundum* sp.], Appendix 4B; Eukaryotic OTU 3842 [*Cirrophorus lyra*], Appendix 4C). Other OTUs yielded dissimilar spline trends from both eDNA and eRNA data, receiving a low (Q3-Q4) EG allocation score (e.g., OTU 1600 [*Vellaria pellucidus*], Appendix 4A; OTU 22713 [*Sulfurovum* sp.], Appendix 4B; and OTU 175 [*Schistomeringos rudolphi*], Appendix 4C). There were also a few cases where eDNA and eRNA datasets produced dissimilar regression spline trends (e.g., OTU 2586 [Allogromiida], Appendix 4A).

Some good examples of high-quality foraminiferal EG-sensitive to enrichment were *Buliminella tenuata* [fo-0041-0-I] and *Psammophaga* sp. [fo-0109-0-I], while *Bathysiphon* sp. [fo-0011-I-IV] and *Vellaria pellucidus* [fo-0030-I-IV] were more common at increased levels of enrichment (Figure 6). *Bolivina* sp. [fo-0129-1-V], *Spirophthalmidium* sp. [fo-0084-1-V] and *Rosalina* sp. [fo-0038-0-V] were less frequently observed, but only occurred in highly (ca. ES 4) or very highly (ca. > ES 5) enriched conditions and as such were considered EG IV or EG V OTUs.

Bacteria were, on the whole, far more prolific in producing good quality EGs than foraminifera and other eukaryotes (Appendix 4B). There were also numerous high-quality (Q0-Q2) EG taxa that were clearly associated with certain enrichment levels (Figure 7). For example the proteobacterial species *Nitrospira* sp. [ba-0139-0-I] and *Thiococcus* sp. [ba-0187-1-I] and the Plantomycetes species *Blastopirellula* sp. [ba-0197-0-I] were very clearly associated with unenriched (ES 1-2.5) conditions (assigned EG I). Other EG taxa were abundant at moderate to high enrichment conditions (ES 4; EG IV), such for example the proteobacterium species *Desulfosarcina* sp. [ba-0122-1-IV] and the Bacteroidetes species *Aestuariicola* sp. [ba-0179-0-IV]. *Dethiosulfovibrio* sp. [ba-0136-1-V] and *Desulfobacter* sp. [ba-0113-0-V] are examples of EG taxa that were almost exclusively associated with highly-enriched sediments (> ES 5; EG V).

A number of excellent bioindicator taxa with high-quality EG scores were also identified among general eukaryotes (Figure 8). The polychaete *Odontosyllis freycinetensis* [eu-0125-0-I], the maxillapod *Bradya* sp. [eu-0098-0-I] and the small (<1 mm) invertebrate *Echinoderes setiger* [eu-0057-0-I] were examples of EG I or enrichment sensitive taxa. A platyhelminthic worm *Microstomum papillosum* [eu-0002-0-IV] was very common at moderate to highly enriched conditions (EG VI) and the first-order opportunistic polychaete *Capitella teleta* [eu-0001-0-V] was prolific at ES 4 to ES 5 (EG V). The marine ciliate *Philasterides armatalis* [eu-0064-1-V] and the copepod *Tachidius triangularis* [eu-0100-1-V] tended to favour extremely enriched conditions (ES 5 or above).

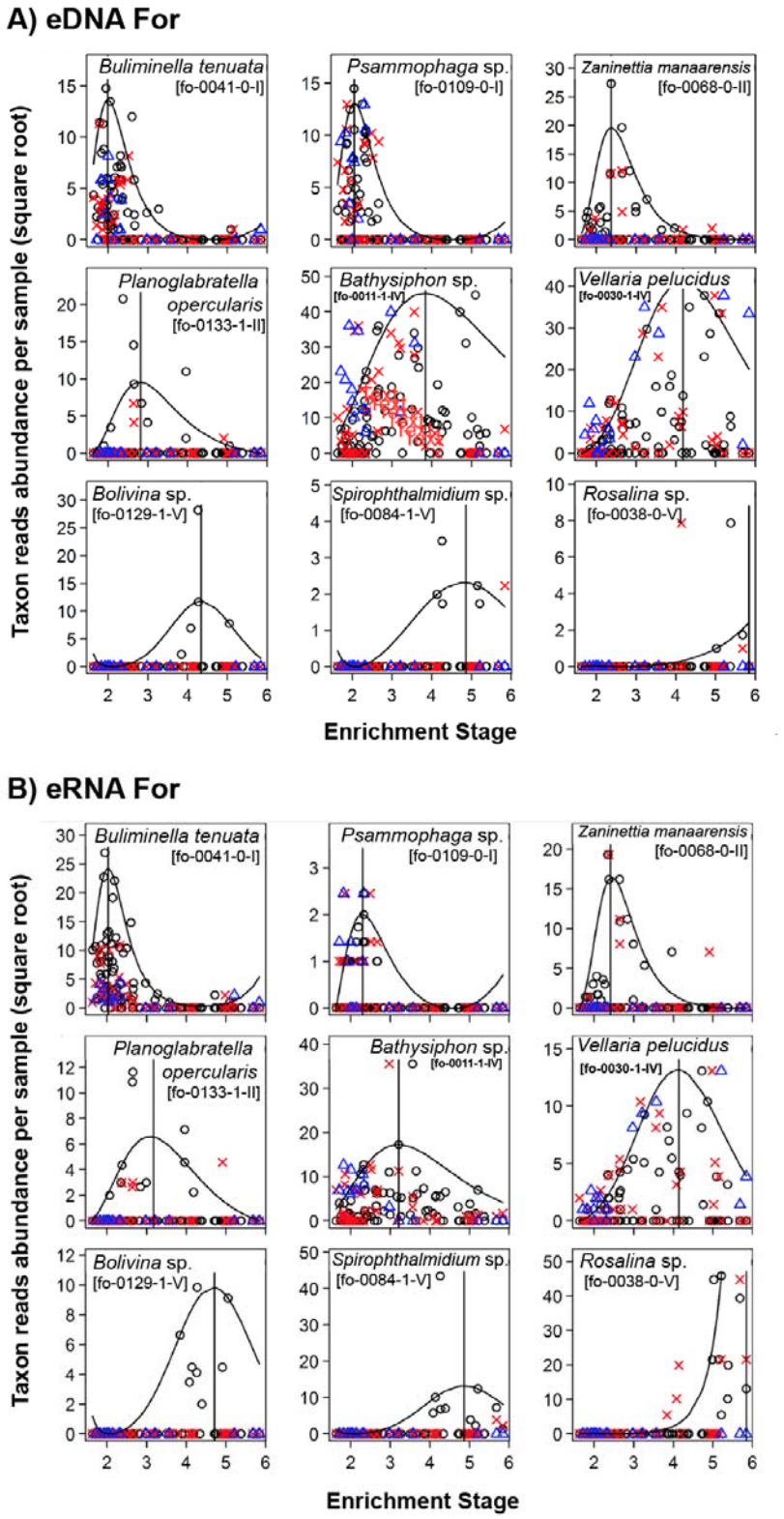


Figure 6. Relative sequence reads abundance of nine selected high-quality (Q0-Q2) foraminiferal (For) Eco-Groups taxa plotted against the Enrichment Stage, using (A) environmental DNA and (B) environmental RNA data. Black circles = MS Y1, red crosses = MS Y2, blue triangles = SI Y2. Quantile regression splines (Tau = 0.95, df = 3, Deg = 3) fitted for all data combined with vertical lines indicating position of peak read abundance on the enrichment scale.

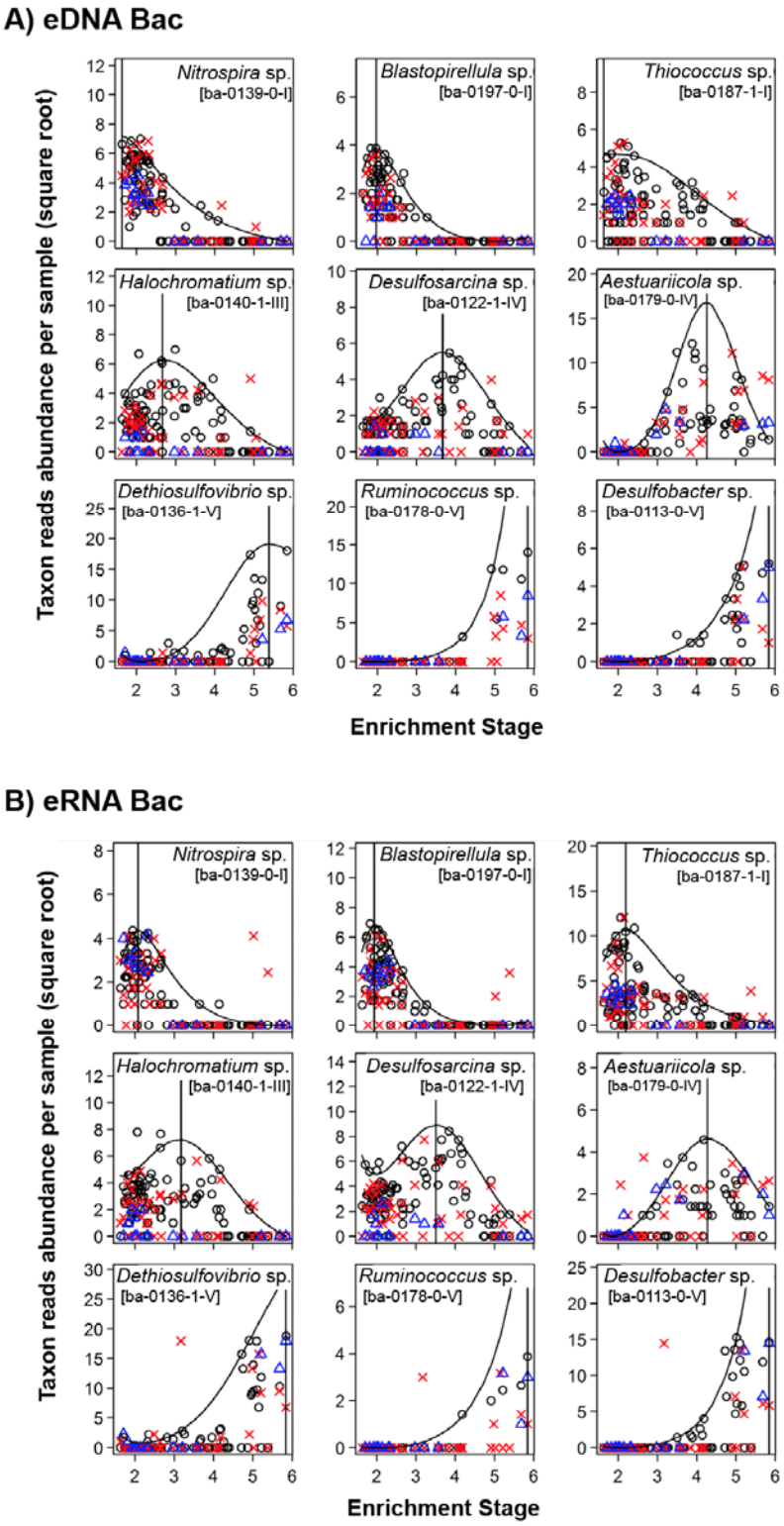


Figure 7. Relative sequence reads abundance of nine selected high-quality (Q0-Q2) bacterial (Bac) Eco-Groups taxa plotted against the Enrichment Stage, using (A) environmental DNA and (B) environmental RNA data. Black circles = MS Y1, red crosses = MS Y2, blue triangles = SI Y2. Quantile regression splines (Tau = 0.95, df = 3, Deg = 3) fitted for all data combined with vertical lines indicating position of peak read abundance on the enrichment scale.

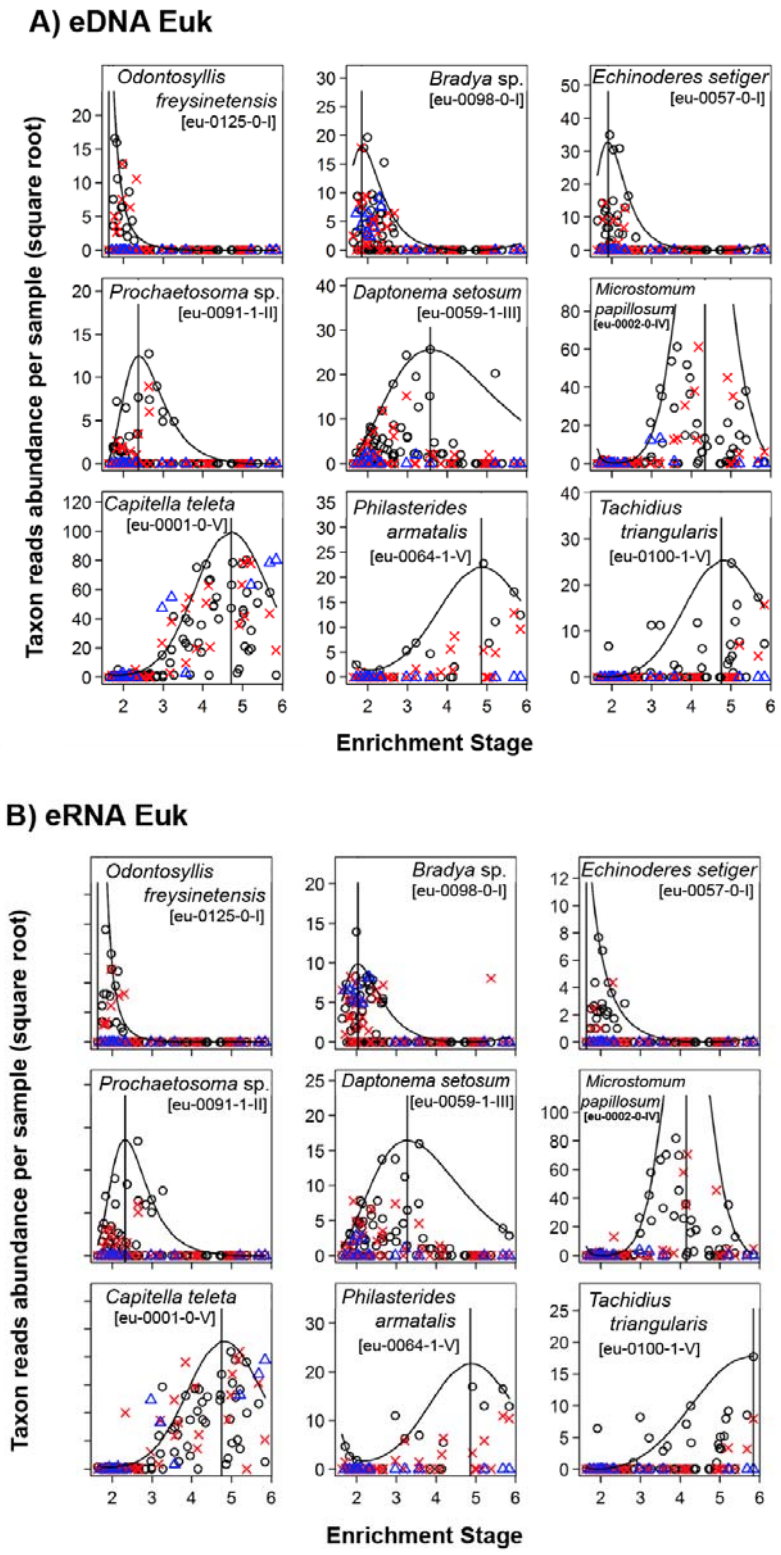


Figure 8. Relative sequence reads abundance of nine selected high-quality (Q0-Q2) eukaryotic (Euk) Eco-Group taxa plotted against the Enrichment Stage, using (A) environmental DNA and (B) environmental RNA data. Black circles = MS Y1, red crosses = MS Y2, blue triangles = SI Y2. Quantile regression splines (Tau = 0.95, df = 3, Deg = 3) fitted for all data combined with vertical lines indicating position of peak read abundance on the enrichment scale.

Out of the < 250 most abundant/prolific molecular OTUs per taxonomic group (Appendix 3 and 4; Figures 6-8), it was possible to assign a total of 551 and 555 bioindicator OTUs to 5 distinct EGs for the eDNA and eRNA datasets, respectively, across the three taxonomic groups (Table 7). The majority of bioindicator OTUs belonged to the EG categories I to III (very sensitive to ubiquitous). Furthermore, the majority of the quality scores attributed to these bioindicators (DNA dataset n = 507; RNA dataset n = 509) ranged between Q0 (Excellent) to Q2 (Good), with a minority (DNA n = 143; RNA n = 140) that were either indifferent or unusable (Table 8). Notably, the foraminiferal dataset accounted for most of the 'unusable' regression splines.

The distribution of assignments between EGs and their corresponding quality scores (Q) was very similar for eDNA and eRNA data in each taxonomic group (Table 7, Figure 9). However, the distribution of EG assignments differed among taxonomic groups, with the foraminiferal and eukaryotic datasets being strongly dominated by EGs I and II (very sensitive to sensitive) with minimal EG V (opportunistic) assignments. By contrast, the bacterial datasets were more evenly spread across the five EGs (Figure 9A and 9C). Compared to the eukaryotic and bacterial datasets, the foraminiferal dataset displayed the lowest number of good quality score allocations and the highest number of unusable EGs (Figure 9B and 9D).

Table 7. Summary of Eco-Group allocation and quality scores for each taxonomic group and molecular dataset (environmental DNA and environmental RNA).

