Multimarker Metabarcoding of Prokaryotic and Eukaryotic Communities in Contaminated UK Estuaries

Abdullah Mohammad Aleidan

A thesis submitted for the degree of

Doctor of Philosophy

The School of Environmental Sciences

University of East Anglia, United Kingdom

February 2025

Abstract

Anthropogenic metal contamination, particularly from historic mining, has led to persistent copper enrichment in estuarine sediments of southwestern England. Although this contamination is well documented, its ecological consequences remain less understood, especially in microbial and meiofaunal communities, which can offer early indicators of pollution. This thesis aimed to determine (i) whether a multi-marker metabarcoding strategy (16S, ITS, 18S, 28S, COI) can better detect pollution-induced changes compared with single-marker approaches (ii) which taxa respond most consistently to elevated metal levels and (iii) whether porewater copper more accurately predicts benthic community shifts than sediment copper concentrations. Field sampling encompassed 12- and 34-site datasets across major estuaries, covering porewater Cu concentrations from 4.6 to over 400 µg/L. Sediment DNA was extracted in triplicate at each site. Amplicon sequencing of multiple markers profiled bacterial, archaeal and eukaryotic assemblages. Statistical analyses linked community data to copper levels (porewater and sediment), revealing thresholds for compositional turnover. Bacterial and archaeal communities underwent threshold-like shifts near 20 µg/L porewater Cu, with archaea showed greater sensitivity in more uniform site subsets. Eukaryotic assemblages, particularly nematodes and alveolates, exhibited marked changes at heavily contaminated sites, corroborating previous morphological observations. Multi-primer metabarcoding captured a broader range of taxa than single markers alone but remained limited by primer biases and incomplete reference databases. These results refine pollution threshold estimates and highlight porewater copper as a strong predictor of community disruption. Archaea emerged as potential bioindicators, responding more sharply to contamination than bacteria. The multi-marker approach significantly improves ecological resolution, underscoring the importance of integrating morphological and molecular data. Moving forward, enhancing primer sets, expanding databases and applying functional assays (e.g. metatranscriptomics) will further strengthen the use of metabarcoding in environmental management and remediation efforts.

ACKNOWLEDGEMENTS

Words cannot fully capture my profound appreciation for Professor Alastair Grant, my principal supervisor. His exceptional guidance and unwavering support have been indispensable throughout this journey. I am immensely grateful for his remarkable technical insight, which not only shaped the meticulous processing of amplicon data and laid the groundwork for the bioinformatics analysis, but also guided me through each crucial step in the laboratory. Under his mentorship, I have grown into the researcher I am today, and I owe him a debt of gratitude beyond measure. I would also like to acknowledge my additional supervisory team; Dr Trevor Tolhurst and Professor Alexei Maklakov for their invaluable insights and steadfast encouragement. I extend my sincere gratitude to Dr. Vera Fonseca and Professor Thomas Mock for their invaluable feedback and constructive suggestions, which have greatly enhanced the quality of this thesis.

To my mother: your constant presence and unconditional devotion have sustained me over the years. Only now, as I experience parenthood myself, do I fully begin to recognise the sacrifices you and my father made on my behalf. To my father, taken too soon: your memory guides and inspires me daily. I endeavour to be the loving, dedicated parent you were, and I am certain you would take pride in my achievements.

I am profoundly grateful to my hero, my beloved wife, Norah. Your unwavering support, patient understanding during countless late-night analyses, tireless encouragement, wise advice when I had doubts, quiet sacrifices at home, and constant optimism helped me through every challenge and kept me going; words cannot truly say how much I cherish you. To my children, Shahad, Abdulaziz, and Mohammad, watching you grow, explore new horizons, and embrace new languages here in the UK has been a continual source of motivation. Your keen curiosity, from lab tales to how microbes respond to metals, bolstered my resolve and offered invaluable support at every stage, even when you appointed yourselves junior proof-readers on the hunt for my typos. You have reminded me that life extends far beyond the laboratory. To my youngest child, Badr, your arrival during my final year of study not only brought immense joy and excitement to my analysis and writing-up stage, but also encouraged me to see life's moments from a fresh perspective. I love you all more than words can ever convey.

My heartfelt thanks extend to my uncle, Abdullah al Salman, for his ongoing personal support, and to my sisters and their wonderful families for their steadfast encouragement. A special mention goes to Salem, who has been more than a brother. I am also deeply appreciative of my friends and colleagues, Dr Hussein Aldosari, Dr Hamdan Alzahrani, Dr Solomon Udochi for his invaluable advice, and Dr Mohammed Albeshr, Salih al Farraj, Mansour al Mansour, and Abdul Wahid AlRifai for their friendship and support during my time at the university.