	Foraminifera		Bacteria		Eukaryota		Total	
Eco-Group	DNA	RNA	DNA	RNA	DNA	RNA	DNA	RNA
I = Very sensitive	53	60	47	59	55	61	155	180
II = Sensitive	35	28	65	50	49	44	149	122
III = Ubiquitous	17	18	48	51	28	26	93	95
IV = Transitory	16	16	52	48	24	25	92	89
V = Opportunistic	13	15	31	39	18	15	62	69
Total	134	137	243	247	174	171	551	555

	Foraminifera		Bacteria		Eukaryota		Total	
Quality Score	DNA	RNA	DNA	RNA	DNA	RNA	DNA	RNA
0 = Excellent	13	9	9	16	32	32	54	57
1 = Very good	60	55	164	151	91	94	315	301
2 = Good	48	69	49	45	41	37	138	151
3 = Indifferent	13	4	21	35	10	8	44	46
4 = Unusable	66	63	7	3	26	29	99	94

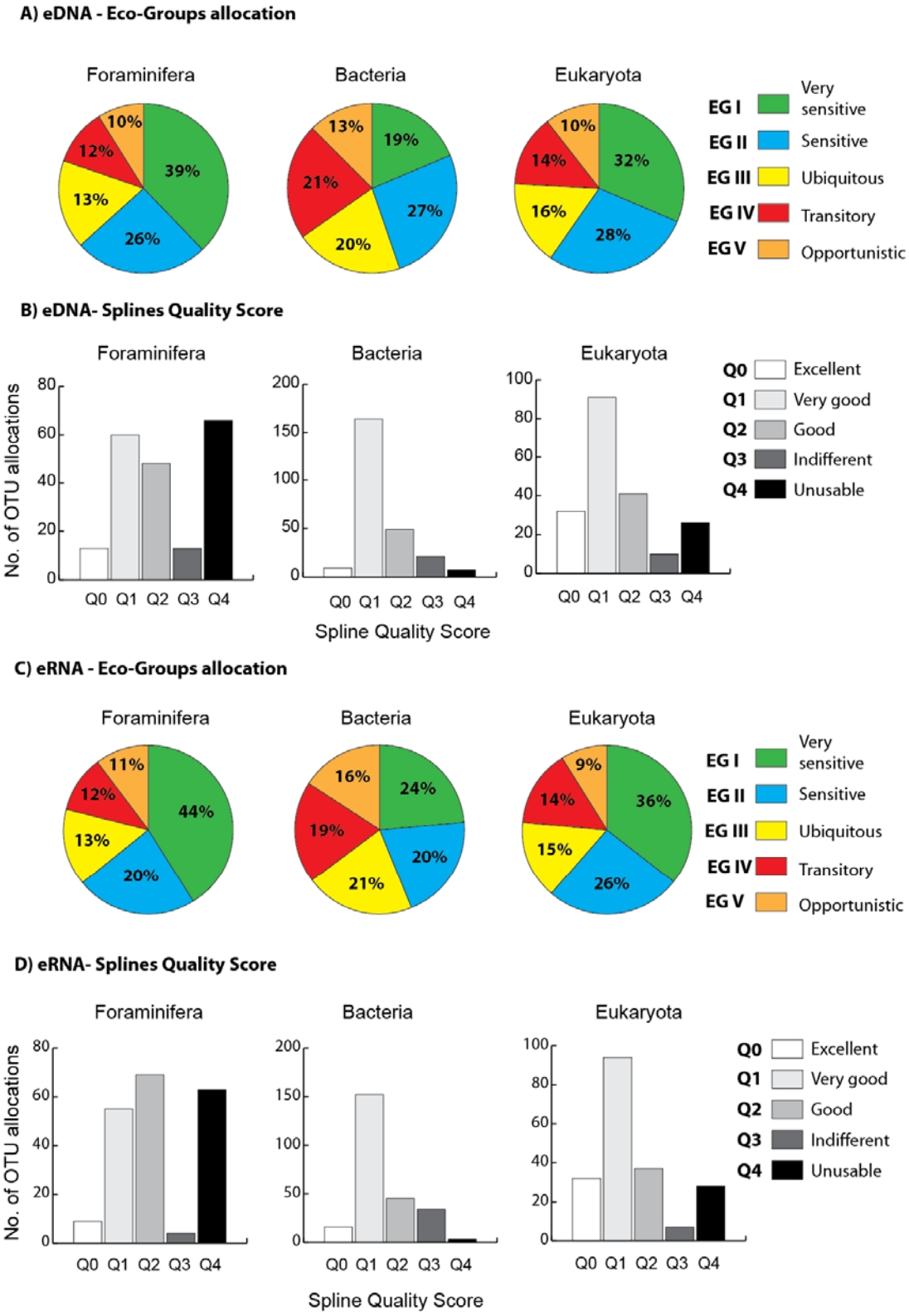


Figure 9. Summary of Eco-Group (EG) assignments (EG I-V) for each taxonomic group using the environmental DNA (A) and environmental RNA (C) datasets, and the corresponding regression splines quality scores (Q) using environmental DNA (B) and environmental RNA (D) data.

For the calculation of the MBI (next section), only bioindicator OTUs assigned to EGs with a quality score ranging from Q0 to Q2 (Excellent to Good) were used, resulting in a total number of incorporated bioindicators as follows: Foraminifera (DNA n = 121; RNA n = 106), Bacteria (DNA n = 230; RNA n = 211); and Eukaryotes (DNA n = 164 ; RNA n = 159); Figure 10. The removal of indifferent (Q3) and unusable (Q4) splines only affected foraminifera, with an apparent increase in the number of foraminiferal EG I and EG II proportions (Figure 10A and 10B).

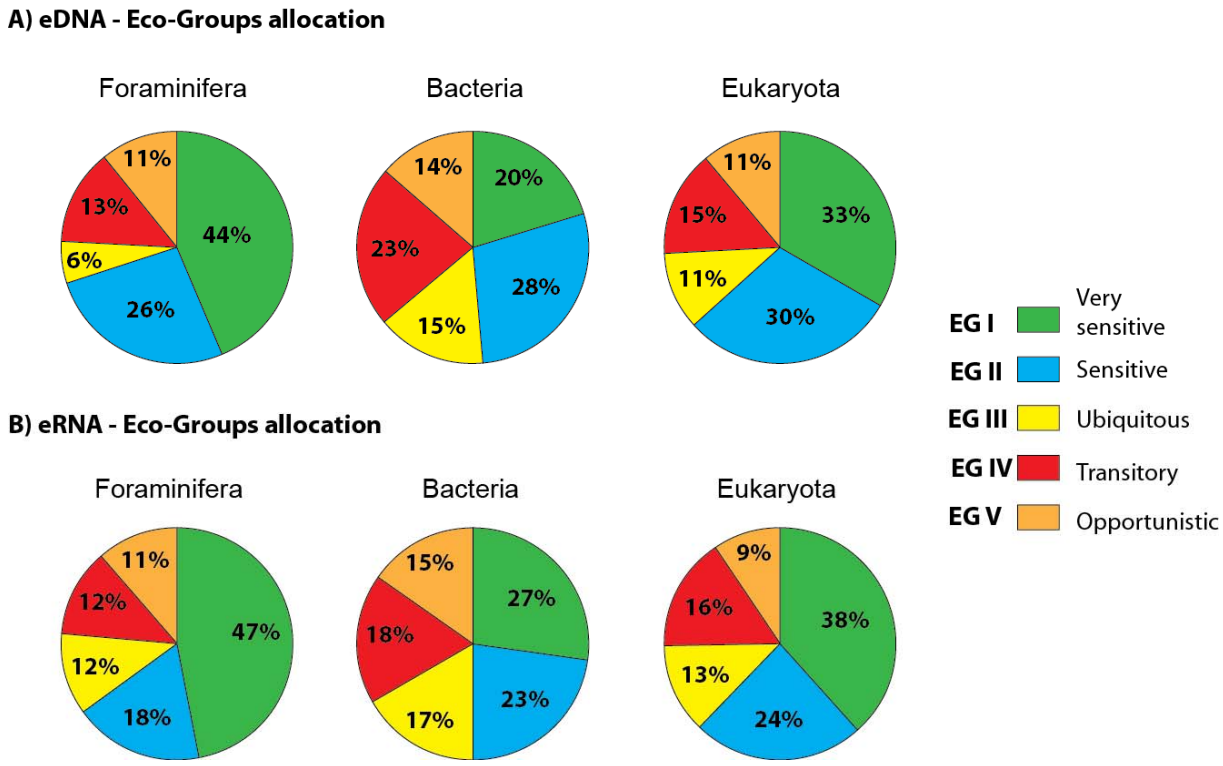


Figure 10. Summary of high-quality Score (Q0-Q2) Eco-Group assignments (EG I-V) for each taxonomic group using the environmental DNA (A) and environmental RNA (B) datasets.

3.4. Development and performance of eDNA and eRNA-based multi-trophic Metabarcoding Biotic Indices

The MBIs were first generated for each taxonomic groups separately, using standardised 4th root transformed eDNA and eRNA data. The MBI values were calculated utilising the EG assignment outlined in Section 3.3, first using the EG assignment data (MS Y1, MS Y2, and SI Y2; Appendix 5). These were then tested on a separate validation dataset (MS Y3; Appendix 6). The MBI values were generally very well correlated with ES values, with significant R^2 values of 0.73 (For), 0.84 (Euk) and 0.94 (Bac) (Appendix 5 and 6).

Following testing of several standardisation methods, it was established that the most severe transformation (presence-absence) resulted in the highest consistency between the resulting MBI values for the different taxonomic groups. This facilitated combining the different taxonomic datasets to be used in conjunction with combined EG datasets to produce a multi-trophic MBI. Individual MBI values were first derived for each taxonomic group and datasets (eDNA vs eRNA), and the linear regression plotted against the ES.

For the assignment dataset (MS Y1, MS Y2, and SI Y2; Figure 11A-J), the weakest relationship was with For-MBI (eDNA, R^2 = 0.777, Figure 11A; eRNA, R^2 = 0.745, Figure 11B), followed by Bac-MBI (eDNA, R^2 = 0.857, Figure 11E; eRNA, R^2 = 0.851, Figure 11F) and then Euk-MBI (eDNA, R^2 = 0.901, Figure 11B; eRNA, R^2 = 0.886, Figure 11C). Very strong relationships were also obtained when all three taxonomic datasets (mt-MBI) were combined (eDNA and eRNA R^2 = 0.900, Figure 11I-J), and when just Bac and Euk data (mt[e+b]-MBI) were combined (eDNA R2 = 0.901, and eRNA R2 = 0.900, Figure 11G, H). The slope of the fitted linear regression also varied slightly between taxonomic groups (eDNA: Euk 1.98x > For 1.87x > Bac 1.17x; eRNA: For 2.02x > Euk 1.91x > Bac 1.17x), indicating a slightly different scale of responses.

When tested on the independent validation dataset (MS Y3; Figure 11K-T), similar strong correspondence between MBI and ES was achieved. The For-MBI was noticeably less well correlated with ES when using eDNA (R^2 = 0.731, Figure 11K) than when using eRNA (R^2 = 0.850, Figure 11L), while markedly higher relationships were obtained with Euk-MBI (eDNA, R^2 = 0.930, Figure 11M; eRNA, R^2 = 0.932, Figure 11N) and with Bac-MBI (eDNA, R^2 = 0.924, Figure 12O); eRNA, R^2 = 0.944, Figure 11P). Correspondence with ES also remained very strong for the mt-MBI and the Euk+Bac-MBI, with R^2 values > 0.926 (Figure 11Q-T).

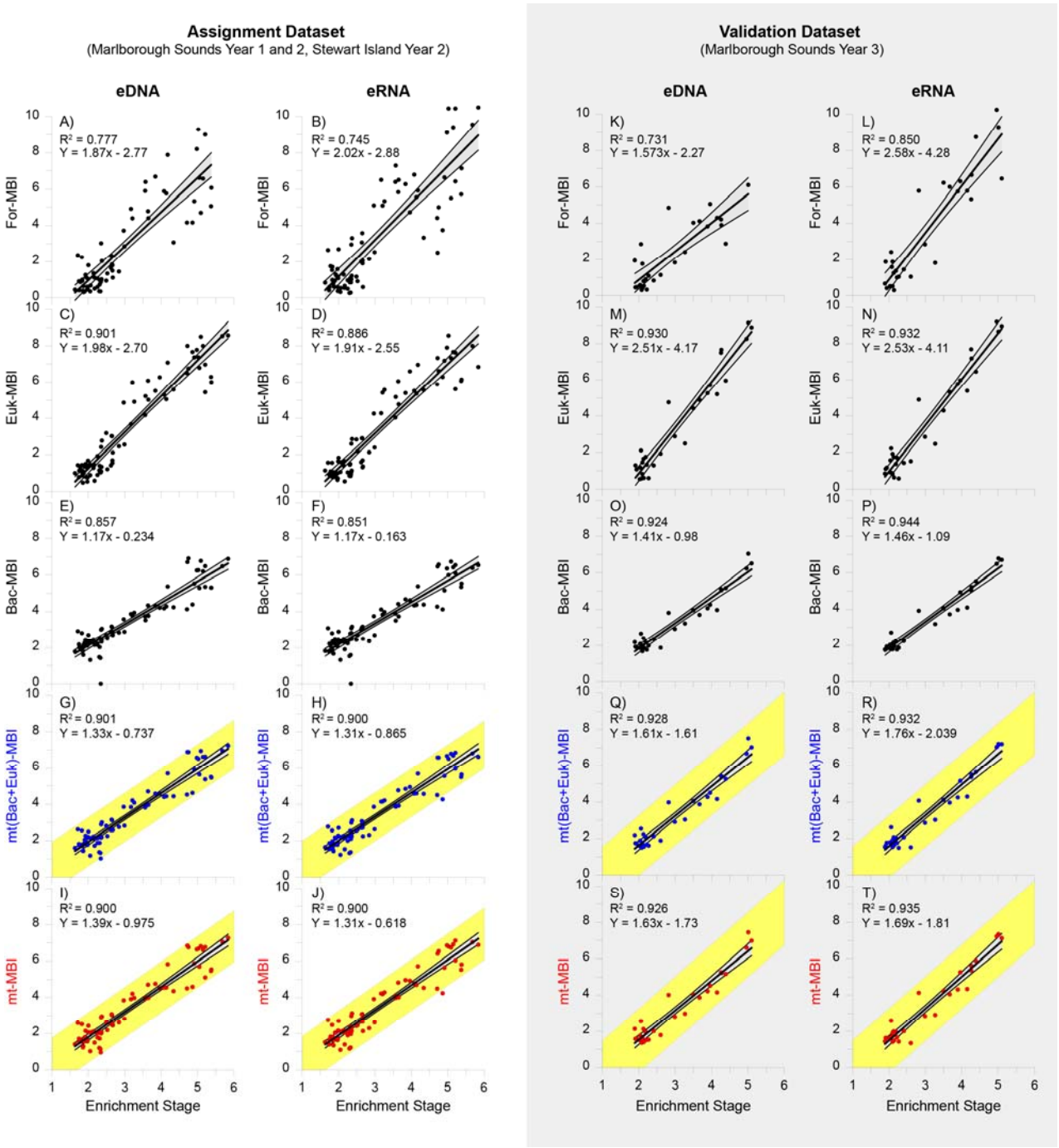


Figure 11. Linear relationships between environmental DNA and RNA-based biotic indices and overall Enrichment Stage calculated from presence-absence read-abundance data. Regression slopes on the left (white area) and right (shaded area) correspond to the Eco-Group assignment dataset and validation dataset, respectively. Confidence interval delineators surrounding each regression slope indicate 95% confidence intervals, yellow region indicates margin where estimates equate to ES ± 1 ES. For = foraminifera, Bac = bacteria, Euk = eukaryota, mt = multi trophic, MBI = Metabarcoding Biotic Index.

An analysis of covariance (ANCOVA) was performed on each regression matrix shown in Figure 11 to evaluate if the eDNA-derived indices would provide the same

robustness (i.e., identical linear regression slope) as the eRNA-derived indices. Table 8 shows that there was no significant difference in the linear regression slopes between any pairs of eDNA- and eRNA-derived indices, except for the f-MBI index in the validation dataset, which produced significantly different regression slopes between datasets. Correspondence between mt-MBI eDNA/eRNA matrices yielded an extremely strong relationship (R^2 = 0.981, data not shown), indicating that equivalent results will be obtained regardless of whether eDNA or eRNA are utilised.

Table 8. Pr(>F) values for slope and intercepts in ANCOVA models testing for difference between eDNA and eRNA regressions for each index for both main Eco-Groups assignment and validation datasets (Figure 11). Model1 (for slope) = aov(Y~ES*DNA_RNA, data), Model2 (for intercept when slope is not significant) = aov(Y~ES+DNA_RNA, data). Significant values in bold.

		Model 1 Pr(>F)		**Model 2** Pr(>F)
		Slope	Intercept	Intercept
Main dataset	***f*-MBI**	0.4340	0.8645	0.135
	***e*-MBI**	0.9354	0.7898	0.6083
	***b*-MBI**	0.5115	0.6793	0.6009
	***mt*-MBI**	0.9488	0.6678	0.3227
	***mt(e+b)*-MBI**	0.7360	0.6242	0.635
Validation dataset	***f*-MBI**	**0.0009**	**0.0264**	**0.0034**
	***e*-MBI**	0.5518	0.7361	0.504
	***b*-MBI**	0.9127	0.9144	0.5199
	***mt*-MBI**	0.2952	0.4509	0.4786
	***mt(e+b)*-MBI**	0.5355	0.6067	0.8322

4. DISCUSSION

In the present study we used eDNA/eRNA samples collected during three separate surveys on successive years from a range of high- and low-flow salmon farms in New Zealand to develop a mt-MBI, incorporating three different taxonomic groups (foraminifera, bacteria, and general eukaryotes).

The development and validation of the mt-MBI required the investigation of a range of important considerations, including (a) multiplexing metabarcoding data from the different taxonomic groups, (b) the exploration of spatio-temporal changes in community composition within each taxonomic group along enrichment gradients of fish farms, and (c) the rigorous selection of key molecular bioindicator taxa attributed to five distinct EG categories to be incorporated into the mt-MBI. Below, we discuss these considerations and the robustness of the mt-MBI. We provide future recommendations for the successful implementation of this new monitoring tool.

4.1. Multiplexing of metabarcoding data

High-throughput sequencing metabarcoding enables a large numbers of samples to be analysed quickly and automatically, and can provide semi-quantitative data on various marine communities. Currently, capturing different taxonomic groups using metabarcoding requires specific PCR amplification of each group independently followed by individual HTS of the particular group(s) under investigation. The assembly and sequencing of HTS libraries for three distinct taxonomic groups can be relatively time-consuming and costly. In the present study, we investigated whether sequencing costs could be reduced by using a multiplex approach consisting of pooling the PCR products from the three groups (foraminifera, bacteria, and eukaryotes) into single HTS libraries for simultaneous sequencing, followed by bioinformatics separation of sequence reads from each taxonomic group. Theoretically, a single Illumina MiSeqTM run can contain three distinct PCR amplicons x 96-wells, totalling 288 samples. Our results showed that while this approach reduced the sequencing cost to ca. $11 per sample, it markedly reduced (i.e., < 10,000) the number of high-quality sequences per sample and many samples required re-sequencing. Our recommendation, based on using the current Illumina MiSeq™ chemistry kits, is that no more than 192 PCR amplicons should be sequenced simultaneously on one plate.

Global biodiversity analysis of the metabarcoding data (refer Figures 2-4) revealed that the multiplexing approach was very successful at capturing a wide range of taxa. Foraminiferal and bacterial diversity data obtained here were consistent with our previous studies derived from some of the same samples collected in 2012. For example, our global foraminiferal diversity chart (refer Figure 2) confirmed the dominance of unilocular monothalamids, followed by multilocular rotaliids, textulariids

and miliolids in similar proportions to our previous findings (Pochon et al. 2015a, 2015b). Similarly to Dowle et al. (2015), our global bacterial data (refer Figure 3) were largely dominated by Proteobacteria, particularly Gammaproteobacteria with a high proportion of unclassified reads at the order level, and with four dominating bacterial genera (*Sulfurovum*, *Thioprofundum*, *Vibrio*, and *Desulfosarcina*). Finally, although no comparative metabarcoding data exist on the genetic diversity of eukaryotic assemblages occurring along enrichment gradient of fish farms in New Zealand, our data (refer Figure 4) yielded similar results than to a previous study by Lejzerowicz et al. (2015) who used metabarcoding to explore metazoan diversity at a Scottish salmon farm. Consistent with Lejzerowicz et al. (2015) the majority of sequences were nematodes, annelids, Platyhelminthes and arthropods. Ciliates were also well represented in the eukaryotic dataset (Figure 4), especially *Trachelorаphis huangi*.

Collectively, these results demonstrate that our multiplexing metabarcoding approach was effective in capturing a wide diversity range of assemblages across the three taxonomic groups, representing a unique opportunity to incorporate small-sized (< 1 mm) taxa (e.g., copepods, ostracods, and gastrotrichs) that are not currently surveyed using traditional methods (Grego et al. 2009). This approach enables the cost-effective analysis of a more diverse biome than previously possible, making it applicable for routine testing.

4.2. Community composition shifts along enrichment gradients of fish farms

When evaluating taxa or communities for the purposes of indicating enrichment levels, a general prerequisite is that they demonstrate clear compositional changes along enrichment gradients (Pearson and Rosenberg 1978). Numerous studies have used multivariate analysis to visualise the extent of such changes, using fish farms as the enrichment point source (e.g., Keeley et al. 2013, 2014; Kutti et al. 2007; Valdemarsen et al. 2015). In the present study, compositional changes were clearly observed for all three metabarcoding-derived taxonomic groups (i.e., For, Bac, and Euk). Consistent directional trends with proximity to the farms were evident, both between the different farms and regions (i.e. in space), and from the repeated surveys (i.e. in time) for all three datasets. Although similar changes have been described previously from metabarcoding data, for example foraminifera (Pochon et al. 2015a, 2015b) and bacteria (Dowle et al. 2015), the analogies between all four groups (MCD included) observed here demonstrates: (a) a level of repeatability and regional transferability, (b) the validity for their incorporation into a multi-trophic level index, and (c) the potential for the resulting index to strongly differentiate samples from different points along an enrichment gradient.

There were, however, some interesting differences among taxonomic groups in terms of community shifts between enriched or unenriched conditions. Many bacterial OTUs

were either associated with highly impacted near-farm sediments, or with un-enriched sediments. For example, several Spirochaeta OTUs (previously commonly referred to as *Beggiatoa;* Verhoeven et al. 2016), which often form white bacterial mats in anoxic conditions beneath salmon farms, were strongly associated with highly enriched sediments (Figure 5F). Additionally, bacterial communities did not appear to be notably influenced by farm, region or flow specificity (expressed on the y axes in Figure 5F). This demonstrates that bacterial bioindicator OTUs are highly consistent between years and are regionally transferable. In contrast, there were a number of eukaryotic taxa whose distribution and abundance appeared to be affected by flow (TEP versus OTA) or region (SI versus MS). For the purpose of indicating enrichment status of a site, these taxa (e.g., *Berthella* sp., Figure 5G) are not very useful, as their distributions and abundances will be confounded by factors other than enrichment.

Foraminiferal community shifts displayed the highest variability between investigated regions, water flows, and sites (Figure 5A). This indicates that foraminiferal assemblages are affected by site-specific environmental conditions. Furthermore, only one species (*Vellaria pellucidus*, Figure 5E) that had been previously identified as a good bioindicator of high enrichment sites (Pochon et al. 2015b) was recovered here, suggesting that temporal shifts in foraminiferal communities occur. These findings indicate that there are likely to be significant challenges when developing a nationally or internationally applicable biotic index that relies on inclusion of foraminifera.

4.3. Eco-Group assignment, development and validation of the multi-trophic Metabarcoding Biotic Index

The approach of plotting read abundance information against overall ES stage which is based on multiple conventional indicators has proven to be a useful method for elucidating ecological tolerances of lesser known (or even unknown) organisms. In this project, it has facilitated the successful development of a multi-trophic EG-based index utilising purely eDNA/eRNA data. Previous studies have described organism associations with certain enrichment states (e.g. bacteria [Dowle et al. 2015] or foraminifera [Pawlowski et al. 2014a, 2014b; Pochon et al. 2015a, 2015b]), but these are generally based on a limited sample pool or one-off assessments, resulting in uncertainty as to whether the observations are applicable at wider spatial and temporal scales. Other researchers utilising eDNA data have relied on established knowledge of ecological tolerances of known macrofaunal species (e.g., AMBI 5.0 software, http://ambi.azti.es), thus limiting the scope to well-known and macroscopic organisms (Aylagas et al. 2014). By sequentially plotting the distribution of individual molecular OTUs recorded in three independent surveys (two being at the same location on consecutive years and the third being from a completely separate region) and checking for consistency we have reliably allocated more than 500 new bioindicator taxa representing a wide range of known and unknown bacterial, foraminiferal and eukaryotic organisms to 5 distinct EG categories.

Proof of concept can be found in how the widely recognised first-order opportunist (EG V) *Capitella capitata*, as identified through morphological assessment, and *Capitella teleta*, identified through metabarcoding, responded so similarly to enrichment (Figure 5G, 5H, 8). It is likely that this is in fact the same species, as only one species of *Capitella* has been identified from these sites. Species level identification is known to be challenging for this polychaete genus (Westheide & Schmidt 2003), and/or there may not be sufficient resolution within the 18S rRNA gene used in this study to differentiate them based on sequence data. Regardless, both sequence read abundances and morphological count abundances increased strongly and analogously in response to increasing enrichment, with peak read abundance occurring at ES 5, and accordingly being assigned EG V. Another example is the OTUs assigned to Spirochaetales (discussed previously), which were allocated an EG IV or V using the spline assignment method. Similarly, the Desulfobacterales that were identified here as a reliable bioindicator of strongly enriched conditions, and was independently assigned an EG V using the splines approach, and is also a known sulphate-reducing bacteria that functions under anaerobic conditions (Garrity et al. 2005). Hence, these observations and the clear result in the validation data support the use of the EG process to classify lesser known organisms.

The allocation and selection of high-quality EG-based bioindicator taxa for constructing the individual (f-MBI, b-MBI and e-MBI) and multi-trophic (mt-MBI) indices were undertaken following in-depth analysis of both eDNA and eRNA data in parallel. Because RNA molecules degrade rapidly after cell death, they are considered to represent a better proxy for measuring recent biodiversity than DNA which may persist in the environment for extended periods of time (Corinaldesi et al. 2008; Dell'Anno & Danovaro 2005). We also collected triplicate samples from each site to minimise potential biological biases due to environmental micro-patchiness (Pawlowski et al. 2016b), and filtered our datasets to keep only OTUs present in at least one of the triplicate samples from each type of molecules. This step enabled us to focus the analysis and selection of EG on biologically active taxa.

A major limitation in using eRNA molecules for routine monitoring is the high-cost associated with RNA extraction and sample processing. Isolating and processing DNA samples is faster and more cost-effective for routine testing, and therefore investigating whether eDNA can provide the same answer as eRNA is of critical importance. Our parallel analysis of both eDNA and eRNA products enabled a robust allocation of EGs that best captured contemporary diversity changes along enrichment gradients. As revealed in our regression spline analysis (Appendix 3), there was an overall strong agreement in spline distribution between the eDNA and eRNA datasets (Figures 6-8), indicating strong consistency in the distribution of the selected key EGs between both molecule types.

The few cases where eDNA- and eRNA-derived OTUs produced dissimilar regression spline trends may indicate differential persistence of eDNA fragments in the sediment. The latter were effectively filtered out using our spline quality score procedure. The resulting list of high-quality bioindicator taxa assigned to EGs and corresponding sequence GenCodeIDs (Appendix 4) is extremely valuable as it can serve as a new sequence reference database to which future eDNA samples will be automatically compared. The extremely robust ($R^2 > 0.9$) relationships between mt-MBI and ES (Figures 11 and 12), and the absence of statistical differences between eDNA- and eRNA-derived mt-MBI regressions (Table 7), further indicate that eDNA samples can provide the same answer as eRNA data when targeting species bioindicator taxa. However, further testing is required to confirm this as in this study only sequences which were present in eDNA and eRNA from each site were used. We recommend that additional samples from the same investigated MS fish farms (e.g., using available samples from our 2015-2016 OTA/TEP surveys) be analysed using eDNA and the mt-MBI further tested for final validation.

4.4. Robustness of the multi-trophic Metabarcoding Biotic Index

Macrofaunal taxonomy has long been regarded as the benchmark environmental indicator for biomonitoring (Diaz et al. 2004; Gray and Pearson 1982; Rice et al. 2012). Well-known biotic indices that have been developed based on macrofaunal data include the AMBI (Borja et al. 2000) and the Infaunal Trophic Index (ITI, Maurer et al. 1999). Other indices combining macrofaunal count data and physico-chemical or biological indicators have been specifically designed and effectively applied for monitoring fish farms (e.g., ES index, MPI 2015; Norwegian Quality Index, Rygg & Norling 2013).

Confidently replacing these established indices with an alternate method necessitates that a very strong relationship be demonstrated between the two approaches, as the consequences of misdiagnosing environmental conditions in a biomonitoring context can be significant. For example, in the case of evaluating enrichment effects from salmon farms, over-estimating impact level status can result in premature ‘fallowing’ (stock removal) of the site for prolonged periods, which has economic implications in the millions of dollars. Conversely, failure to identify severely degraded conditions can have significant social and ecological ramifications, and greatly prolong benthic recovery (Brooks et al. 2004; Keeley et al. 2015).

To date, several studies have demonstrated the potential for metabarcoding data from certain taxa groups to correlate with some traditional benthic indicator variables (Pawlowski et al. 2014a, 2014b; Dowle et al. 2015; Pawlowski et al. 2016a, 2016b; Pochon et al. 2015a, 2015b). However, many of the key relationships exhibit more scatter, and therefore greater uncertainty in the result than would be desirable in order to accept a metabarcoding-based tool. For example, R-squared values at or below ca.

0.8 implies that at least 20% of the variability is unaccounted for and as such, the room for misinterpreting actual benthic status is 1 in 5. The strong and significant correlations obtained in the present study between overall ES and mt-MBI confirm that metabarcoding has the potential to be used as a proxy for ES in the near future.

These stronger relationships are largely due to the EG approach (Borja et al. 2000; Borja and Muxika 2005), which eliminates many of the challenges associated with using simple metrics, diversity-based approaches and multivariate analyses. Previous attempts to use diversity-based measures such as Shannon diversity (H') and Chao indices (Gotelli and Colwell 2011; Pawlowski et al. 2016b) are vulnerable to sequence abundance biases associated with PCR and sequencing errors. Data normalisation procedures are usually required in order to obtain consistency and comparability between samples (see Pochon et al. 2015a), all of which strongly influence the resulting diversity score. Multivariate analyses enable overall differences between samples based on species-abundance data to be identified, whereby each analysis very effectively arranges the samples in 2- or 3-dimensional space. However, these arrangements are essentially 'unitless', and the frame of reference changes for each analysis. This makes it very challenging to relate the differences to the sort of fixed categorical scale that is necessary for the purposes of setting biomonitoring thresholds and standards. Both approaches also effectively ignore important ecological sensitivity information (e.g. EG, Borja et al. 2000; HS50, Rosenberg et al. 2004), which can infer valuable information about environmental conditions.

In the present study, we avoided this problem by (a) selecting OTUs that could be repeatedly and reliably identified, (b) assigning ecological sensitivities (e.g., eco-barcoding, Pawlowski et al. 2016b), and (c) treating the data on a presence / absence level thus avoiding the need for sample normalisation, and overcoming any uncertainties related to differences in read abundances between samples due to sequencing efforts. The 'eco-barcoding' reference library component of this approach is also concurrent with recommendations from other related studies (Aylagas et al. 2014; Pawlowski et al. 2016b). The results presented here are based on the assignment of ca. 500 molecular bioindicator taxa assigned to EG categories; we anticipate this database will continue to expand as more reliable GenCodeIDs are identified. A larger database should equate to more useful bioindicator OTUs being identified and barcoded, thereby increasing the accuracy, reliability and robustness of this approach.

Another important consideration with respect to replacing a long established environmental indicator such as macrofauna taxonomy, is the time scale over which different indicators integrate. This aspect provides a strong argument for incorporating multiple trophic level organisms into a single indicator. Macrofauna used in routine monitoring globally encompass a range of organisms with body sizes usually in the order of 1–20 mm (with the occasional larger-bodied organism), and with that a wide range in life expectancies, and therefore, durations over which they indicate their

response to suitable/unsuitable conditions (by their presence). Opportunistic polychaetes can have reproductive turnover on the time scale of weeks (Gremare et al. 1989), whereas, at the other extreme, a large-bodied bivalve like *Arctica islandica*, can live for 100s of years (Schöne et al. 2005).

In this study we have considered three broad taxonomic groups: foraminifera, bacteria and the diverse eukaryotic lineage including multicellular organisms, and found the latter two groups to have very strong correlations with overall ES, both individually and in combination (i.e. a true multi-trophic index). The turnover time of bacteria can be from minutes to days (Luna et al. 2002) and while this bioindicator group alone seems to correlate very well with traditional indicators, it is important to recognise the relatively contemporary implications. Likewise, some of the eukaryotes identified through metabarcoding are also smaller than the conventional macrofauna minimum size (1 mm sieve), organisms such as ciliates and the small enrichment-sensitive invertebrate *Echinoderes setiger*, which presumably have relatively rapid turnover rates. These small, fast-responding organisms are used here in conjunction with a variety of OTUs from larger-bodied organisms, such as urchins, polychaetes and small crustaceans, to give a more holistic, time-integrated assessment. Interestingly, some bacteria are thought to share a synergistic relationship with capitellid worms (e.g., *Vibrio cyclitrophicus*; Wada et al. 2008), supporting their inclusion as a useful bioindicator organism. The incorporation of small organisms, that are not easily identified through regular taxonomy, is advantageous in that they tend to be highly numerous and spread throughout the sediment, and are therefore, well sampled from small sediment volumes. Employing such a multi-trophic approach ensures that the biomonitoring results capture both very recent short-term changes in sediment condition (geochemistry) as well as longer-term chronic pressures.

While the incorporation of multiple taxonomic groups arguably improves the ability of the mt-MBI to capture subtle environmental changes along the enrichment gradient, our results showed that foraminifera underperformed compared to bacterial and eukaryotic data (Figures 11 and 12). The exclusion of foraminiferal data did not affect the overall performance of the mt-MBI as evidenced by the same strong relationships (R^2>0.9) obtained with the mt(b+e)-MBI. Consequently, we suggest that foraminifera be excluded from the mt-MBI. Focusing the analysis on bacterial and eukaryotic assemblages will reduce analytical time and costs, while providing extremely robust assessment of the impacts of fish farm on benthic communities.

5. CONCLUSIONS AND FUTURE RECOMMENDATIONS

We have developed and tested the robustness of an mt-MBI for monitoring the impacts of salmon farming on benthic environments in New Zealand. The new index was developed using samples collected over a period of three years from a range of high- and low-flow salmon farms from two distinct regions in New Zealand, and was validated against the established ES index currently used for fish farm monitoring in the Marlborough Sounds. Based on the results, the tool is now ready to use and complements current monitoring using the ES methodology. The mt-MBI should be used in parallel to current methods to further ground-truth the tool and in turn facilitate its incorporation into best management practices and consent monitoring programmes.

The key findings and future recommendations from this study are:

- A multiplexed HTS approach was used to cost-effectively characterise three taxonomic groups occurring along enrichment gradients of fish farms. We recommend that no more than two different genes are multiplexed during the HTS step to ensure sufficient sequencing data is obtained and the diversity rarefaction curves reach a plateau. The metabarcoding approach used in this study captured a wide diversity of marine organisms including many micro-organisms that cannot be assessed using current morphologically-based approaches. The surveyed taxa have a large range of faunal biomasses, longevities and response times. Employing a multi-trophic approach ensures that the biomonitoring results capture both very recent short-term changes in sediment condition (geochemistry) as well as longer-term chronic pressures.
- Similar spatio-temporal community changes in response to enrichment pressure were observed among the three taxonomic groups studied. Bacterial communities were least affected by factors such as flow and year. Bacterial communities varied significantly along the enrichment gradients regardless of space and time, revealing their potential as stable bioindicators nationwide. In contrast, the distribution of foraminiferal assemblages was sporadic and appeared to differ spatially and temporally, limiting their potential to be used as robust bioindicator taxa.
- The EG allocations approach used in the present study enabled us to robustly identify 500 key bioindicator taxa from the three taxonomic groups. These were then used to design individual and combined (multi-trophic) indices. The multi-trophic (mt-MBI) indices performed better than individual indices, yielding extremely robust ($R^2 > 0.9$; i.e. > 90%) relationships with the traditional ES index. The exclusion of foraminifera did not reduce the strength of the multi-trophic approach, indicating that the combined use of bacteria and eukaryotes is sufficient for effective monitoring of fish farms with this method. We recommend that foraminifera be excluded from the mt-MBI. Focusing on bacteria and eukaryotes is

sufficiently robust and will reduce the costs of sample analysis for use in the mt-MBI.

- Each high-quality bioindicator taxon was assigned a unique GenCodeID. This will form the genesis of a reference database against which future environmental samples will be directly screened for rapid and automatic EG identification and mt-MBI calculation. This approach will circumvent the need for recurrent, time-consuming statistical analysis of each new dataset.
- Similar relationships with ES were observed when eDNA or eRNA molecules were used. Because DNA can be analysed much more rapidly and cost-effectively than RNA, we suggest that for monitoring purposes only eDNA be used in the future. A caveat regarding this suggestion is that in this study any sequences not occurring in both eDNA/eRNA were removed (i.e., only biologically active taxa were analysed). We suspect that using DNA data only (i.e. without the data screening step) will produce similar results; however we recommend further testing with new datasets to confirm this.
- Implementation for compliance purposes needs to be carried out with caution, due to the potential commercial and environmental consequences of the assessments. We therefore suggest a 'phase-in strategy' whereby eDNA samples are collected in conjunction with normal routine sampling for two additional years of sampling and analysis of the past two years using archived samples.
- We estimate that the 'phase-in period' would involve the analysis of 30 samples per year (2015–2018) from Otanerau and Te Pangu farms, as well as from two other additional farms, and the multiplexed sequencing of bacterial and eukaryotic assemblages. For the first two years both eDNA and eRNA will be extracted to ensure that the same results can be obtained using eDNA without the data screening step.
- At the end of this period, the ability of mt-MBI to estimate enrichment stage and with it, reliably determine farm compliance, will be tested. Based on this assessment recommendations will be made as to how the mt-MBI is best utilised for monitoring purposes. The most likely option it to use mt-MBI as a substitute for macrofauna analyses (and associated metrics) in the calculation of ES. Or, if the relationship between ES and mt-MBI continues to be very good, it may be possible to estimate ES based solely on eDNA data, thereby also negating the need for sulphide and redox analyses. This phase-in period would also permit the identification of additional useful OTUs for EG assignment, thereby expanding the versatility of the GenCodeID database.
- At the same stage it would be appropriate to consider how it is used with respect to Type 1, 2 or 3 monitoring, as described in the Benthic BMP (MPI 2015). The recommended changes can then be proposed to the benthic standards working group for consideration and formal implementation.
- A further step that would reduce costs associated with the mt-MBI would be the use of bacteria only. We suspect that if the Bac-MBI was used, sample sizes could

be reduced from 2 g to ca. 0.25 g. This allows cheaper and much faster eDNA extraction techniques to be used. Assessment of whether similar results could be obtained using this method could be undertaken over the first two years of the 'phase-in testing period' recommended above.