I thank the technical team in ENV for their wonderful assistance. Amongst others, Mr Simon Ellis provided great help during multiple trips to Cornwall and Great Yarmouth. Mr Rob Utting and Mr Nicholas Garrard supplied the essential equipment and consumables for my laboratory work, while Mr Graham Chilvers assisted with ICP-MS analyses. My gratitude also goes to Dr Xuewei Mao and Dr Charlotte Davies for their laboratory training.

Lastly, I wish to express my appreciation to the Government of Saudi Arabia and King Saud University for granting me the scholarship that made my studies at the University of East Anglia possible. Without their generous support, the achievements celebrated here would not have come to fruition.

List of Contents

List of tables

List of Figures

Abbreviations and glossary of terms

AEM	Acid-Extractable Metals.
ANOSIM	Analysis of Similarities.
ANOVA	Analysis of Variance.
ASV	Amplicon Sequence Variant.
AVS	Acid Volatile Sulphides.
BIOENV	Biota-Environment correlation analysis.
COI	Cytochrome c oxidase subunit I.
D50	Median grain size.
DESS	Dimethyl Sulphoxide–EDTA–Salt solution.
DNA	Deoxyribonucleic Acid.
DOC	Dissolved Organic Carbon.
DMSO	Dimethyl Sulphoxide.
EDTA	Ethylenediaminetetraacetic Acid.
eqpCu	Equilibrium Partitioning Copper.
GTDB	Genome Taxonomy Database.
HCl	Hydrochloric Acid.
HNO3	Nitric Acid.
ICP - MS - QQQ	Triple Quadrupole Inductively Coupled Plasma Mass Spectrometry.
ITS	Internal Transcribed Spacer.
LCA	Lowest Common Ancestor.
LotuS2	less OTU scripts bioinformatics pipeline.
LT63	Particle size fraction <63 µm.
MiSeq	Illumina sequencing platform.
MIDORI	Reference database for COI sequences.
NMDS	Non-metric Multidimensional Scaling.
NovaSeq	Illumina sequencing system.
OTU	Operational Taxonomic Unit.
PCA	Principal Components Analysis.

PCoA	Principal Coordinates Analysis.
PCR	Polymerase Chain Reaction.
PW	Porewater.
PW_OC_Cu	Dissolved Organic Carbon normalised PWCu.
R	Programming language.
RNA	Ribonucleic Acid.
SEM	Simultaneously Extracted Metals.
SIMPER	Similarity Percentage.
SILVA	rRNA gene sequence database.
SYBR	Fluorescent dye for nucleic acid staining.
TOC	Total Organic Carbon.
Tris	Tris(hydroxymethyl)aminomethane.
UNOISE	Denoising algorithm.
UPARSE	Sequence clustering algorithm.
UPW	Ultrapure Water.
EPS	Extremely poorly sorted.
MS	Moderately sorted.
PS	Poorly sorted.
VPS	Very poorly sorted.
VWS	Very well sorted.
WS	Well sorted.
pH	Measure of acidity or alkalinity.

Chapter 1:

General Introduction

1.1 Overview of Marine Pollution in Coastal Ecosystems

Anthropogenic inputs in marine environments include a suite of pollutants such as metals, hydrocarbons and organic enrichments. Many derive from point sources, such as wastewater discharges or industrial plants, while others arrive through diffuse pathways, including agricultural runoff and atmospheric deposition (Kennish, 2002). Although some contaminants break down over time, metals persist indefinitely because they cannot be biologically or chemically degraded to benign forms (Alloway, 2012). Metals like copper, zinc and lead thus accumulate in sediments, where they can remain bound to particles or become remobilised into the water column depending on redox conditions, grain size and biological activity (Eggleton and Thomas, 2004). Mining has historically been among the most significant contributors to metal contamination, generating tailings that are sometimes deposited directly into rivers or seep slowly from old spoil heaps. Where tidal influence is strong, estuaries can trap these contaminants in lower-energy areas, creating hotspots of long-term enrichment (Bryan and Gibbs, 1983). Although modern legislation and improved waste management have reduced direct inputs in many industrialised regions, past mining often has left a chemical legacy that continues to shape sediment composition and benthic ecology long after mining operations ceased (Johnson, 1986, Rollinson et al., 2007).

1.2 The Mining Legacy in Southwestern England

Southwestern England epitomises this phenomenon. Cornish mining, at its peak between the eighteenth and twentieth centuries, generated large volumes of copper, tin and arsenic rich wastes that flowed into watercourses (Barton, 1961). Estuaries like Restronguet Creek and the Hayle became known for extreme copper and other metal concentrations in their sediments (Bryan and Hummerstone, 1971, Bryan and Gibbs, 1983). Although active mining has largely ended, the sediments have not reverted to baseline conditions. This stability offers a relatively

simple system in the sense that there are fewer industries now depositing metals and contamination is dominated by a small number of metals, principally copper, zinc and arsenic (Bryan and Gibbs, 1983, Grant, 2010). Nonetheless, the distribution of metals can vary markedly across short distances, creating gradations that allow investigators to assess how ecological communities shift along these gradients (Millward and Grant, 2000, Ogilvie and Grant, 2008). For instance, the Hayle may exhibit lower total sediment copper concentrations than Restronguet Creek but higher porewater copper, which can produce greater biological effects if metals in porewater are more bioavailable to benthic species (Millward and Grant, 1995, Chen et al., 2022). The apparently stable and well-defined nature of these pollution gradients means that southwestern estuaries have served as a longstanding model for studying metal ecotoxicology (Bryan and Hummerstone, 1971, Somerfield et al., 1994a, Rainbow, 2020).