- Another advantage of the MBI developed in this project is that the sample size required (ca. < 5 g) is significantly less than for macrofaunal assessments. We have designed a prototype sampler (Figure 12) that would at least half current sampling times at each site. We recommend refining this design, building a prototype and undertaking field trials. This could further reduce compliance monitoring costs.

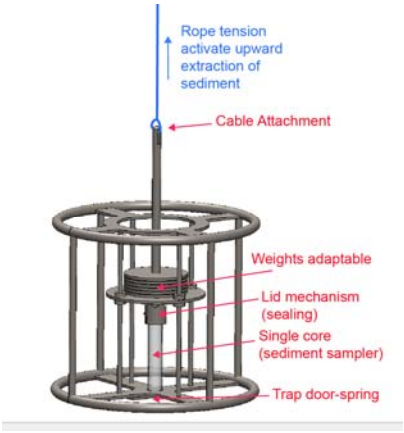


Figure 12. Single corer prototype specifically adapted for molecular applications, and able to take virtually undisturbed samples with short lowering and lifting times (designed by Ace Engineering, Brightwater, New Zealand)

- The mt-MBI has been developed and validated using samples from the Marlborough Sounds and Stewart Island. We recommend that it is tested using samples from other regions and in conjunction with other marine activities that result in organic enrichment (e.g. mussel farms, albeit less severe gradients; see Firth of Thames report, Pochon et al. 2016).

6. ACKNOWLEDGEMENTS

We are grateful to Prof. Jan Pawlowski and Franck Lejzerowicz (University of Geneva, Switzerland) for valuable discussions and bioinformatics advice. We thank New Zealand Genomics Limited for sequencing the data, and in particular Luke Boyle and Alex Stuckey for bioinformatics analyses and support. Special thanks to Jono Drew, Laura Biessy and Eddy Dowle (Cawthron) for their help during field work, sample processing and sequence libraries preparation. We are particularly grateful to Mark Gillard at New Zealand King Salmon Co. Limited, and Sanford Limited for access to their farms, samples and in-kind support.

7. REFERENCES

Aquaculture New Zealand magazine (Oct 2013). (PDF).

Anderson MJ 2008. Animal-sediment relationships re-visited: characterising species' distributions along an environmental gradient using canonical analysis and quantile regression splines. Journal of Experimental Marine Biology and Ecology 366: 1627.

Aylagas E, Borja A, Rodriguez-Ezpeleta N 2014. Environmental status assessment using DNA metabarcoding: towards a genetics based Marine Biotic Index (gAMBI). PLOS One 9, e90529.

Baird DJ, Hajibabaei M 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Molecular Ecology 21: 2039-2044.

Bissett A, Bowman J, Burke C 2006. Bacterial diversity in organically enriched fish farm sediments. FEMS Microbiology and Ecology 55: 48-56.

Borja A, Franco J, Perez V 2000. A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Marine Pollution Bulletin 40: 1100-1114.

Borja A, Muxika H 2005. Guidelines for the use of AMBI (AZTI's Marine Biotic Index) in the assessment of the benthic ecological quality. Marine Pollution Bulletin 50: 787-789.

Brooks KM, Stierns AR, Backman C 2004. Seven year remediation study at the Carrie Bay Atlantic salmon (*Salmo salar*) farm in the Broughton Archipelago, British Columbia, Canada. Aquaculture 239: 81-123.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7: 335-336.

Caporaso JG, Lauber CL, Walters WA, et al. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Science USA 108: 4516-22.

Clarke KR, Gorley RN 2015. PRIMER v7: User Manual/Tutorial. PRIMER-E, Plymouth, 296 pp.

Cole JR, Wang Q, Fish JA, et al. 2014. Ribosomal database project: data and tools for high throughput rRNA analysis. Nucleic Acids Research 42: D633-42.

Corinaldesi C, Beolchini F, Dell'Anno A 2008 Damage and degradation rates of extracellular DNA in marine sediments: implications for the preservation of gene sequences. Molecular Ecology 17: 3939–3951.

Dell'Anno A, Danovaro R 2005. Extracellular DNA plays a key role in deep-sea ecosystem functioning. Science 309: 2179.

Diaz RJ, Solan M, Valente RM 2004. A review of approaches for classifying benthic habitats and evaluating habitat quality. Journal of Environmental Management 73: 165-181.

Dowle E, Pochon X, Keeley N, Wood SA 2015. Assessing the effects of salmon farming seabed enrichment using bacterial community diversity and high-throughput sequencing. FEMS Microbiology and Ecology 91: fiv089.

Edgar RC 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460-2461.

Edgar RC, Flyvbjerg H 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31: 3476-3482.

Edgar RC, Haas BJ, Clemente JC, et al. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27: 2194-200.

Garrity GM, Brenner DJ, Krieg NR, Staley JT 2005. Bergey's Manual of Systematic Bacteriology, Volume Two: The Proteobacteria, Part C: The Alpha-, Beta-, Delta-, and Epsilonproteobacteria. Springer, New York.

Gray JS, Pearson TH 1982. Objective selection of species indicative of pollution-inducing change in benthic communities: I Comparative methodology. Marine Ecology Progress Series 9: 111-119.

Gotelli N, Colwell R 2011. Estimating species richness. In: Magurran AE, McGill BJ (eds) Frontiers in measuring biodiversity. Oxford University Press, New York, NY, p. 39−54.

Grego M, De Troch M, Forte J, Malej A 2009. Main meiofauna taxa as an indicator for assessing the spatial and seasonal impact of fish farming. Marine Pollution Bulletin 58: 1178-1186.

Gremare A, Marsh AG, Tenore KR 1989. Secondary production and reproduction of *Capitella capitata* type I (Annelida: Polychaeta) during a population cycle. Marine Ecology Progress Series 51: 99-105.

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, De Vargas C, Decelle J, Del Campo J 2012. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Research 27: gks1160.

Ji Y, Ashton L, Pedley SM, et al. 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecology Letters 16: 1245-1257.

Jones FC 2008. Taxonomic sufficiency: the influence of taxonomic resolution on freshwater bioassessments using benthic macroinvertebrates. Environmental Reviews 16: 45-69.

Keeley N, Forrest B, Crawford C, Macleod C 2012a. Exploiting salmon farm benthic enrichment gradients to evaluate the regional performance of biotic indices and environmental indicators. Ecological Indicators 23:453-466.

Keeley N, MacLeod C, Forrest B 2012b. Combining best professional judgement and quantile regression splines to improve characterisation of macrofaunal responses to enrichment. Ecological Indicators 12:154-166.

Keeley N, Forrest BM, Macleod C 2013. Novel observations of benthic enrichment in contrasting flow regimes with implications for marine farm monitoring and management. Marine Pollution Bulletin 66:105-116.

Keeley N, Forrest BM, Crawford C, Hopkins GA, MacLeod C 2014. Spatial and temporal dynamics in macrobenthos during recovery from salmon farm induced organic enrichment: when is recovery complete. Marine Pollution Bulletin 80:250-262.

Keeley N, Forrest B, MacLeod C 2015. Benthic recovery and re-impact responses from salmon farm enrichment: implications for farm management. Aquaculture 435:412-423.

Kobayashi M, Msangi S, Batka M, Vannuccini S, Dey MM, Anderson JL 2015. Fish to 2030: The role and opportunity for aquaculture. Aquaculture Economics & Management 19(3):282-300.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. Applied Environmental Microbiology 79: 5112-5120.

Koenker R 2005. Quantile Regression. Cambridge University Press, p. 342 http://dx.doi.org/10.1017/cbo9780511754098.

Koenker R, Ng P, Portnoy S 1994. Quantile smoothing splines. Biometrika 81: 673-680. http://dx.doi.org/10.1093/biomet/81.4.673.

Kutti T, Hansen PK, Ervik A, Høisæter T, Johannessen P 2007. Effects of organic effluents from a salmon farm on a fjord system. II. Temporal and spatial patterns in infauna community composition. Aquaculture 262: 355-366.

Langlet D, Geslin E, Baal C, Metzger E, Lejzerowicz F, Riedel B, Zuschin M, Pawlowski J et al., 2013. Foraminiferal survival after long-term *in situ* experimentally induced anoxia. Biogeosciences 10: 7463-7480.

Laroche O, Wood SA, Tremblay LA, Ellis JI, Lejzerowicz F, Pawlowski J, Lear G, Atalah J, Pochon X 2016. First evaluation of foraminiferal metabarcoding for monitoring environmental impact from an offshore oil drilling site. Marine Environmental Research 120: 225–235.

Lejzerowicz F, Esling P, Pillet L, Wilding TA, Black KD, Pawlowski J 2015. High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. Scientific Reports 5:13932.

Luna GM, Manini E, Danovaro R 2002. Large fraction of dead and inactive bacteria in coastal marine sediments: comparison of protocols for determination and ecological significance. Applied and Environmental Microbiology 68: 3509-3513.

Macleod C, Forbes S 2004. Guide to the assessment of sediment condition at marine fish farms in Tasmania. Aquafin CRC Project 4.1. Tasmanian Fisheries and Aquaculture Institute, p. 63.

Maurer D, Nguyen H, Robertson G, Gerlinger T 1999. The Infaunal Trophic Index (ITI): its suitability for marine environmental monitoring. Ecological Applications 9: 699-713.

McCaig AE, Phillips CJ, Stephen JR, et al. 1999. Nitrogen cycling and community structure of proteobacterial β-subgroup ammonia-oxidizing bacteria within polluted marine fish farm sediments. Applied Environmental Microbiology 65: 213-20.

MPI 2015. Best Management Practice guidelines for salmon farms in the Marlborough Sounds. Part 1: Benthic environmental quality standards and monitoring protocol (Version 1.0 January 2015). Prepared by the Benthic Standards Working Group: Keeley N., Gillard M., Broekhuizen N., Ford R., Schuckard R., Urlich S., Ministry for Primary Industries.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. 2014. Vegan: Community Ecology Package Version 2.2-0. http://cran.r-project.org/web/packages/vegan/vegan.pdf (Accessed 1 November 2016).

Pawlowski J, Lejzerowicz F, Esling P 2014a. Next-generation environmental diversity surveys of Foraminifera: preparing the future. Biological Bulletin 227:93-106.

Pawlowski J, Esling P, Lejzerowicz F, Cedhagen T, Wilding TA 2014b. Environmental monitoring through protist NGS metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. Molecular Ecological Resources 14: 1129-1140.

Pawlowski J, Esling P, Lejzerowicz F, Cordier T, Visco JA, Martins CIM., Kvalvik A., Staven K, Cedhagen T 2016a. Benthic monitoring of salmon farms in Norway using foraminiferal metabarcoding. Aquaculture Environment Interactions 8:371-386.

Pawlowski J, Lejzerowicz F, Apotheloz-Perret-Gentil L, Visco J, Esling P 2016b. Protist metabarcoding and environmental biomonitoring: Time for change. European Journal of Protistology, 55:12-25.

Pearson, T.H., Rosenberg, R., 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. Oceanography and Marine Biology Annual Review 16:229-311.

Pinedo S, Jordana E, Salas F, Subida MD, Adiego EG, Torres J 2012. Testing MEDOCC and BOPA indices in shallow soft-bottom communities in the Spanish Mediterranean coastal waters. Ecological Indicators 19: 98-105.

Pochon X, Wood SA, Keeley NB, Lejzerowicz F, Esling P, Drew J, Pawlowski J 2015a. Accurate assessment of the impact of salmon farming on benthic sediment enrichment using foraminiferal metabarcoding. Marine Pollution Bulletin 100: 370-382.

Pochon X, Atalah J, Keeley L, Drew J, Wood S 2015b. Preliminary assessment of a new molecular tool for biomonitoring New Zealand's fish farms. Prepared for Ministry for Primary Industries. Cawthron Report No. 2734. 15 p.

Pochon X, Zaiko A, Hopkins GA, Banks JC and Wood SA 2015. Early detection of eukaryotic communities from marine biofilm using high-throughput sequencing: an assessment of different sampling devices. Biofouling. Volume 31, Issue 3. Pages 241-251.

Pochon X, Cornelisen C, Morrisey D, Wood S 2016. Survey of microbial seabed communities in the Firth of Thames and Hauraki Gulf using a DNA metabarcoding approach. Prepared for Waikato Regional Council. Cawthron Report No. 2880. 21 p.

R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2014. http://www.R-project.org (10 November 2016, date last accessed).

Rice J, Arvanitidis C, Borja A, Frid C, Hiddink JG, Krause J, Lorance P, Ragnarsson SÁ, Sköld M, Trabucco B, Enserink L, Norkko A 2012. Indicators for sea-floor integrity under the European Marine Strategy Framework Directive. Ecological Indicators 12:174-184.

Rygg B, Norling K 2013. Norwegian Sensitivity Index (NSI) for marine macroinvertebrates, and an update of Indicator Species Index (ISI). In: NIVA-rapport;6475. Norsk institutt for vannforskning.

Rognes T, et al. (2015) VSEARCH. (https://github.com/torognes/vsearch)

Rosenberg R, Blomqvist M, Nilsson C, Cederwall H, Dimming A 2004. Marine quality assessment by use of benthic species-abundance distributions: a proposed new protocol within the European Union Water Framework Directive. Marine Pollution Bulletin 49: 728-739.

Schöne BR, Fiebig J, Pfeiffer M, Gleβ R, Hickson J, Johnson ALA, Dreyer W, Oschmann W 2005. Climate records from a bivalved Methuselah (*Arctica islandica*, Mollusca; Iceland). Palaeogeography, Palaeoclimatology, Palaeoecology 228: 130-148.

Simboura N 2003. Bentix Index vs Biotic Index in monitoring: an answer to Borja et al. 2003. Marine Pollution Bulletin 48: 403-404.

Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH 2012. Environmental DNA. Molecular Ecology 21: 1789-1793.

Valdemarsen T, Hansen PK, Ervik A, Bannister RJ 2015. Impact of deep-water fish farms on benthic macrofauna communities under different hydrodynamic conditions. Marine Pollution Bulletin 101: 776-783.

Valentini A, Pompanon F, Taberlet P 2009. DNA barcoding for ecologists. Trends in Ecology and Evolution 24: 110-117.

Verhoeven JTP, Salvo F, Hamoutene D, Dufour SC 2016. Bacterial community composition of flocculent matter under a salmonid aquaculture site in Newfoundland, Canada. Aquaculture Environment Interactions 8: 637-646.

Wada M, Zhang D, Do H-K, Nishimura M, Tsutsumi H, Kogure K, 2008. Co-inoculation of *Capitella* sp. I with its synergistic bacteria enhances degradation of organic matter in organically enriched sediment below fish farms. Marine Pollution Bulletin 57: 86-93.

Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied Environmental Microbiology 73: 5261-5267.

Westheide W, Schmidt H 2003. Cosmopolitan versus cryptic meiofaunal polychaete species: an approach to a molecular taxonomy. Helgoland Marine Research 57: 1-6.

Zhan A, Hulk M, Sylvester F, Huang X, Adebayo AA, Abbott C, Adamowicz SJ, Heath DD, Cristescu ME, MacIsaac HJ 2013. High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. Methods in Ecology and Evolution 4: 558-565.

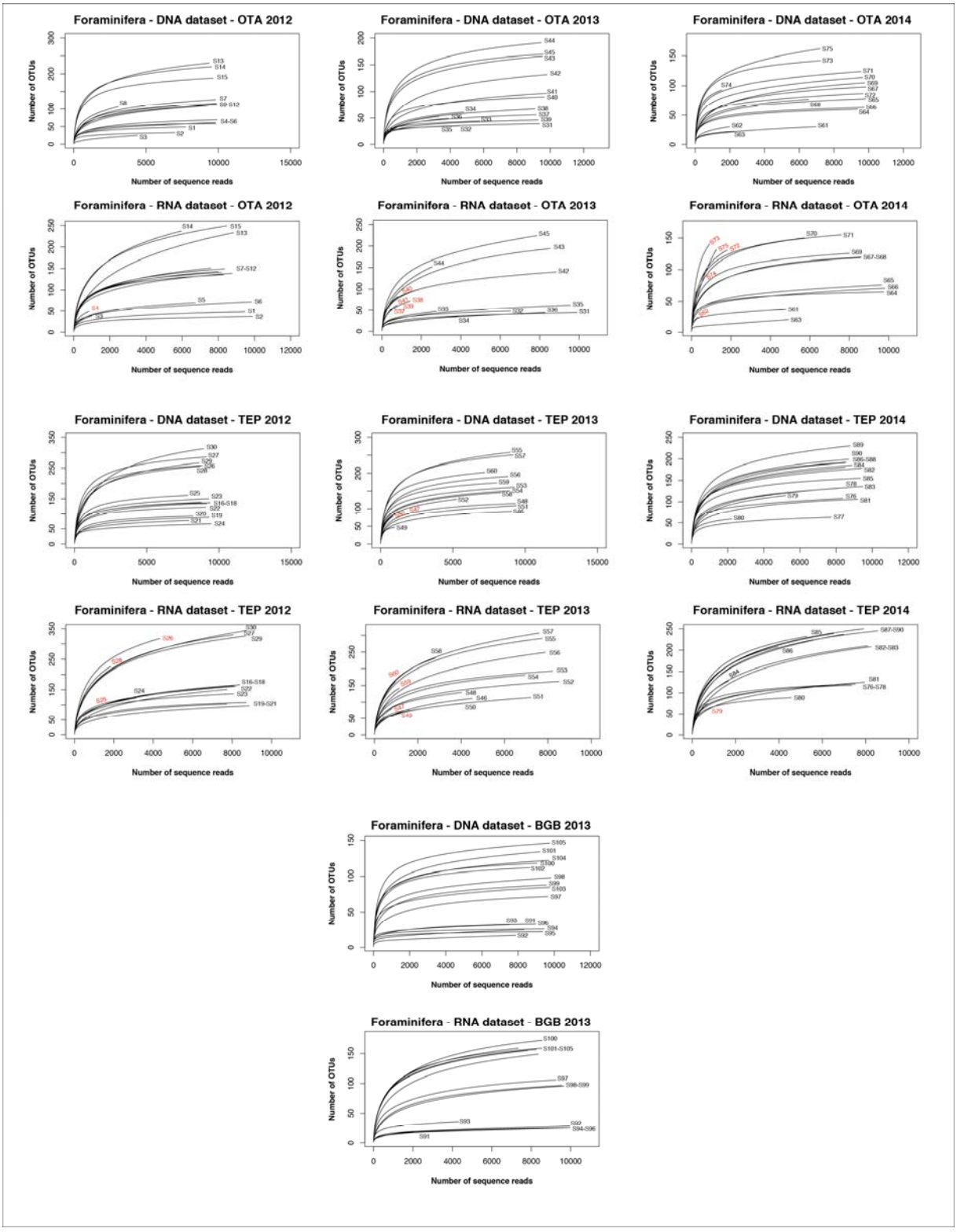
Zaiko A, Schimanski K, Pochon X, Hopkins G, Goldstein S, Floerl O, Wood SA 2016. Metabarcoding reveals early shifts in biofouling community composition: implications for monitoring non-indigenous marine species. Biofouling 32: 671-684.

8. APPENDICES

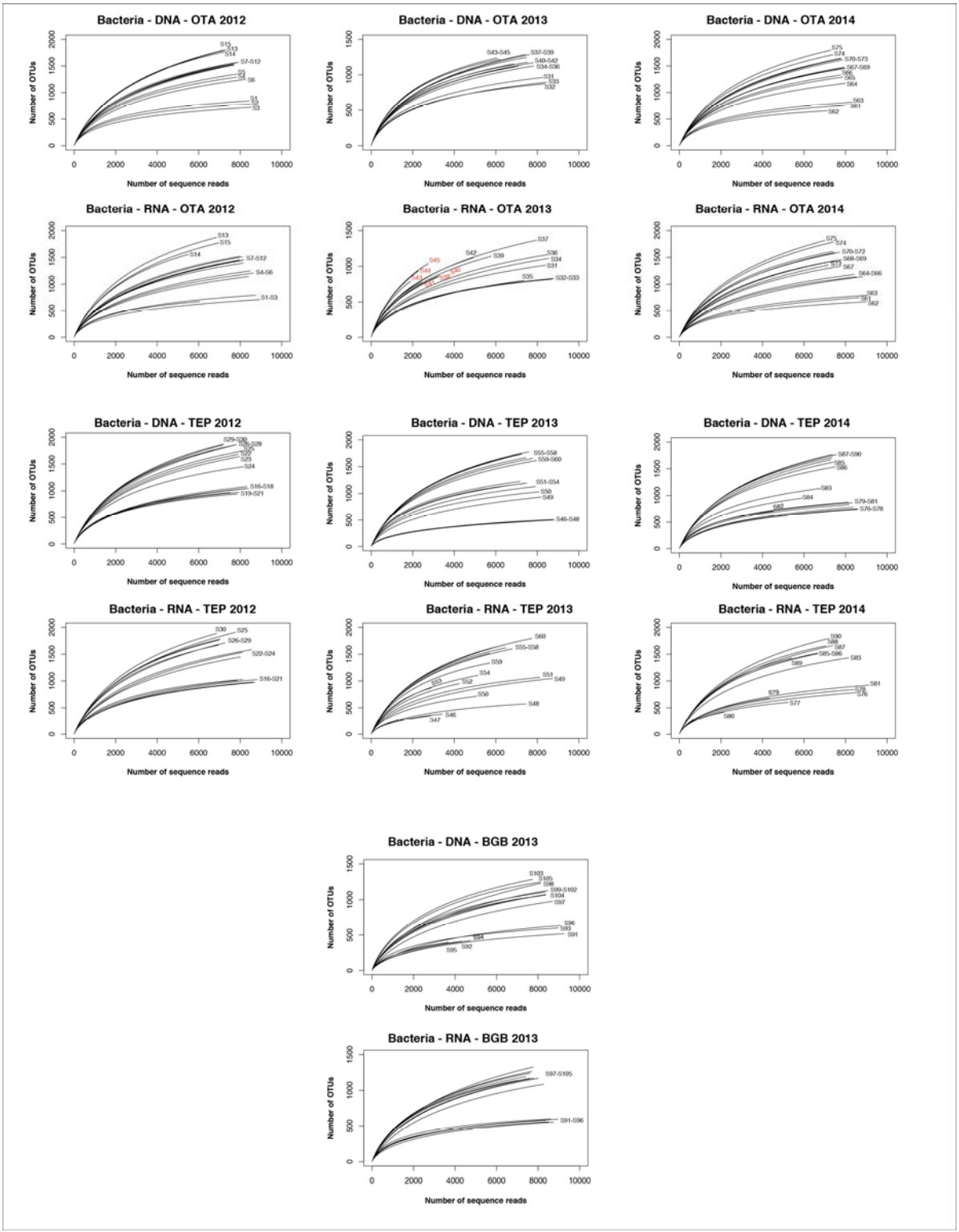
Appendix 1. List of PCR primers and thermo-cycling conditions used to amplify gene fragments for total bacteria, total eukaryotes, and total foraminifera. bp = base pair.

Primer Name	Direction 5'-3'	Gene/Taxa	Thermocycling condition	amplicons size	Reference
BacF	CCTACGGGNGGCWGCAG (forward)	Nuclear 16S - bacteria	27 cycles of 94°C (30s), 54°C (30s), 68°C (45s), and 68°C extension (7m)	~450 bp	Caporaso et al. 2011
806R	GACTACHVGGGTATCTAATCC (reverse)	Nuclear 16S - bacteria		~450 bp	
Uni18SF	AGGGCAAKYCTGGTGCCAGC (forward)	Nuclear 18S - Total Eukaryotes	30 cycles of 94°C (30s), 55°C (30s), 72°C (90s), and 72°C extension (7m)	~450 bp	Zhan et al. 2013
Uni18SR	GRCGGTATCTRATCGYCTT (reverse)	Nuclear 18S - Total Eukaryotes		~450 bp	
F1	AAGGGCACCACAAGAACGC (forward)	Nuclear 18S – Foraminifera	<45 cycles of 94°C (20s), 52°C (20s), 72°C (20s), and 72°C extension (7m)	~250 bp	Pawlowski et al. 2014a
15.3	CCACCTATCACAYAATCATG (reverse)	Nuclear 18S - Foraminifera		~250 bp	

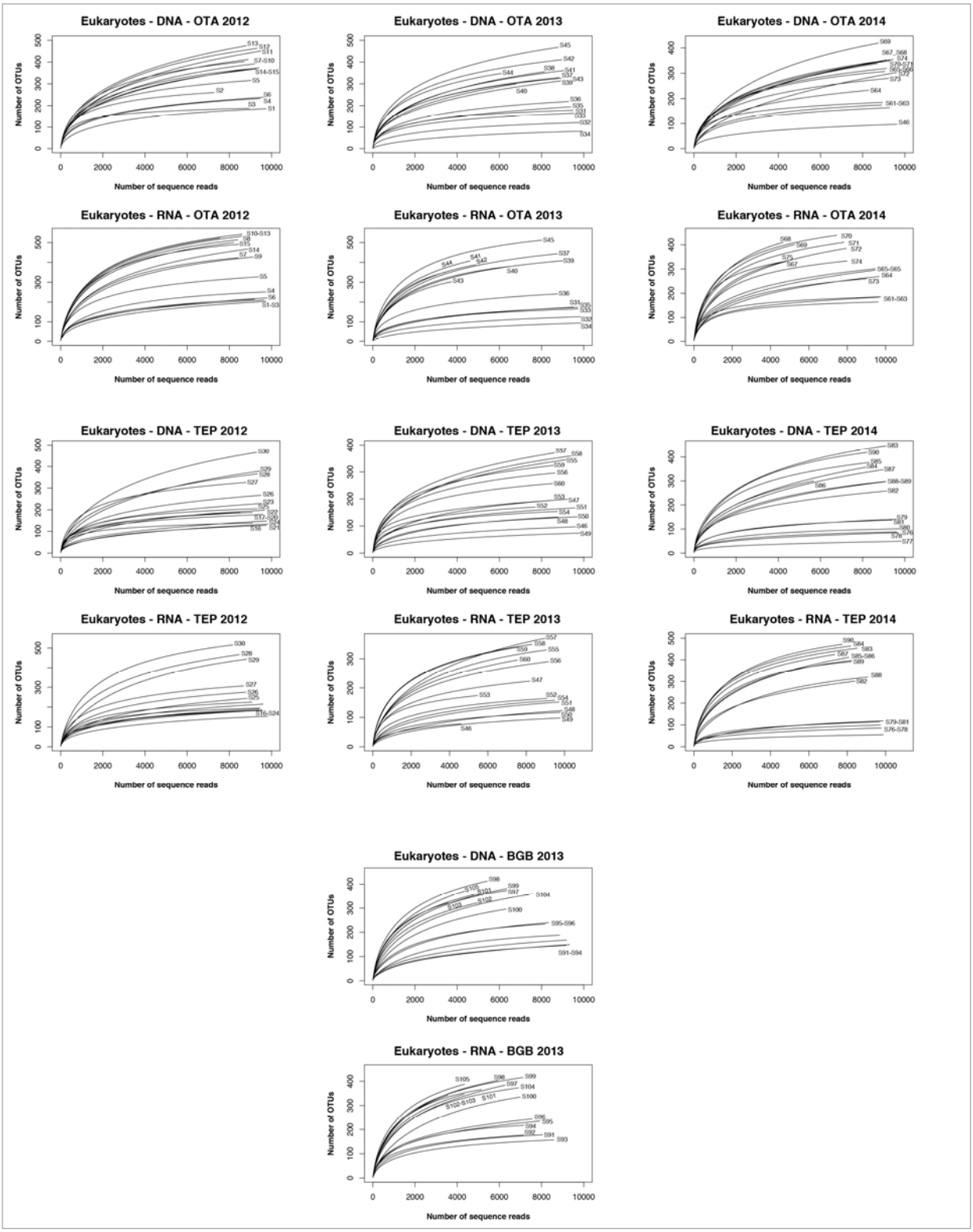
Appendix 2A Rarefaction curves plotting the number of reads by the numbers of identified foraminiferal Operational taxonomic Unit (OTU) for each investigated sample and from both DNA/RNA molecules. Samples highlighted in red did not reach saturation. OTA = Otanerau, TEP = Te Pangu; BGB = Big Glory Bay.



Appendix 2B. Rarefaction curves plotting the number of reads by the numbers of identified bacterial Operational taxonomic Unit (OTU) for each investigated sample and from both DNA/RNA molecules. Samples highlighted in red did not reach saturation. OTA = Otanerau, TEP = Te Pangu; BGB = Big Glory Bay.



Appendix 2C. Rarefaction curves plotting the number of reads by the numbers of identified eukaryotic Operational taxonomic Unit (OTU) for each investigated sample and from both DNA/RNA molecules. OTA = Otanerau, TEP = Te Pangu; BGB = Big Glory Bay.



Appendix 3A. List of the foraminiferal (For) Operational Taxonomic Units (OTUs) and corresponding taxa names with high-quality scores (Q0-Q2) and which received a specific GenCodeID. Taxa names are indicated at lowest possible taxonomic level. EG = Eco-Group.

OTU	Taxa Name	DNA EG	RNA EG	DNA Quality	RNA Quality	GenCodeID
OTU_910	*Buliminella tenuata*	I	I	0	1	fo-0041-0-I
OTU_2363	*Psammophaga* sp.	I	II	0	0	fo-0109-0-I
OTU_1163	*Zaninettia manaarensis*	II	II	0	2	fo-0068-0-II
OTU_1499	*Planoglabratella opercularis*	II	III	1	1	fo-0133-1-II
OTU_1878	*Bathysiphon* sp.	IV	IV	1	2	fo-0011-1-IV
OTU_1745	*Vellaria pellucidus*	IV	IV	1	1	fo-0030-1-IV
OTU_2231	*Bolivina* sp.	V	V	1	1	fo-0129-1-V
OTU_2129	*Spirophthalmidium* sp.	V	V	1	0	fo-0084-1-V
OTU_1487	*Rosalina* sp.	V	V	0	0	fo-0038-0-V
OTU_1033	*Psammophaga* sp.	I	III	2	2	fo-0003-2-I
OTU_1055	*Astrammina rara*	I	I	1	1	fo-0131-1-I
OTU_1067	*Nemogullmia* sp.	I	I	1	0	fo-0083-1-I
OTU_1132	*Bathysiphon* sp.	II	NA	2	4	fo-0059-2-II
OTU_1133	*Nemogullmia* sp.	I	NA	1	4	fo-0035-1-I
OTU_1144	Unknown taxa	II	I	1	0	fo-0081-1-II
OTU_115	*Astrammina rara*	I	II	1	2	fo-0093-1-I
OTU_1191	Unknown taxa	I	I	1	1	fo-0100-1-I
OTU_1283	Unknown taxa	II	I	1	1	fo-0039-1-II
OTU_1289	Unknown taxa	I	I	1	1	fo-0028-1-I
OTU_1293	Unknown taxa	II	I	1	1	fo-0073-1-II
OTU_1347	*Bathysiphon* sp.	II	NA	2	4	fo-0102-2-II
OTU_1363	Unknown Rotaliid	II	NA	2	4	fo-0116-2-II
OTU_1364	*Rosalina* sp.	V	V	1	1	fo-0008-1-V
OTU_1395	Allogromida	V	IV	2	1	fo-0053-2-V
OTU_1453	Unknown Rotaliid	I	I	2	2	fo-0101-2-I
OTU_1480	*Rosalina* sp.	V	V	1	1	fo-0017-1-V
OTU_1481	*Rosalina* sp.	V	V	1	1	fo-0051-1-V
OTU_1488	*Notorotalia finlayi*	I	II	2	2	fo-0043-2-I
OTU_1494	Saccamminidae	III	III	2	2	fo-0117-2-III
OTU_1497	*Cribrostomoides* sp.	I	I	1	1	fo-0094-1-I
OTU_1500	*Hippocrepinella hirudinea*	I	I	0	2	fo-0049-0-I
OTU_1503	*Rosalina* sp.	IV	V	2	2	fo-0067-2-IV
OTU_1519	*Vellaria* sp.	IV	IV	1	2	fo-0032-1-IV
OTU_1529	Saccamminidae	I	II	1	1	fo-0114-1-I
OTU_1544	*Astrammina rara*	I	I	1	2	fo-0012-1-I
OTU_1556	*Astrammina rara*	I	I	0	1	fo-0027-0-I
OTU_1560	*Rosalina* sp.	V	V	1	0	fo-0029-1-V
OTU_1567	*Vellaria pellucidus*	IV	NA	2	4	fo-0075-2-IV
OTU_1585	*Vellaria pellucidus*	V	IV	2	1	fo-0013-2-V
OTU_1593	*Planoglabratella opercularis*	I	I	1	2	fo-0010-1-I
OTU_1612	*Planoglabratella opercularis*	I	II	1	2	fo-0033-1-I
OTU_1622	*Vellaria pellucidus*	IV	IV	2	2	fo-0005-2-IV
OTU_1634	*Planoglabratella opercularis*	I	I	1	1	fo-0125-1-I
OTU_1635	*Planoglabratella opercularis*	I	NA	1	4	fo-0086-1-I
OTU_1640	*Planoglabratella opercularis*	I	I	0	2	fo-0105-0-I
OTU_1644	*Hippocrepinella hirudinea*	I	I	0	1	fo-0096-0-I
OTU_1689	*Bathysiphon* sp.	I	I	0	1	fo-0026-0-I
OTU_1690	*Planoglabratella opercularis*	I	I	1	2	fo-0089-1-I
OTU_1696	*Vellaria pellucidus*	III	IV	2	2	fo-0016-2-III
OTU_1747	*Bolivina* sp.	V	NA	2	4	fo-0113-2-V
OTU_1749	*Vellaria pellucidus*	IV	III	1	3	fo-0022-1-IV
OTU_1756	*Psammophaga* sp.	I	I	0	1	fo-0065-0-I
OTU_1764	*Vellaria pellucidus*	IV	III	2	2	fo-0055-2-IV
OTU_1766	*Bathysiphon* sp.	I	I	0	1	fo-0107-0-I
OTU_1782	*Planoglabratella opercularis*	II	III	2	2	fo-0085-2-II
OTU_1882	*Cribrostomoides* sp.	IV	IV	1	2	fo-0134-1-IV
OTU_1902	Monothalamea	II	I	1	2	fo-0066-1-II
OTU_1905	*Micrometula* sp.	II	II	2	2	fo-0045-2-II
OTU_1908	*Bathysiphon* sp.	III	IV	2	2	fo-0019-2-III
OTU_1911	Saccamminidae	I	I	1	2	fo-0080-1-I
OTU_1924	*Nouria polymorphinoides*	I	I	1	2	fo-0087-1-I