1.3 Defining Contamination and Pollution under GESAMP

One of the conceptual foundations that guide these studies is the distinction between contamination and pollution made by the United Nations Joint Group of Experts on the Scientific Aspects of Marine Pollution (GESAMP) (www.gesamp.org). According to GESAMP, contamination refers to the presence of a substance above natural background levels, while pollution implies that such contamination impairs organisms, communities or ecosystem processes. This difference is of practical importance because demonstrating that concentrations of copper or other metals are elevated above their preindustrial background does not, on its own, confirm that communities experience harmful effects. Many southwestern England estuaries are demonstrably contaminated, but they are only classified as polluted if there is concrete evidence that biological structure or function is altered (Bryan and Hummerstone, 1971, Grant, 2010). The GESAMP framework thus compels researchers to look beyond chemical measurements and address whether metal exposure leads to shifts in community composition, declines in species richness or changes in key functions such as feeding rates. Such data are crucial to determining whether management or remediation is necessary and to clarifying whether metals rather than other environmental parameters are responsible for observed biological changes (Chapman and Wang, 2001).

1.4 Impacts of Pollutants on Marine Communities

A substantial literature has accumulated on the effects of pollutants in marine systems, often highlighting both direct toxicity and more subtle alterations in community assembly. Some studies focused on macrofauna, examining how species diversity and abundance changed with rising metal levels or increasing hydrocarbon inputs (Olsgard and Gray, 1995). Others documented community rearrangements, such as the dominance of opportunistic species or the loss of sensitive taxa in severely impacted sites (Pearson, 1978). The Ekofisk oilfield in the North Sea is frequently cited as an example where discharges associated with drilling significantly reduced benthic biodiversity at the most contaminated sites and coased more subtle changes in community composition at less contaminated sites (Gray et al., 1990, Warwick and Clarke, 1991, Olsgard and Gray, 1995). In southwestern England, changes in benthic diversity or the disappearance of certain bivalves and crustacea have been attributed to mining-derived metals, though confounding variables occasionally limit straightforward interpretations (Bryan and Hummerstone, 1971, Warwick et al., 1991, Warwick and Clarke, 1991). Some authors reported that moderate contamination could bolster a few tolerant taxa, leading to unexpected increases in total abundance if tolerant species flourished without competition (Blanck, 2002). Therefore, the notion that pollution always causes uniform declines in diversity is not always borne out by empirical data and subtle community shifts can be as ecologically significant as outright losses of species (Dauvin, 2007).

1.5 Environmental Complexity in Southwestern Estuaries

Investigators studying southwestern England estuaries have faced additional complications associated with the interplay of metals, grain size and salinity gradients. Sediment copper is often correlated with the proportion of fine particles, since finer grains have a larger surface area to which metals can adsorb (Eggleton and Thomas, 2004). Depositional zones with extensive mud can accumulate more copper, yet low oxygen penetration into the sediments and high organic matter may mitigate or magnify the metal's bioavailability depending on redox states (Kennish, 2002). The Hayle's high porewater copper but lower total sediment copper exemplifies how geochemical processes can elevate the fraction of metal accessible to benthic infauna despite smaller total loads (Millward and Grant, 2000). Meanwhile, salinity itself can vary along the estuarine gradient, further influencing metal speciation and the tolerance

thresholds of local species (Millward and Grant, 1995). Correlating metal concentrations to biological effects thus requires site-specific knowledge of sediment characteristics, porewater conditions and the feeding or burrowing habits of local taxa (Grant et al., 1989). Several authors have recommended combining multiple lines of evidence, such as total metal measurements, porewater analyses and advanced models like AVS or equilibrium partitioning, to capture the intricate nature of metal binding and release (Di Toro et al., 1991). However, implementing these models can be resource-intensive and each site may present distinct geochemical nuances that complicate straightforward interpretation.

1.6 Evidence of Metal-Induced Ecological Change

Southwestern estuaries have revealed multiple lines of evidence that metals cause ecological changes. Bryan (1971) and Bryan and Gibbs (1983) noted that certain species, such as the amphipod *Corophium volutator* and the bivalve *Scrobicularia plana*, were scarce or missing in Restronguet Creek, presumably due to high copper toxicity. They included a wide range of taxa, from crustaceans and molluscs to fish, annelids and algae, demonstrating how metal-enriched conditions reduced the abundance of sensitive organisms. Warwick (2001) directly compared the Fal's intertidal macrobenthos with other southwestern estuaries and noted that the Fal's macrofauna were conspicuously missing metal-sensitive crustaceans such as *Corophium volutator* and *Cyathura carinata* while exhibiting elevated abundances of opportunistic annelids. These patterns contrasted sharply with similarly muddy habitats elsewhere in the region, suggesting that long-term metal contamination overshadowed confounding variables such as salinity or grain size. Similarly, Bryan et al. (1987) emphasised that the Fal's elevated copper, zinc and organotin concentrations which was originated from historical mining, had governed local benthic distributions over extended periods, further highlighting the dominant influence of metals on species composition in this estuary. Meiofauna, especially nematodes, exhibited clear community shifts at sites such as Restronguet Creek and the Hayle, where copper levels were elevated (Somerfield et al., 1994a, Millward and Grant, 1995). Millward and Grant (1995) showed that nematode copper tolerance was significantly higher in Restronguet, Pill and St Just's Creeks compared with reference sites, paralleling morphological evidence from Warwick (2001) macrofaunal studies in these same creeks. Microbial communities likewise displayed enhanced copper tolerance in Restronguet, the Hayle and Pill once porewater concentrations exceeded around 15 µg/L (Ogilvie and Grant,

2008). Meanwhile, the polychaete *Nereis diversicolor* demonstrated inherited tolerance confined to Restronguet Creek and the Hayle, but this tolerance weakened in the lower parts of Restronguet Creek (Grant et al., 1989). Collectively, these findings underscore the widespread ecological repercussions of copper contamination across multiple benthic groups. Furthermore, Emily Shipp's feeding rate study highlighted how deposit feeders like *Hydrobia* can exhibit lowered feeding and growth in certain creeks with high copper, although the effect was also influenced by sediment organic content (Shipp and Grant, 2006). Pill and St Just's creeks, for instance, displayed both reduced feeding and growth, suggesting the presence of genuine pollution effects rather than mere contamination. Percuil, by contrast, had reduced feeding rates but no significant impact on growth and the explanation for that difference aligned with the sediment's high organic matter content, illustrating once more the complexity of linking measured metals to ecological outcomes. Overall, these studies provided clear evidence that copper contamination has repeatedly caused shifts in macrofaunal, meiofaunal and microbial assemblages throughout southwestern estuaries, highlighting metals as a primary driver of ecological change in these systems.

1.7 Measures of Metal Contamination

The question of which sites are most affected by metals and which measures best predict impacts has not been fully resolved. Multiple approaches have been employed to characterise metal contamination in southwestern estuaries, focusing chiefly on total sediment copper and porewater copper. Restronguet Creek often appears more heavily contaminated by total copper, whereas the Hayle can appear less contaminated by total sediment metrics yet exhibits higher porewater copper that may harm deposit feeders (Grant, 2010, Udochi, 2020). For instance, at Restronguet Creek, total copper in sediments can approach 3,000 µg/g but yield a porewater value of about 130 µg/L, whereas the Hayle, sediments containing 862 µg/g total copper can record a far higher porewater copper concentration of over 220 µg/L (Udochi, 2020). These contrasts emphasised that porewater concentrations may reflect bioavailable fractions more accurately than total copper analyses alone (Millward and Grant, 2000, Ogilvie and Grant, 2008). Others found that sediment analyses remain relevant, particularly for deposit feeders that consume sediment particles (Bryan and Langston, 1992). More complicated models, such as those incorporating AVS or equilibrium partitioning, added nuance but demand robust geochemical data (Hall Jr and Anderson, 2022). Certain estuaries may also have site-specific

anomalies: the Helford River is often used as a control system because it lacks the same intensity of historical mining activity, yet some areas there also show unexpected disturbance. In particular, Udochi (2020) analysis of Helford River nematode communities noted that they clustered close to a reference site (Breydon Water) in multivariate space, but identified a total porewater Cu threshold of about 3.5 µg/L and a sediment Cu threshold of about 215 µg/g for shifts in community composition, indicating that even moderate copper levels can elicit ecological impacts. This variability underscores the importance of having multiple reference or control sites when trying to infer pollution-induced patterns of change.