OTU_1963	*Psammophaga* sp.	I	I	1	1	fo-0108-1-I
OTU_1990	*Rosalina* sp.	V	V	2	1	fo-0014-2-V
OTU_1994	*Psammophaga* sp.	I	II	2	2	fo-0127-2-I
OTU_2039	Saccamminidae	IV	NA	1	4	fo-0058-1-IV
OTU_2059	*Nouria polymorphinoides*	II	I	1	1	fo-0110-1-II
OTU_2062	*Bolivina* sp	IV	IV	2	1	fo-0044-2-IV
OTU_208	*Notodendrodes antarctikos*	I	I	1	1	fo-0077-1-I
OTU_2087	Saccamminidae	IV	III	2	2	fo-0054-2-IV
OTU_2091	*Rosalina* sp.	IV	V	2	2	fo-0064-2-IV
OTU_2146	*Cribrostomoides* sp.	II	I	2	2	fo-0074-2-II
OTU_2175	*Bolivina* sp.	IV	V	2	1	fo-0042-2-IV
OTU_2185	*Psammophaga* sp.	III	III	2	4	fo-0098-2-III
OTU_2253	*Nonionoides grateloupii*	I	NA	1	4	fo-0132-1-I
OTU_2291	*Tinogullmia* sp.	I	II	1	2	fo-0115-1-I
OTU_2299	*Psammophaga* sp.	III	III	2	2	fo-0050-2-III
OTU_2328	*Rubratella* sp.	I	I	1	1	fo-0092-1-I
OTU_2351	*Nonion* sp.	I	I	0	1	fo-0031-0-I
OTU_2352	*Nonion* sp.	I	I	1	0	fo-0071-1-I
OTU_2368	*Psammophaga* sp.	II	NA	2	4	fo-0070-2-II
OTU_2404	*Crithionina hispida*	III	III	2	2	fo-0082-2-III
OTU_2423	*Pulleniatina obliquiloculata*	II	II	2	2	fo-0090-2-II
OTU_2426	*Crithionina hispida*	II	II	2	1	fo-0112-2-II
OTU_2442	*Reophax* sp.	I	I	2	2	fo-0056-2-I
OTU_2456	*Pulleniatina obliquiloculata*	II	II	1	1	fo-0079-1-II
OTU_2463	*Cribrothalammina alba*	I	I	1	1	fo-0062-1-I
OTU_2465	*Pulleniatina obliquiloculata*	II	II	2	1	fo-0128-2-II
OTU_2473	*Cribrothalammina alba*	I	I	1	2	fo-0061-1-I
OTU_248	*Notodendrodes antarctikos*	I	I	1	0	fo-0091-1-I
OTU_2480	*Cribrothalammina alba*	II	I	1	2	fo-0057-1-II
OTU_2484	*Pulleniatina obliquiloculata*	I	I	1	2	fo-0097-1-I
OTU_2520	Saccamminidae	II	II	2	2	fo-0015-2-II
OTU_2521	Saccamminidae	II	II	2	2	fo-0023-2-II
OTU_2525	Allogromida	I	II	2	2	fo-0001-2-I
OTU_2528	Saccamminidae	II	I	2	1	fo-0021-2-II
OTU_2550	*Psammophaga* sp.	II	I	2	1	fo-0106-2-II
OTU_2551	*Psammophaga* sp.	I	NA	1	4	fo-0121-1-I
OTU_2553	*Psammophaga* sp.	I	I	1	1	fo-0104-1-I
OTU_2572	Saccamminidae	V	V	1	2	fo-0024-1-V
OTU_2574	Allogromiida	II	I	1	2	fo-0099-1-II
OTU_2588	Allogromida	II	II	2	2	fo-0009-2-II
OTU_2595	Saccamminidae	IV	IV	2	1	fo-0006-2-IV
OTU_2615	Saccamminidae	V	IV	2	2	fo-0004-2-V
OTU_2616	Saccamminidae	IV	IV	2	2	fo-0037-2-IV
OTU_299	*Notodendrodes antarctikos*	I	NA	2	4	fo-0122-2-I
OTU_341	*Pulleniatina obliquiloculata*	III	III	2	2	fo-0119-2-III
OTU_362	Unknown taxa	II	NA	2	4	fo-0046-2-II
OTU_41	*Notodendrodes antarctikos*	I	NA	2	4	fo-0118-2-I
OTU_412	*Nemogullmia longevariabilis*	II	I	1	2	fo-0052-1-II
OTU_441	*Siphonaperta* sp.	I	I	1	2	fo-0103-1-I
OTU_450	*Astrammina rara*	I	I	1	1	fo-0111-1-I
OTU_455	Crithionina hispida	II	I	0	1	fo-0095-0-II
OTU_464	Crithionina hispida	I	I	1	1	fo-0130-1-I
OTU_483	*Siphonaperta* sp.	I	I	1	1	fo-0126-1-I
OTU_584	*Cribrostomoides* sp.	II	NA	1	4	fo-0120-1-II
OTU_691	*Nemogullmia* sp.	I	I	1	0	fo-0072-1-I
OTU_707	*Notodendrodes antarctikos*	II	II	1	2	fo-0047-1-II
OTU_716	*Bulimina aculeata*	I	II	1	1	fo-0076-1-I
OTU_792	*Psammophaga* sp.	I	I	2	4	fo-0025-2-I
OTU_793	*Nouria polymorphinoides*	II	III	2	1	fo-0078-2-II
OTU_919	*Buliminella tenuata*	II	I	1	2	fo-0069-1-II
OTU_1600	*Vellaria pellucidus*	III	IV	3	2	fo-0123-3-III
OTU_1796	*Vellaria pellucidus*	IV	III	3	3	fo-0124-3-III
OTU_2330	Saccamminidae	III	II	3	2	fo-0122-3-III

Appendix 3B. List of the bacterial (Bac) Operational Taxonomic Units (OTUs) and corresponding taxa names with high-quality scores (Q0-Q2) and which received a specific GenCodeID. Taxa names are indicated at lowest possible taxonomic level. EG = Eco-Group.

OTU	Taxa Name	DNA EG	RNA EG	DNA Quality	RNA Quality	GenCodeID
OTU_6507	*Nitrosospira*	I	I	0	0	ba-0139-0-I
OTU_14913	*Blastopirellula*	I	I	0	0	ba-0197-0-I
OTU_10512	*Thiococcus*	I	II	1	2	ba-0087-1-I
OTU_7118	*Halochromatium*	III	III	1	1	ba-0140-1-III
OTU_1103	*Desulfosarcina*	IV	IV	1	1	ba-0122-1-IV
OTU_15783	*Aestuariicola*	IV	IV	0	0	ba-0179-0-IV
OTU_19316	*Dethiosulfovibrio*	V	V	1	0	ba-0036-1-V
OTU_21052	*Ruminococcus*	V	V	0	0	ba-0178-0-V
OTU_1055	*Desulfobacter*	V	V	0	0	ba-0113-0-V
OTU_909	*Desulfacinum*	I	I	0	0	ba-0115-0-I
OTU_14910	*Rhodopirellula*	I	I	1	1	ba-0183-1-I
OTU_1435	*Desulfosarcina*	II	I	1	1	ba-0123-1-II
OTU_19160	Acidobacteria_Gp21	II	II	1	1	ba-0212-1-II
OTU_14914	*Blastopirellula*	II	II	1	1	ba-0203-1-II
OTU_10774	*Thioprofundum*	II	II	1	1	ba-0157-1-II
OTU_852	*Desulfonema*	II	II	0	1	ba-0075-0-II
OTU_902	*Syntrophobacter*	II	I	0	0	ba-0220-0-II
OTU_21028	*Thiohalobacter*	III	III	1	1	ba-0030-1-III
OTU_1192	*Saccharospirillum*	IV	IV	1	2	ba-0001-1-IV
OTU_1183	*Psychromonas*	IV	IV	1	1	ba-0040-1-IV
OTU_10629	*Thioprofundum*	IV	IV	1	2	ba-0108-1-IV
OTU_21064	*Thiohalobacter*	IV	IV	1	1	ba-0214-1-IV
OTU_19186	*Tindallia*	V	V	1	0	ba-0046-1-V
OTU_19231	*Thermanaerovibrio*	V	V	1	0	ba-0211-1-V
OTU_858	*Desulfomicrobium*	V	V	1	1	ba-0120-1-V
OTU_6440	*Thioprofundum*	IV	IV	1	1	ba-0002-1-IV
OTU_19163	*Sulfurovum*	V	V	1	1	ba-0003-1-V
OTU_6441	*Thioprofundum*	I	I	1	1	ba-0004-1-I
OTU_21025	*Sulfurovum*	IV	V	1	1	ba-0005-1-IV
OTU_19140	*Sulfurovum*	IV	V	1	1	ba-0006-1-IV
OTU_1207	*Vibrio*	III		3	1	ba-0007-3-III
OTU_2245	*Cystobacter*	II	II	1	1	ba-0008-1-II
OTU_22444	*Sulfurovum*	IV	V	1	1	ba-0009-1-IV
OTU_20309	*Sulfurovum*	IV	V	2	1	ba-0010-2-IV
OTU_6446	*Byssovorax*	II	III	1	1	ba-0011-1-II
OTU_22713	*Sulfurovum*	III	IV	1	1	ba-0012-1-III
OTU_850	*Thiohalomonas*	II	II	1	1	ba-0013-1-II
OTU_16161	*Aestuariicola*	V	V	1	1	ba-0014-1-V
OTU_988	*Desulfatiferula*	V	V	1	1	ba-0015-1-V
OTU_2553	*Thiohalocapsa*	IV	IV	1	1	ba-0016-1-IV
OTU_2361	*Thioprofundum*	II	II	1	1	ba-0017-1-II
OTU_6443	*Haliea*	IV	V	2	1	ba-0018-2-IV
OTU_2572	*Alkalispirillum*	III	IV	1	1	ba-0019-1-III
OTU_18677	*Thermodesulfovibrio*	I	I	1	1	ba-0020-1-I
OTU_15859	*Lewinella*	V	V	1	1	ba-0021-1-V
OTU_2646	*Thioprofundum*	II	I	1	1	ba-0022-1-II
OTU_846	*Pelobacter*	I	I	1	1	ba-0023-1-I
OTU_16162	*Sunxiuqinia*	IV	IV	2	1	ba-0024-2-IV
OTU_847	*Desulfonema*	II	I	1	1	ba-0025-1-II
OTU_266	*Desulfosarcina*	III	III	1	1	ba-0026-1-III
OTU_2659	*Thiohalophilus*	IV	IV	1	1	ba-0027-1-IV
OTU_7081	*Thioprofundum*	II	II	1	1	ba-0028-1-II
OTU_6466	*Congregibacter*	I	II	1	2	ba-0029-1-I
OTU_2783	*Thiococcus*	IV	IV	1	1	ba-0031-1-IV
OTU_839	*Desulfobacula*	IV	IV	1	1	ba-0032-1-IV
OTU_4783	*Desulfosarcina*	III	III	1	3	ba-0033-1-III
OTU_1991	*Desulfosarcina*	IV	V	2	2	ba-0034-2-IV

OTU_6445	*Marinobacterium*	IV	IV	2	1	ba-0035-2-IV
OTU_5198	*Psychromonas*	IV	V	1	2	ba-0037-1-IV
OTU_21436	*Sulfurovum*	V	V	1	1	ba-0038-1-V
OTU_6541	*Thioprofundum*	I	I	1	1	ba-0039-1-I
OTU_854	*Desulfotalea*	V	IV	1	2	ba-0041-1-V
OTU_851	*Desulfobulbus*	I	I	1	1	ba-0042-1-I
OTU_183	Acidobacteria_Gp23	I	II	1	2	ba-0043-1-I
OTU_6470	*Psychromonas*	IV	III	1	2	ba-0044-1-IV
OTU_363	*Anaeromyxobacter*	II	II	1	1	ba-0045-1-II
OTU_1561	*Thiohalomonas*	II	I	1	1	ba-0047-1-II
OTU_860	*Desulfosalsimonas*	II	I	1	1	ba-0048-1-II
OTU_10828	*Cystobacter*	I	I	1	2	ba-0049-1-I
OTU_6444	*Thiohalocapsa*	IV	IV	1	1	ba-0050-1-IV
OTU_6474	*Thioprofundum*	I	I	1	1	ba-0051-1-I
OTU_7593	*Oleispira*	III	III	2	1	ba-0052-2-III
OTU_20006	*Ilumatobacter*	III	II	2	1	ba-0053-2-III
OTU_848	*Desulfosarcina*	V	I	1	0	ba-0054-1-V
OTU_20906	*Thiohalobacter*	II	I	1	1	ba-0055-1-II
OTU_6468	*Thioprofundum*	II	II	1	1	ba-0056-1-II
OTU_6442	*Thiohalocapsa*	I	I	1	1	ba-0057-1-I
OTU_8492	*Syntrophorhabdus*	II	II	1	1	ba-0058-1-II
OTU_10304	*Thioprofundum*	III	II	1	2	ba-0059-1-III
OTU_16168	*Mucilaginibacter*	V	V	1	2	ba-0060-1-V
OTU_6483	*Thiohalocapsa*	III		3	1	ba-0061-3-III
OTU_845	*Desulfonema*	V	V	1	2	ba-0062-1-V
OTU_6450	*Spongiibacter*	I	I	1	1	ba-0063-1-I
OTU_37	*Desulfobacterium*	IV	III	2	3	ba-0064-2-IV
OTU_870	*Megasphaera*	IV	III	2	3	ba-0065-2-IV
OTU_6463	*Thioprofundum*	II	II	1	2	ba-0066-1-II
OTU_372	*Desulfosarcina*	V	V	1	2	ba-0067-1-V
OTU_7820	*Marinobacter*	II	II	2	2	ba-0068-2-II
OTU_2717	*Thiohalomonas*	III	III	1	2	ba-0069-1-III
OTU_6447	*Halochromatium*	III	IV	1	1	ba-0070-1-III
OTU_6453	*Thioalkalivibrio*	II	I	1	1	ba-0071-1-II
OTU_20863	*Sulfurimonas*	V	IV	1	1	ba-0072-1-V
OTU_6584	*Thiococcus*	I	II	2	2	ba-0073-2-I
OTU_6780	*Thioprofundum*	II	II	1	1	ba-0074-1-II
OTU_1902	*Desulfosarcina*	II	I	1	2	ba-0076-1-II
OTU_21032	*Pelagibius*	I	I	1	1	ba-0077-1-I
OTU_10241	*Thiohalocapsa*	IV	IV	2	1	ba-0078-2-IV
OTU_6462	*Pseudomonas*	II	IV	1	1	ba-0079-1-II
OTU_668	*Desulfobulbus*	III	III	1	1	ba-0080-1-III
OTU_865	*Desulfobacter*	V	V	1	1	ba-0081-1-V
OTU_6460	*Thiohalocapsa*	III	IV	1	1	ba-0082-1-III
OTU_882	*Desulfacinum*	I	I	1	1	ba-0083-1-I
OTU_9814	*Haliea*	IV	IV	2	1	ba-0084-2-IV
OTU_4404	*Thiohalomonas*	II	II	1	1	ba-0085-1-II
OTU_6455	*Thioprofundum*	I	I	2	2	ba-0086-2-I
OTU_7009	*Saccharospirillum*	II	II	1	2	ba-0088-1-II
OTU_7584	*Thiohalocapsa*	IV	IV	1	1	ba-0089-1-IV
OTU_855	*Desulfopila*	IV	V	1	3	ba-0090-1-IV
OTU_21031	*Thalassobius*	IV	IV	1	1	ba-0091-1-IV
OTU_10080	*Thiohalocapsa*	IV	IV	2	1	ba-0092-2-IV
OTU_849	*Desulforhopalus*	V	V	1	1	ba-0093-1-V
OTU_857	*Desulfobacterium*	II	II	2	1	ba-0094-2-II
OTU_17898	*Fulvivirga*	II	II	1	2	ba-0095-1-II
OTU_995	*Desulfobulbus*	I	I	1	1	ba-0096-1-I
OTU_68	*Marinobacter*	II		3	1	ba-0097-3-II
OTU_19230	*Nitrospina*	I	I	1	0	ba-0098-1-I
OTU_4770	*Desulfosalsimonas*	I	II	2	1	ba-0099-2-I
OTU_8335	*Thioprofundum*	III	II	1	1	ba-0100-1-III
OTU_862	*Desulforhopalus*	I	III	1	1	ba-0101-1-I
OTU_7427	*Thioprofundum*	I	I	1	1	ba-0102-1-I
OTU_871	*Desulforhopalus*	V	V	1	1	ba-0103-1-V

OTU_4766	*Desulfosarcina*	II	I	1	1	ba-0104-1-II
OTU_3637	*Halochromatium*	III	III	1	1	ba-0105-1-III
OTU_21051	*Thiohalobacter*	III	III	1	1	ba-0106-1-III
OTU_21029	*Geminicoccus*	I	I	1	0	ba-0107-1-I
OTU_711	*Desulfobulbus*	IV	IV	1	1	ba-0109-1-IV
OTU_21036	*Thiohalobacter*	II	I	1	1	ba-0110-1-II
OTU_2072	*Desulfosarcina*	V	V	1	1	ba-0111-1-V
OTU_1955	*Anaeromyxobacter*	I	II	1	0	ba-0112-1-I
OTU_2721	*Desulfobulbus*	II	II	1	1	ba-0114-1-II
OTU_19668	*Ilumatobacter*	II	I	0	1	ba-0116-0-II
OTU_11390	*Thiococcus*	III	III	1	2	ba-0117-1-III
OTU_404	*Syntrophobacter*	I	II	1	1	ba-0118-1-I
OTU_6467	*Marinobacter*	I	II	2	2	ba-0119-2-I
OTU_6459	*Thiohalocapsa*	IV	III	1	3	ba-0121-1-IV
OTU_5	*Desulfosalsimonas*	II	I	2	1	ba-0124-2-II
OTU_6477	*Psychromonas*	IV	IV	1	1	ba-0125-1-IV
OTU_867	*Desulfobacterium*	V		2	2	ba-0126-2-V
OTU_6485	*Halochromatium*	IV	IV	1	1	ba-0127-1-IV
OTU_671	*Thioprofundum*	II	II	1	1	ba-0128-1-II
OTU_7008	*Thioprofundum*	I	I	1	1	ba-0129-1-I
OTU_873	*Desulfosalsimonas*	II	I	1	1	ba-0130-1-II
OTU_20009	*Acetivibrio*	IV	III	1	3	ba-0131-1-IV
OTU_20004	*Aminomonas*	II	III	1	3	ba-0132-1-II
OTU_3499	*Thioalkalivibrio*	IV	IV	1	1	ba-0133-1-IV
OTU_6451	*Haliea*	III	IV	2	2	ba-0134-2-III
OTU_45	*Desulfosarcina*	II	II	2	1	ba-0135-2-II
OTU_576	*Geobacter*	II	II	1	1	ba-0136-1-II
OTU_20445	*Pseudoruegeria*	II	I	2	2	ba-0137-2-II
OTU_8052	*Anaeromyxobacter*	II	I	1	1	ba-0138-1-II
OTU_21030	*Pelagibius*	II	I	1	1	ba-0141-1-II
OTU_7164	*Thioprofundum*	II	I	1	2	ba-0142-1-II
OTU_859	*Desulfosarcina*	V	V	1	1	ba-0143-1-V
OTU_895	*Syntrophus*	I	I	1	1	ba-0144-1-I
OTU_452	*Desulforhopalus*	IV	V	1	1	ba-0145-1-IV
OTU_4864	*Congregibacter*	I	II	1	1	ba-0146-1-I
OTU_2798	*Thiohalocapsa*	IV	IV	1	1	ba-0147-1-IV
OTU_6515	*Thiohalophilus*	IV	IV	1	1	ba-0148-1-IV
OTU_94	*Thioprofundum*	I	I	1	1	ba-0149-1-I
OTU_6473	*Halochromatium*	II	I	1	1	ba-0150-1-II
OTU_19667	*Alkaliphilus*	V	III	1	3	ba-0151-1-V
OTU_2249	*Thiohalocapsa*	III	II	2	2	ba-0152-2-III
OTU_2376	*Desulfosarcina*	III	III	1	3	ba-0153-1-III
OTU_16166	*Lewinella*	V		2	2	ba-0154-2-V
OTU_353	*Latescibacteria*	IV	III	1	3	ba-0155-1-IV
OTU_16167	*Aestuariicola*	V		2	2	ba-0156-2-V
OTU_7203	*Thioprofundum*	I	I	1	1	ba-0158-1-I
OTU_10511	*Thiohalocapsa*	III		3	1	ba-0159-3-III
OTU_1175	*Desulfosarcina*	II	II	1	2	ba-0160-1-II
OTU_21060	*Natranaerovirga*	V	IV	1	1	ba-0161-1-V
OTU_16317	*Aestuariicola*	IV	V	1	1	ba-0162-1-IV
OTU_20717	*Gaetbulicola*	II	III	2	3	ba-0163-2-II
OTU_853	*Desulforhopalus*	IV		2	4	ba-0164-2-IV
OTU_6456	*Haliea*	III	III	1	3	ba-0165-1-III
OTU_21039	*Pelagibius*	II	I	1	1	ba-0166-1-II
OTU_6449	*Thiohalomonas*	III	III	1	3	ba-0167-1-III
OTU_6556	*Thioprofundum*	I	I	1	1	ba-0168-1-I
OTU_16169	*Owenweeksia*	IV	IV	1	1	ba-0169-1-IV
OTU_21468	*Thiohalobacter*	III	III	2	1	ba-0170-2-III
OTU_6558	Acidobacteria_Gp17	II	II	1	1	ba-0171-1-II
OTU_2511	*Congregibacter*	IV		2	2	ba-0172-2-IV
OTU_9153	*Thiohalocapsa*	III		4	1	ba-0173-4-III
OTU_875	*Desulfuromonas*	II	I	1	1	ba-0174-1-II
OTU_4947	*Desulfatiferula*	V	V	1	1	ba-0175-1-V
OTU_6549	*Psychromonas*	V	IV	1	1	ba-0176-1-V