1.8 Traditional Monitoring and Its Limitations

Traditional morphological monitoring has relied heavily on macrofauna retained on a 0.5 mm sieve (Kendall and Widdicombe, 1999). Macrofauna provide tangible endpoints and have been studied over many decades, but these methods can miss smaller taxa like meiobenthos or microbial eukaryotes that may respond differently or more rapidly to pollution (Moore and Bett, 1989, Coull, 1992). Nematodes, for instance, have often proven more sensitive than larger invertebrates, though morphological identification of nematodes demands exceptional expertise and extensive effort (Warwick and Uncles, 1980, Somerfield et al., 1994c). Bacteria and archaea, as revealed by classic T-RFLP or culturing work, sometimes display altered tolerance in polluted estuaries, but compositional changes are less apparent if the fingerprinting methods lack taxonomic resolution (Ogilvie and Grant, 2008). Over the past decade, attention has turned to DNA-based metabarcoding, which can capture a broader spectrum of organisms in a single analysis (Taberlet et al., 2012, Pawlowski et al., 2018). This approach holds substantial promise for detecting subtle shifts in community composition that may signify contamination moving toward pollution. Despite this promise, the application of metabarcoding to estuarine metal pollution lags behind its usage in other areas, such as invasive species detection and general biodiversity monitoring (Bik et al., 2012). Although preliminary efforts to correlate metabarcoding outputs with metal exposure are now emerging (Corcoll et al., 2019, Yan et al., 2020), large-scale investigations along well-defined pollution gradients remain rare.

1.9 Primer Selection and Coverage Biases

One key methodological consideration is the choice of primers and target genes. Bacterial surveys traditionally use the 16S rRNA gene, which has both conserved and variable regions (Cole et al., 2005). However, the commonly used 97% similarity threshold for species delimitation is sometimes criticised for lumping distinct taxa, prompting suggestions of more stringent cutoffs (Edgar, 2018). Archaea pose another challenge because neither "universal" prokaryote nor many Archaea specific primers successfully capture the full range of archaeal phylogenetic diversity(Grant et al., 2023). Bahram et al. (2019) designed primers that amplified archaeal lineages missed by older sets and Karst et al. (2018) identified highly divergent archaea that might represent entirely new phyla. Eukaryotes lack a single universally accepted target for use in metabarcoding. The 18S rRNA gene is widely adopted, but partial fragments may not distinguish closely related species (Chittavichai et al., 2025). Gaonkar and Campbell (2024) noted that full-length 18S is needed for many groups, yet longer fragments can be less practical for high-throughput sequencing. Some eukaryotic surveys relied on the more variable ITS or COI sequences, but these markers also present biases or incomplete coverage in public databases (Bik et al., 2012, Zhang et al., 2014). These issues become especially relevant in pollution studies where researchers hope to detect subtle but ecologically relevant changes. If the primers or reference data fail to identify key taxa, important pollution responses could remain undetected.

1.10 Influence of Sample Processing on Metabarcoding

In addition to primer design, sample processing can influence metabarcoding outcomes. Whole-sediment extractions may yield broad community profiles but risk underrepresenting certain fragile taxa if the extraction protocol was harsh (Deiner et al., 2015, Hermans et al., 2018). Sieving or elutriation can isolate meiofauna but could lose microbial eukaryotes or archaea. DNA from dead cells or extracellular sources might inflate perceived diversity, though its importance depends on local sediment turnover rates (Taberlet et al., 2012). Despite these caveats, metabarcoding offers a remarkable opportunity to move beyond single-taxon or morphological approaches, potentially capturing a wide range of prokaryotic and eukaryotic diversity in one workflow (Pawlowski et al., 2018). By combining these methods with morphological data on macrofauna or nematodes, it might be possible to identify how

consistently contamination triggers shifts across different size classes and trophic strategies (Warwick and Clarke, 1991, Somerfield et al., 1994a). This thesis would harness multi-target metabarcoding, including 16S, ITS, 18S, 28S and COI, to examine the distribution of bacteria, archaea, fungi, metazoans and protists in southwestern estuaries with well-characterised metal gradients.

1.11 Key Research Questions

The thesis aimed to address several unresolved questions. First, it would examine which metrics of metal contamination correlated best with changes in community structure, particularly asking whether porewater copper stood out as a more accurate predictor than sediment copper concentrations. Second, it would explore which taxonomic groups exhibited the strongest and most consistent shifts under elevated metal concentrations. Some earlier evidence suggested that nematodes and microbes may be extremely sensitive, but the role of eukaryotic microbes or cryptic archaeal lineages was unclear (Corcoll et al., 2019, Yan et al., 2020, Zhang et al., 2024a). Third, it would investigate how the choice of primer sets influenced the detection of pollution-induced shifts in community composition, assessing whether newly designed primers genuinely improved coverage or risk amplifying certain clades at the expense of others. These methodological concerns are critical to developing robust protocols for future monitoring programs. Fourth, the research would identify whether the patterns in southwestern estuaries aligned with or diverged from those reported in more moderately contaminated or less complex systems, thereby contributing to broader ecological understanding.

1.12 Thesis Structure and Methodological Outline

The material in **chapters 2** and **3** of this thesis elaborated on sampling, DNA extraction and primer selection, including practical considerations such as single-step PCR with barcoded primers to reduce contamination risk. Chapter 4 examined bacterial communities and their response to metals, building on earlier T-RFLP findings that showed limited compositional change despite strong tolerance gradients (Ogilvie and Grant, 2008). **Chapter 5** addressed archaea, asking if these frequently overlooked prokaryotes exhibited distinct or more resilient responses to metals. **Chapter 6** focused on eukaryotic assemblages, employing multi-primer

metabarcoding to assess whether some lineages consistently tracked metal gradients. This final part of the thesis evaluated whether multi-target metabarcoding genuinely exceeded single-marker or morphological methods in detecting early ecological changes associated with contamination. Additionally, factors such as salinity, sediment texture and organic matter, already explored in **chapters 4** and **5** for bacterial and archaeal communities, were further examined to determine how they might obscure or amplify these pollution effects in eukaryotic assemblages.

1.13 Advancing the Field Through Multi-Target Approaches

This work thus aimed to exceed the scope of previous studies by combining a wide set of molecular markers, a large sample size from different sites and established morphological and biochemical metrics of pollution impact. The intention was to determine which combinations of metrics and methods best captured the nuanced ways in which metals disturbed benthic ecosystems. By doing so, the thesis provided new insights into the ecological consequences of historic mining in southwestern England and offered methodological refinements for researchers and environmental managers seeking to detect metal pollution elsewhere. A core argument was that metals often created subtle shifts in community composition that morphological surveys or single-marker approaches might have only partially revealed. Demonstrating this convincingly required careful integration of geochemical, morphological and molecular data, consistent with the GESAMP view that, contamination only constituted pollution if measurable harm was demonstrated. Subsequent chapters detailed how each component of this approach had been implemented, beginning with the sampling design and culminating in a comparative analysis across multiple molecular markers.

1.14 Morphological and Molecular Approaches for Community Assessment

1.14.1 Integrative Identification Methods

Throughout history, most taxonomic studies of microorganisms used conventional morphological identification, which described an organism's visible form and structure. This approach has underpinned numerous investigations of diatom (Moseley and Manoylov, 2012), fungal (Shearer and Lane, 1983), bacterial (Hollaway et al., 1980) and nematode (Somerfield

et al., 1994c) community composition. However, morphological analysis could be labour-intensive and depended heavily on taxonomic expertise (Hajibabaei et al., 2011). Tytgat et al. (2019) compared morphological and metabarcoding approaches for nematodes sampled from diverse land uses, concluding that DNA-based methods sometimes underestimated species numbers because certain morphologically distinct species shared identical 18S and COI sequences. On the other hand, morphological identification could struggle with cryptic or damaged specimens, again highlighting the need for complementary molecular tools. Comparative studies indicated that, despite their distinct biases, molecular and morphological surveys often revealed consistent ecological patterns. Kim et al. (2025) showed that both approaches tracked similar plankton community trends, with metabarcoding offering finer resolution for some calanoid copepods. Cahill et al. (2018) reported a close correspondence between morphological counts and metabarcoding profiles, which strengthened further when taxonomic resolution was harmonised. In benthic sediments, den Bulcke Laure et al. (2024) found that bulk-DNA metabarcoding mirrored morphological patterns of alpha- and beta-diversity following sand-extraction disturbance. A systematic review by Keck et al. (2022) likewise concluded that, while each method captured certain taxa the other missed, they generally converged on community-level trends. Together, these findings supported an integrative strategy in which metabarcoding augmented, rather than replaced, classical morphology in biodiversity assessments.

High throughput sequencing (HTS), also known as next-generation or massively parallel sequencing, became pivotal in microbial ecology (Liu et al., 2012). It arose after decades of capillary (Sanger et al., 1977) and gel-based sequencing, offering unparalleled throughput and enabling the simultaneous decoding of up to one billion DNA molecules (Schuster, 2008). High-throughput sequencing (HTS)is especially useful for uncovering the diversity of smaller or hard-to-culture organisms such as viruses, bacteria or ancient samples (Krych et al., 2019). The capacity to generate large datasets has fostered a wide range of amplicon-based studies in humans (Yap et al., 2016), animals (Rausch et al., 2016) and soils (El Khawand et al., 2016). Despite its power, HTS can introduce biases in PCR amplification and may under-detect certain taxa if they have mismatched primer sites or lower template abundance (Berry et al., 2011, O'Donnell et al., 2016).