OTU_901	*Thioprofundum*	I	I	1	0	ba-0177-1-I
OTU_2964	*Desulfosalsimonas*	I	I	2	1	ba-0180-2-I
OTU_16170	*Pedobacter*	IV		2	1	ba-0181-2-IV
OTU_885	*Jahnella*	II	II	1	1	ba-0182-1-II
OTU_17597	*Paludibacter*	IV	III	1	3	ba-0184-1-IV
OTU_16724	*Sunxiuqinia*	III	III	1	3	ba-0185-1-III
OTU_6550	*Litorimonas*	II	I	1	1	ba-0186-1-II
OTU_1587	*Desulfosarcina*	I	II	1	1	ba-0187-1-I
OTU_1468	*Thioprofundum*	II	I	1	1	ba-0188-1-II
OTU_303	*Desulfosalsimonas*	I	II	2	1	ba-0189-2-I
OTU_181	*Latescibacteria*	III	III	1	3	ba-0190-1-III
OTU_16605	*Aestuariicola*	II	III	2	3	ba-0191-2-II
OTU_912	*Thiohalomonas*	II	III	2	3	ba-0192-2-II
OTU_6464	*Thioprofundum*	I	III	2	1	ba-0193-2-I
OTU_10158	*Thiohalocapsa*	III		4	1	ba-0194-4-III
OTU_17034	*Sunxiuqinia*	IV	V	1	1	ba-0195-1-IV
OTU_21070	*Celeribacter*	IV		2	3	ba-0196-2-IV
OTU_6537	*Oleiphilus*	III		3	1	ba-0198-3-III
OTU_6535	*Thioprofundum*	I	I	1	1	ba-0199-1-I
OTU_4895	*Desulfosarcina*	IV		2	1	ba-0200-2-IV
OTU_12861	*Thiohalophilus*	III	III	2	2	ba-0201-2-III
OTU_21057	*Sulfurimonas*	III	IV	2	2	ba-0202-2-III
OTU_7622	*Thiohalocapsa*	III	IV	2	1	ba-0204-2-III
OTU_9909	*Thiohalocapsa*	IV		2	1	ba-0205-2-IV
OTU_20007	*Ilumatobacter*	II	II	1	2	ba-0206-1-II
OTU_17118	*Aestuariicola*	I	I	2	2	ba-0207-2-I
OTU_9179	*Thiohalocapsa*	IV		4	1	ba-0208-4-IV
OTU_8215	*Thioprofundum*	II	II	1	1	ba-0209-1-II
OTU_12749	*Thiohalomonas*	I	I	1	1	ba-0210-1-I
OTU_16263	*Maribacter*	II	III	2	3	ba-0213-2-II
OTU_21056	*Parvibaculum*	I	I	1	1	ba-0215-1-I
OTU_21037	*Anderseniella*	I	II	1	1	ba-0216-1-I
OTU_1965	*Desulfosarcina*	V		2	2	ba-0217-2-V
OTU_21035	*Sulfurovum*	V	V	1	1	ba-0218-1-V
OTU_13836	*Thiohalobacter*	II	III	2	3	ba-0219-2-II
OTU_3502	*Desulfobulbus*	II	I	1	1	ba-0221-1-II
OTU_21034	*Maritimibacter*	I	II	2	1	ba-0222-2-I
OTU_6527	Acidobacteria_Gp17	II	II	1	2	ba-0223-1-II
OTU_6490	*Thalassomonas*	II	III	2	2	ba-0224-2-II
OTU_339	*Desulfobacca*	II	II	1	1	ba-0225-1-II
OTU_19611	*Alkaliphilus*	IV		2	3	ba-0226-2-IV
OTU_4371	*Halochromatium*	III	II	1	1	ba-0227-1-III
OTU_4429	*Syntrophobacter*	II	I	1	0	ba-0228-1-II
OTU_9606	*Thioprofundum*	I	I	1	1	ba-0229-1-I
OTU_1966	Acidobacteria_Gp10	II	II	1	1	ba-0230-1-II

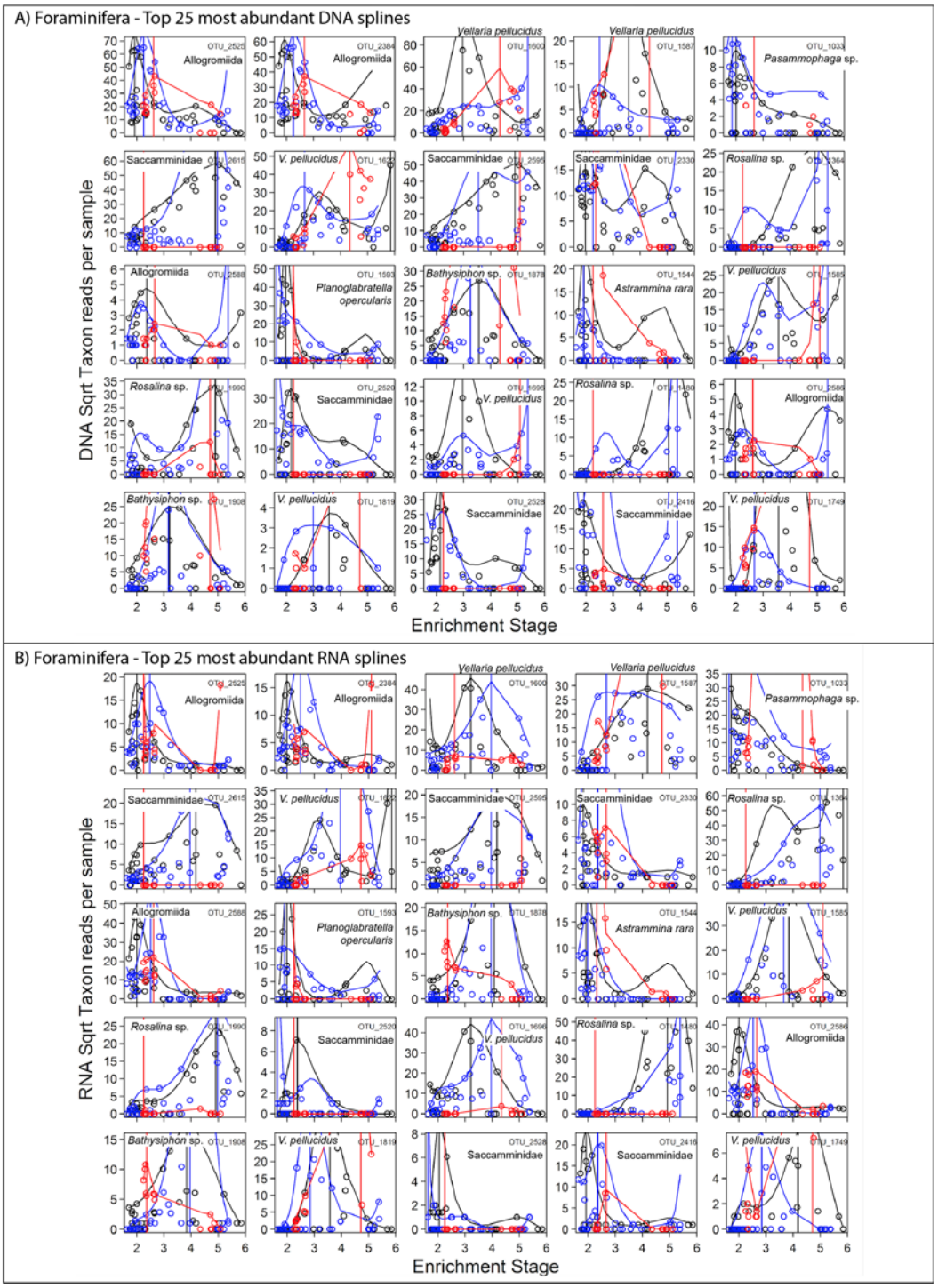
Appendix 3C. List of the eukaryotic (Euk) Operational Taxonomic Units (OTUs) and corresponding taxa names with high-quality scores (Q0-Q2) and which received a specific GenCodeID. Taxa names are indicated at lowest possible taxonomic level. EG = Eco-Groups.

OTU	Taxa Name	DNA EG	RNA EG	DNA Quality	RNA_ Quality	GenCodeID
OTU_2950	*Odontosyllis freycinetensis*	I	I	0	0	eu-0125-0-I
OTU_1540	*Bradya* sp.	I	I	0	0	eu-0098-0-I
OTU_4338	*Echinoderes setiger*	I	I	0	1	eu-0057-0-I
OTU_3663	*Prochaetosoma* sp.	II	II	1	1	eu-0092-1-II
OTU_3789	*Daptonema setosum*	III	III	1	1	eu-0059-1-III
OTU_2348	*Microstomum papillosum*	IV	IV	0	0	eu-0002-0-IV
OTU_301	*Capitella teleta*	V	V	0	0	eu-0001-0-V
OTU_224	*Philasterides armatalis*	V	V	1	1	eu-0064-1-V
OTU_1008	*Tachidius triangularis*	V	V	1	1	eu-0100-1-V
OTU_1077	*Eurycletodes laticauda*	II	III	1	1	eu-0089-1-II
OTU_1292	*Sabatieria pulchra*	V	IV	1	1	eu-0006-1-V
OTU_133	*Mesoglossus* sp.	I	I	1	1	eu-0020-1-I
OTU_175	*Schistomeringos rudolphi*	IV	IV	1	1	eu-0003-1-IV
OTU_1756	*Heterolepidoderma loricatum*	II	II	2	1	eu-0017-2-II
OTU_1955	*Diarthrodes* sp.	I	I	0	0	eu-0068-0-I
OTU_2043	*Macoma balthica*	III	I	1	1	eu-0126-1-III
OTU_2075	*Amphicorina ascidicola*	I	I	1	1	eu-0097-1-I
OTU_2081	*Cyclopina gracilis*	II	NA	2	4	eu-0023-2-II
OTU_2116	Chromadorea	IV	IV	1	1	eu-0052-1-IV
OTU_2125	*Diarthrodes* sp.	I	I	0	0	eu-0030-0-I
OTU_2141	*Stenhelia* sp.	II	II	1	1	eu-0015-1-II
OTU_2161	*Nemertoderma westbladi*	I	I	0	0	eu-0094-0-I
OTU_2171	*Canuella perplexa*	I	I	1	0	eu-0079-1-I
OTU_2182	Maxillopoda	III	IV	1	1	eu-0034-1-III
OTU_2200	*Argestigens* sp.	I	I	1	1	eu-0047-1-I
OTU_2202	*Cryothecomonas aestivalis*	II	II	2	2	eu-0122-2-II
OTU_2208	Maxillopoda	I	I	1	2	eu-0153-1-I
OTU_2241	*Prosphaerosyllis longipapillata*	II	II	1	1	eu-0010-1-II
OTU_2245	*Diarthrodes* sp.	I	I	0	0	eu-0049-0-I
OTU_2320	*Heterolepidoderma loricatum*	II	II	1	2	eu-0036-1-II
OTU_2405	*Sphaerosyllis hirsuta*	I	I	0	1	eu-0022-0-I
OTU_2408	*Trimastix* sp.	V	V	0	1	eu-0090-0-V
OTU_2453	*Erinaceusyllis kathrynae*	II	II	1	1	eu-0018-1-II
OTU_2499	Chromadorea	I	I	1	1	eu-0056-1-I
OTU_2500	Chromadorea	I	I	1	0	eu-0141-1-I
OTU_2623	*Berthella* sp.	V	V	1	1	eu-0046-1-V
OTU_2852	Chromadorea	II	I	1	1	eu-0035-1-II
OTU_2873	Chromadorea	IV	IV	1	1	eu-0075-1-IV
OTU_2875	Chromadorea	IV	IV	1	1	eu-0161-1-IV
OTU_2883	*Gyratrix* sp.	II	NA	1	4	eu-0139-1-II
OTU_2896	*Sabatieria celtica*	II	II	1	1	eu-0154-1-II
OTU_2940	*Paracalanus parvus*	III	III	2	2	eu-0073-2-III
OTU_2941	*Archimacrostomum rubrocinctum*	V	V	1	1	eu-0051-1-V
OTU_2942	*Anticomidae* sp.	IV	IV	1	1	eu-0101-1-IV
OTU_2943	Stephanoecidae	IV	V	2	1	eu-0093-2-IV
OTU_2944	*Anticoma* sp.	II	II	1	1	eu-0074-1-II
OTU_2946	*Notomastus* sp.	II	II	2	1	eu-0159-2-II
OTU_2948	*Sabatieria punctata*	II	II	1	1	eu-0132-1-II
OTU_2949	*Cirrifera dumosa*	I	I	0	1	eu-0108-0-I
OTU_2961	*Anticomidae* sp.	I	I	2	1	eu-0163-2-I
OTU_2990	*Sabatieria* sp.	II	II	2	1	eu-0137-2-II
OTU_3008	*Sabatieria punctata*	I	I	0	0	eu-0157-0-I
OTU_3012	*Sabatieria punctata*	IV	IV	1	1	eu-0021-1-IV
OTU_3072	*Sabatieria punctata*	II	II	2	1	eu-0111-2-II
OTU_3087	*Sabatieria punctata*	I	I	1	0	eu-0063-1-I
OTU_3097	Chromadorea	I	I	0	1	eu-0128-0-I
OTU_345	Dothideomycetes	II	I	2	2	eu-0158-2-II