Metabarcoding is a PCR-based approach that used broadly conserved "universal" primers to amplify short, high-copy barcode loci from environmental DNA (Emilson et al., 2017), producing community profiles that are best interpreted as relative rather than absolute abundance estimates (Pompanon et al., 2012**).** The method now underpins biodiversity assessments across diverse habitats, ranging from benthic meiofaunal communities (Leary et al., 2014, Pochon et al., 2015, Zaiko et al., 2016) to bacterial (Xue et al., 2014, Lawes et al., 2016) and eukaryotic (Leary et al., 2014, Pochon et al., 2015, Zaiko et al., 2016) communities, including biofouling assemblages (von Ammon et al., 2018). Quantitative accuracy was hindered by primer mismatches, differential amplification efficiencies and variations of several orders of magnitude in rRNA or mitochondrial gene-copy number, especially among multicellular taxa (Gotelli and Chao, 2013, Aylagas et al., 2014, Bucklin et al., 2016). Polymerase choice and amplicon GC content can skew relative read counts, and ≥ 10 independent PCR replicates are often needed to detect rare taxa reliably (Fonseca, 2018, Nichols et al., 2018). Empirical studies showed that read counts frequently correlated weakly with specimen biomass unless taxon-specific correction factors were applied (Elbrecht and Leese, 2015, Lamb et al., 2019). Even when internal DNA standards were introduced, PCR bias means that read numbers cannot be converted into absolute organismal abundances; at best they provide a semi-quantitative indication of between-sample changes (Luo et al., 2023). Given these mixed findings, this thesis treated metabarcoding reads as semi-quantitative and emphasised conservatively interpreted relative-abundance patterns rather than absolute counts (Krehenwinkel et al., 2017, Deagle et al., 2019). Deploying multiple loci or targeting distinct regions within a gene helped average out locus-specific biases and broadened taxonomic coverage (Pompanon et al., 2012).

Metagenomics, or shotgun sequencing of total environmental DNA without PCR, eliminated primer bias by sequencing all fragment present (Eloe-Fadrosh et al., 2016). This untargeted strategy detected a wider spectrum of life, including uncultivable bacteria, archaea and viruses, while simultaneously recovering genes that reveal metabolic pathways, biogeochemical cycling and antibiotic-resistance determinants (Quince et al., 2017). At sufficient depth it can also yield near-complete mitochondrial genomes, thereby enriching reference libraries and improving downstream identification (Elbrecht et al., 2017). Read depth, however, scaled with genome size while cell-wall architecture affected DNA extraction efficiency, so taxa with small genomes or fragile cells tended to be over-represented whereas large-genome or recalcitrant

organisms were under-sampled (Beszteri et al., 2010, Shi et al., 2011)**.** Because barcode loci form only a small proportion of shotgun libraries and curated references remained sparse, taxonomic annotation was often poorer than in targeted metabarcoding (Elbrecht et al., 2017, Cribdon et al., 2020), despite the deeper sequencing and more complex bioinformatics required (Riesenfeld et al., 2004, Temperton et al., 2009). When coverage is adequate, metagenomics excelled at linking community composition to ecological function, but for routine biodiversity audits its cost and computational overheads meant that it continued to complement rather than replace primer-based approaches (Carew et al., 2018)**.**

Metatranscriptomics involves sequencing community RNA. It is most commonly used to quantify functional gene expression after removing ribosomal RNA. However, if RNASeq is carried out on total RNA without ribosomal depletion, information could be obtained on community composition without the need for primer amplification, as approximately 90% of RNA is ribosomal, split approximately equally between SSU and LSU RNA(Urich et al., 2008, Helbling et al., 2012, Turner et al., 2013). Ribosomal read depth scales approximately with biomass and metabolic activity (Bailly et al., 2007, Semmouri et al., 2020), partially mitigating the genome-size bias of metagenomics. This approach can reduce PCR-based biases and capture prokaryotic diversity accurately, yet its coverage of eukaryotic taxa often remained shallow or restricted to higher taxonomic levels, in part because the sequenced fragments were distributed across the whole length of rRNA molecules, rather than being focussed on hypervariable regions targeted by metabarcoding (Turner et al., 2013). It also remains costlier and more technically demanding than metabarcoding because it requires extra RNA extraction and rRNA-depletion steps (Shakya et al., 2019, Wilson et al., 2019).

Taken together, these molecular methods offered complementary insights that can exceed those of purely morphological approaches, particularly for low-abundance or cryptic taxa. Yet each technique introduced distinct biases and practical constraints, meaning that choosing an optimal strategy often involves balancing depth of coverage, costs and the level of taxonomic resolution required.

1.14.2 Genetic Markers for Metabarcoding

Multiple genetic loci were commonly used in metabarcoding to capture diverse taxonomic groups, each offering distinct advantages (Taberlet et al., 2012, Ficetola et al., 2021, Wang et al., 2023b):

- **16S rRNA**: Primarily for bacteria and archaea. Hypervariable regions enable broad taxonomic resolution and many reference databases exist (Wade, 2002, Webster et al., 2003, Cole et al., 2005).
- **Internal Transcribed Spacer (ITS)**: Highly variable spacer region, especially informative for fungi and some nematodes (Hugall et al., 1999, Elbadri et al., 2002). Reference databases remain incomplete for certain groups, yet ITS can achieve species-level resolution when data are available (Félix et al., 2014).
- **18S rRNA**: Targets eukaryotes broadly, from protists to nematodes. Although 18S can struggle to resolve closely related species, it remains standard for overall eukaryotic community analysis (Lallias et al., 2015, Latz et al., 2022, Zimmermann et al., 2024).
- **28S rRNA**: Offers finer discrimination among certain eukaryotes compared with 18S, but fewer reference sequences exist (Pereira et al., 2010, Vogt et al., 2014).
- **Cytochrome oxidase subunit I (COI)**: Often referred to as the "barcode gene" for animals (Leray and Knowlton, 2015). It can differentiate congeneric species but lacks truly universal primers, complicating large-scale invertebrate surveys (Derycke et al., 2010, Macheriotou et al., 2019).