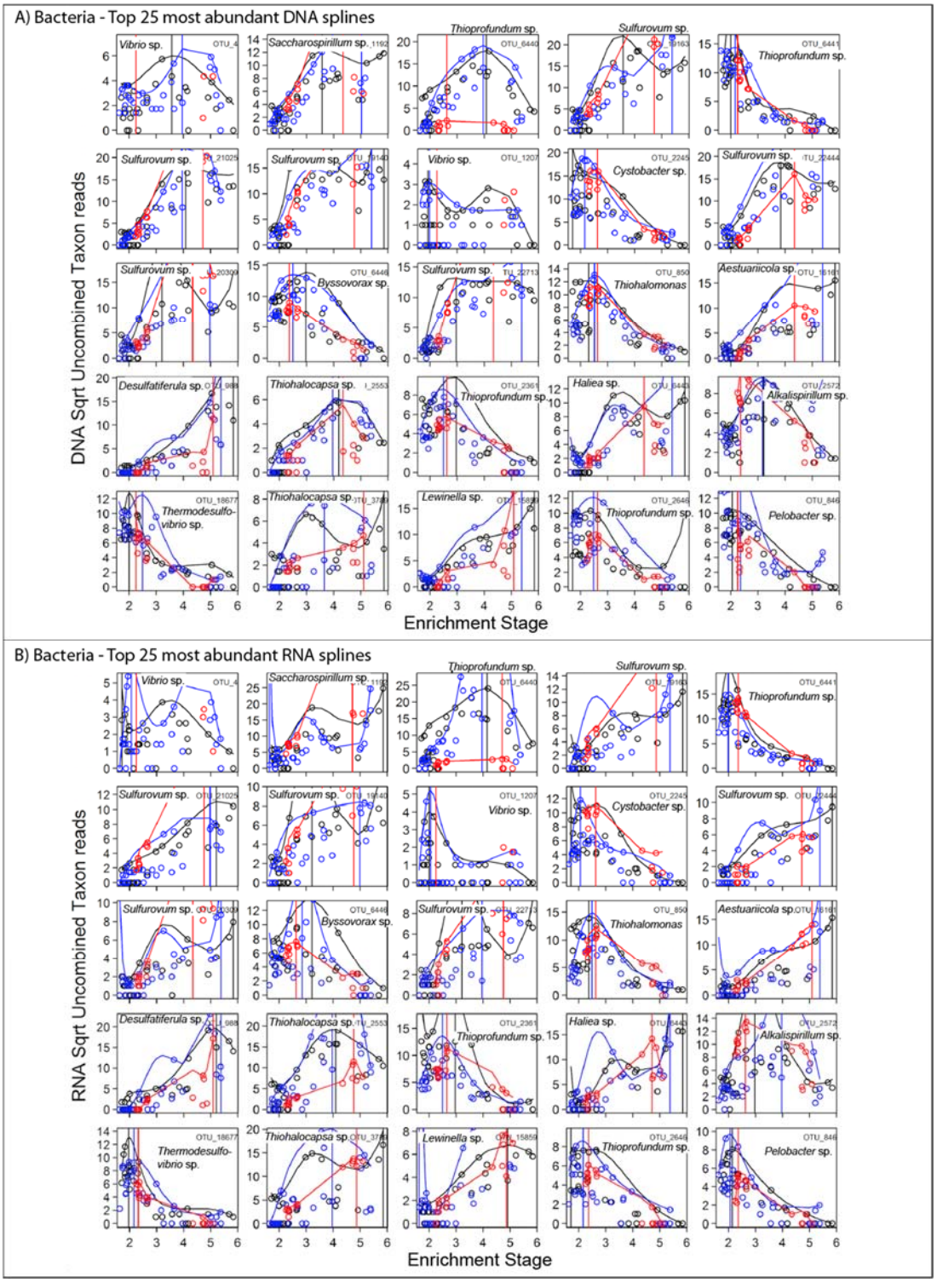
OTU_3638	*Sabatieria punctata*	IV	IV	2	2	eu-0025-2-IV
OTU_3658	*Sabatieria punctata*	V	IV	1	1	eu-0042-1-V
OTU_3667	*Cirrifera dumosa*	I	I	0	0	eu-0096-0-I
OTU_3733	Thoracostomopsidae	II	II	1	1	eu-0120-1-II
OTU_3752	*Sabatieria punctata*	IV	IV	1	1	eu-0055-1-IV
OTU_3759	*Sabatieria punctata*	IV	IV	1	1	eu-0048-1-IV
OTU_3763	*Sabatieria* sp.	II	II	2	2	eu-0104-2-II
OTU_3767	*Sabatieria punctata*	I	I	1	1	eu-0131-1-I
OTU_3768	*Chromadoropsis vivipara*	I	I	0	0	eu-0031-0-I
OTU_3781	*Arbacia* sp.	III	III	1	2	eu-0076-1-III
OTU_3785	*Apistobranchus* sp.	I	I	1	0	eu-0085-1-I
OTU_3788	*Terschellingia longicaudata*	II	II	2	2	eu-0087-2-II
OTU_3791	Chromadorea	II	II	1	1	eu-0147-1-II
OTU_3794	Chromadorea	I	I	0	0	eu-0133-0-I
OTU_3797	*Daptonema normandicum*	I	II	1	1	eu-0155-1-I
OTU_3825	*Danorhynchus* sp.	I	NA	0	4	eu-0156-0-I
OTU_3842	*Cirrophorus lyra*	I	I	0	1	eu-0008-0-I
OTU_3879	*Echinoderes capitatus*	I	I	1	1	eu-0014-1-I
OTU_3910	Thoracostomopsidae	I	I	1	1	eu-0117-1-I
OTU_3931	*Ophiodromus pugettensis*	V	-	1	4	eu-0032-1-V
OTU_3997	*Kotoracythere inconspicua*	II	II	1	1	eu-0045-1-II
OTU_4008	*Brunetorhynchus canariensis*	II	I	1	1	eu-0065-1-II
OTU_4028	Achromadora	I	I	1	1	eu-0058-1-I
OTU_4046	Oncholaimidae	III	IV	2	1	eu-0007-2-III
OTU_4058	Chromadorea	II	I	0	0	eu-0060-0-II
OTU_4132	*Pycnophyes communis*	I	I	1	1	eu-0044-1-I
OTU_4139	Dino-Group-I-Clade-1	II	III	2	2	eu-0115-2-II
OTU_4140	Rhyzophidiales	II	II	1	2	eu-0070-1-II
OTU_4142	Chromadorea	II	II	1	1	eu-0116-1-II
OTU_4143	*Theristus acer*	II	I	1	1	eu-0134-1-II
OTU_4144	Dino-Group-I-Clade-1	II	III	2	2	eu-0118-2-II
OTU_4158	*Leptocythere polymorpha*	II	II	1	1	eu-0162-1-II
OTU_4315	Oncholaimidae	IV	IV	2	1	eu-0041-2-IV
OTU_4393	Oncholaimidae	II	II	1	1	eu-0013-1-II
OTU_4464	*Astomonema* sp.	II	II	1	1	eu-0160-1-II
OTU_4491	Oncholaimidae	II	II	2	1	eu-0029-2-II
OTU_4604	*Echinoderes setiger*	I	I	0	0	eu-0077-0-I
OTU_4621	*Leptolaimus* sp.	I	III	2	2	eu-0011-2-I
OTU_4651	Chromadorea	I	I	2	2	eu-0071-2-I
OTU_4656	*Pontocypris mytiloides*	I	I	1	1	eu-0152-1-I
OTU_4704	*Leptocythere polymorpha*	I	I	1	1	eu-0130-1-I
OTU_4711	*Dolerocypria taalensis*	I	I	1	NA	eu-0004-1-I
OTU_4729	*Terschellingia longicaudata*	III	II	1	2	eu-0106-1-III
OTU_4804	Chromadorea	I	I	1	1	eu-0149-1-I
OTU_4821	*Corymorpha intermedia*	II	II	2	1	eu-0033-2-II
OTU_4830	*Aphelochaeta marioni*	I	I	1	1	eu-0078-1-I
OTU_484	*Holosticha diademata*	III	III	2	3	eu-0121-2-III
OTU_4843	Chromadorea	I	I	0	0	eu-0039-0-I
OTU_4916	*Dichromadora* sp.	I	I	1	2	eu-0082-1-I
OTU_5016	*Anthothoe chilensis*	IV	IV	1	1	eu-0144-1-IV
OTU_5017	*Ectopleura larynx*	IV	IV	1	1	eu-0019-1-IV
OTU_5025	*Kotoracythere inconspicua*	II	II	1	0	eu-0028-1-II
OTU_5031	*Cylindrolaimus communis*	II	I	2	1	eu-0140-2-II
OTU_5119	*Cytherelloidea munechikai*	I	I	1	1	eu-0095-1-I
OTU_5195	*Cephalothrix rufifrons*	I	NA	1	4	eu-0026-1-I
OTU_5218	*Halalaimus* sp.	I	I	0	0	eu-0164-0-I
OTU_5309	*Dichromadora* sp.	III	III	1	3	eu-0127-1-III
OTU_5332	*Leptolaimus* sp.	II	II	1	1	eu-0119-1-II
OTU_5499	*Leptolaimus* sp.	III	III	1	1	eu-0083-1-III
OTU_5674	*Aspidisca* sp.	V	V	1	1	eu-0066-1-V
OTU_5709	*Calomicrolaimus* sp.	II	II	2	1	eu-0054-2-II
OTU_5809	*Plagiostomum vittatum*	I	I	0	0	eu-0037-0-I
OTU_5888	Hypotrichia	II	III	2	2	eu-0129-2-II
OTU_5913	*Calomicrolaimus* sp.	II	II	1	1	eu-0072-1-II

OTU_5917	*Nothoholosticha fasciola*	IV	IV	1	1	eu-0081-1-IV
OTU_5946	Synhymeniid	II	I	0	1	eu-0136-0-II
OTU_597	*Paramphiascella fulvofasciata*	I	I	1	1	eu-0067-1-I
OTU_5986	*Aspidisca* sp.	III	III	2	3	eu-0080-2-III
OTU_5996	*Plagiostomum vittatum*	I	I	1	0	eu-0142-1-I
OTU_6051	*Plagiostomum vittatum*	I	I	0	0	eu-0102-0-I
OTU_6061	Hypotrichia	V	V	1	1	eu-0145-1-V
OTU_6073	Hypotrichia	I	II	2	2	eu-0062-2-I
OTU_6122	Hypotrichia	III	II	2	2	eu-0084-2-III
OTU_6225	Ciliophora	I	I	1	1	eu-0150-1-I
OTU_6294	Prostomatea	III	II	2	2	eu-0088-2-III
OTU_6301	*Tiarina fusa*	I	I	1	1	eu-0151-1-I
OTU_6334	Plagiopylidae	V	V	1	1	eu-0012-1-V
OTU_6380	Prostomatea	IV	III	2	2	eu-0146-2-IV
OTU_6455	*Cardiostomatella vermiformis*	II	I	1	1	eu-0148-1-II
OTU_6489	Prostomatea	III	III	2	2	eu-0114-2-III
OTU_6591	Ciliophora	II	III	2	2	eu-0107-2-II
OTU_6736	*Peritromus kahli*	V	IV	1	0	eu-0091-1-V
OTU_688	*Sabatieria pulchra*	V	IV	2	1	eu-0005-2-V
OTU_6890	*Condylostoma* sp.	IV	III	2	2	eu-0040-2-IV
OTU_6908	*Condylostoma* sp.	III	III	2	2	eu-0099-2-III
OTU_6964	Novel-Apicomplexa-Class	IV	V	1	1	eu-0123-1-IV
OTU_7205	*Aspidisca magna*	IV	IV	0	0	eu-0112-0-IV
OTU_7265	Ciliophora	V	V	1	1	eu-0105-1-V
OTU_7344	Unknown eukaryote	III	II	2	2	eu-0138-2-III
OTU_7345	Dinophyceae	II	I	1	1	eu-0103-1-II
OTU_7447	Protostomatida	III	III	0	0	eu-0109-0-III
OTU_7542	Ciliophora	V	V	0	0	eu-0124-0-V
OTU_7597	Novel-Apicomplexa-Class	II	II	1	2	eu-0038-1-II
OTU_7652	Novel-Apicomplexa-Class	II	II	2	1	eu-0024-2-II
OTU_768	*Heterolepidoderma loricatum*	II	I	1	1	eu-0016-1-II
OTU_7680	*Tracheloraphis huangi*	IV	IV	2	2	eu-0009-2-IV
OTU_7692	*Kovalevaia sulcata*	II	II	2	2	eu-0113-2-II
OTU_7715	*Tracheloraphis huangi*	IV	IV	1	1	eu-0027-1-IV
OTU_7717	*Tracheloraphis huangi*	IV	III	2	2	eu-0043-2-IV
OTU_7724	*Tracheloraphis huangi*	IV	IV	1	1	eu-0135-1-IV
OTU_7725	*Tracheloraphis huangi*	IV	IV	1	1	eu-0053-1-IV
OTU_7851	Haptoria	III	III	2	2	eu-0143-2-III
OTU_7974	*Epispathidium papilliferum*	V	V	0	0	eu-0069-0-V
OTU_7986	*Epispathidium papilliferum*	V	V	0	0	eu-0061-0-V
OTU_855	*Prosphaerosyllis isabellae*	I	I	1	0	eu-0086-1-I
OTU_856	*Eurycletodes laticauda*	I	I	1	1	eu-0050-1-I
OTU_930	*Bradya* sp.	I	I	0	0	eu-0110-0-I

Appendix 4A. Relative sequence reads abundance of the 25 most abundant foraminiferal (For) Operational Taxonomic Units (OTUs) plotted against the Enrichment Stage, using (A) environmental DNA and (B) environmental RNA data. Black circles = Marlborough Sounds (MS) Year 1, Red circles = MS Year 2, Blue circles = Stewart Island Year 2. Quantile regression splines (Tau = 0.95, df = 3, Deg = 3) fitted for each dataset with vertical lines indicating position of peak read abundance on the enrichment scale.

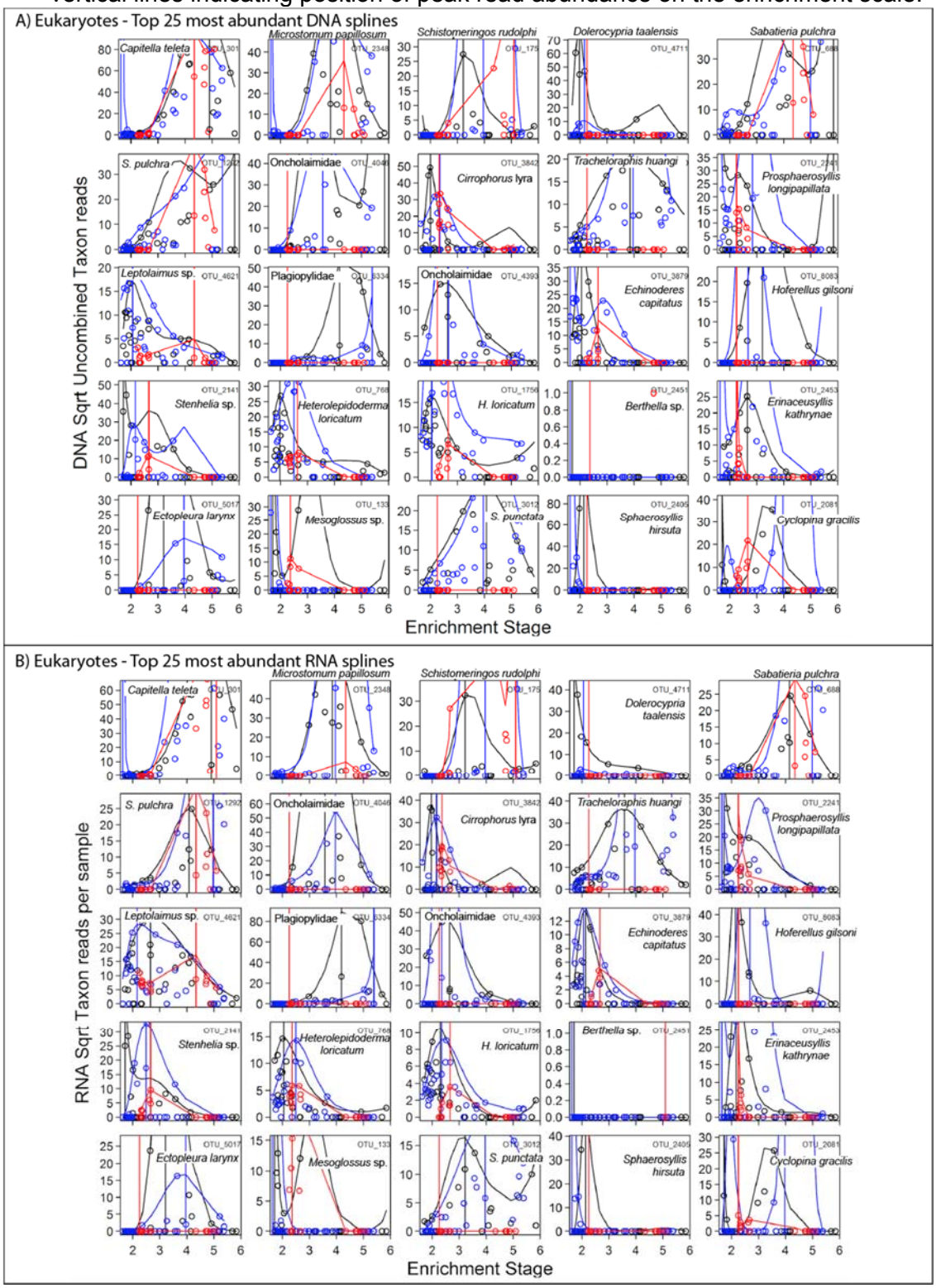


Appendix 4B. Relative sequence reads abundance of the 25 most abundant bacterial (Bac) Operational Taxonomic Units (OTUs) plotted against the Enrichment Stage, using (A) environmental DNA and (B) environmental RNA data. Black circles = Marlborough Sounds (MS) Year 1, Red circles = MS Y2, Blue circles = Stewart Island Year 2. Quantile regression splines (Tau = 0.95, df = 3, Deg = 3) fitted for each dataset with vertical lines indicating position of peak read abundance on the enrichment scale.



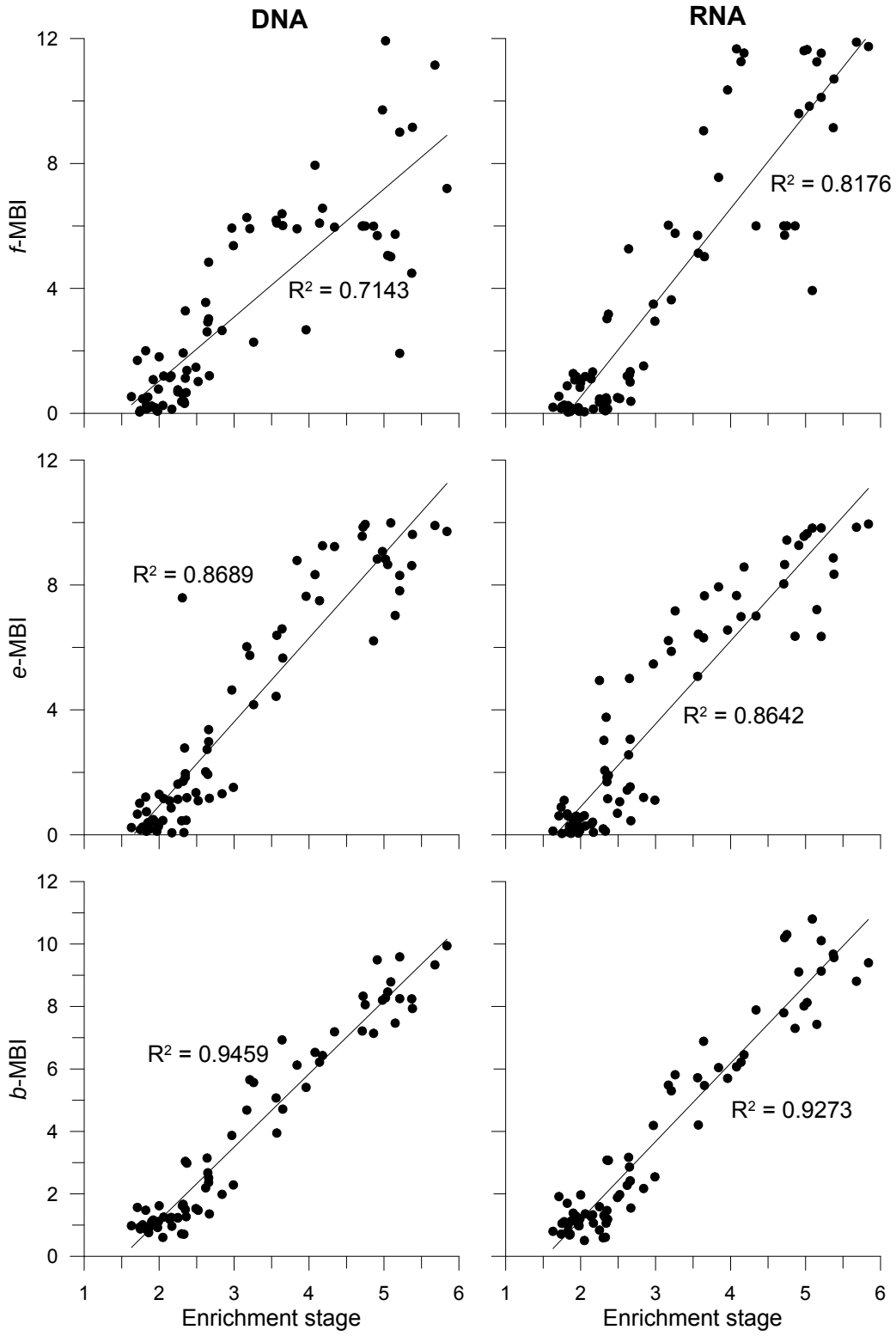
.

Appendix 4C. Relative sequence reads abundance of the 25 most abundant eukaryotic (Euk) Operational Taxonomic Units (OTUs) plotted against the Enrichment Stage (ES), using (A) environmental DNA and (B) eRNA data. Black circles = Marlborough Sounds (MS) Year 1, Red circles = MS Year 2, Blue circles = Steward Island Year 2. Quantile regression splines (Tau = 0.95, df = 3, Deg = 3) fitted for each dataset with vertical lines indicating position of peak read abundance on the enrichment scale.



.

Appendix 5. Linear regressions based on 4th root transformed data of the Eco-Group assignment dataset: Marlborough Sounds (MS) Year 1, MS Year 2 and Stewart Island (SI). Regression slopes between individual biotic indices (foraminifera, f-MBI; bacteria, b-MBI; eukaryote, e-MBI) and Enrichment Stage index are shown for both e environmental DNA and environmental RNA data.



Appendix 6. Linear regressions based on 4th root transformed data of validation dataset: Marlborough Sounds Year 3. Regression slopes between individual biotic indices (foraminifera, f-MBI; bacteria, b-MBI; eukaryote, e-MBI) and Enrichment Stage index are shown for both environmental DNA and environmental RNA data.