By combining these markers, studies can capture both prokaryotic and eukaryotic diversity and cross-check ambiguous assignments. The decision to emphasise particular loci often depends on whether the main interest was in abundant bacteria, fungi or large metazoans (Bik et al., 2012).

1.14.3 Using PICT as a Sensitive Indicator of Metal Pollution

Metal pollution significantly affects marine microbial communities, posing one of the highest threats to benthic ecosystems (Di Cesare et al., 2020). Different microorganisms exhibit varying tolerance levels, which may increase in response to chronic or acute metal exposures (Chen et al., 2019, Fang et al., 2019, Reddy and Dubey, 2019). Pollution-induced community

tolerance (PICT) has thus been widely employed to assess how heavy metals, among other pollutants, reshape microbial assemblages (Blanck et al., 1988, Ogilvie and Grant, 2008). Although several studies have examined links between metal resistance genes (MRGs) and sediment metal concentrations, the findings remain inconsistent. Some have observed marked shifts in microbial community structure at highly contaminated sites (Gough and Stahl, 2011, Zhang et al., 2017, Beale et al., 2018, Chen et al., 2019, Lin et al., 2019), while others report only modest changes, possibly due to the influence of nutrients or other confounding environmental factors (Ogilvie and Grant, 2008, Gołębiewski et al., 2014, Gubelit et al., 2016, Song et al., 2019). Recent work by Di Cesare et al. (2020) found that microbial diversity increased with higher sediment metal loads, yet overall community composition remained largely unchanged, suggesting that metals may act indirectly through nutrient pathways rather than exerting direct toxicity. These contrasting results highlight the importance of site-specific assessments, as PICT may capture adaptive responses in some systems but appear muted where other ecological drivers prevail.

1.14.4 Microcosm Experiment

Microcosm experiments were widely employed in ecological research because they provide controlled environments in which complex interactions and processes can be replicated and studied at manageable scales (Cao et al., 2021). By simulating natural conditions, these systems enable precise tests of hypotheses pertaining to community dynamics, biodiversity and ecosystem functions (Srivastava et al., 2004). They have been used to explore microbial responses to various contaminants including naturally occurring radioactive materials, thus revealing important aspects of microbial resilience and adaptation (Mackay-Roberts et al., 2024). Following the sediment bioassay approach of Austen and Somerfield (1997), this thesis employed microcosms to test whether copper sensitivity observed in situ correlates with shifts in community structure under controlled conditions. Sediments from five sites spanning low to high porewater copper concentrations were defaunated via freeze-thaw and allocated into control and treatment groups. In one treatment, ‘clean’ and ‘contaminated’ sediments were cross-inoculated to assess whether populations from less-polluted sites could adapt to elevated copper loads; in another, contaminated sediments were partly diluted with unpolluted material to gauge the tolerance thresholds of the native fauna. This design facilitates a direct comparison of community responses across a known metal gradient, minimising confounding factors such

as salinity and grain size. A number of studies have demonstrated that microcosm-based studies can uncover early-stage adaptations and shifts in microbial and faunal communities before they become evident in large-scale field surveys (Austen and Somerfield, 1997, Mitchell et al., 2009, Koeppel et al., 2013).

Chapter 2:

Method Development for Assessing Metal Contamination and Microbial Diversity in Estuarine Environments

2.1 Study Design

To investigate metal contamination gradients across Southwest England estuaries, samples were collected from a total of 48 sites during three distinct sampling phases (**Fig. 2.1**) and (**Table 2.1**). The first set, gathered in March 2020, comprised initial samples from two sites in Breydon Water, Norfolk to be used in developing and evaluating methodologies. The second set, sampled in September 2017 by Udochi (2020), included 12 sites: eleven sites covering the entire metal contamination gradient across the Fal and Hayle estuaries in Cornwall, Southwest England, as detailed in (Udochi, 2020) and an uncontaminated reference site at Breydon Water in Norfolk, England (Greenwood, 2001). The third set, collected in March 2022, consisted of 34 sites selected to cover the entire metal contamination gradient across the Southwest England estuaries, as documented in previous studies (Bryan and Hummerstone, 1971, Bryan and Gibbs, 1983, Millward and Grant, 2000, Rollinson et al., 2007, Grant, 2010) and detailed in **Table 2.1**. Uncontaminated estuaries such as the Avon in South Devon, Tamar and the Helford River were sampled as references (Bryan and Hummerstone, 1971, Bryan and Gibbs, 1983). The Percuil River, within the Fal System, was selected as a relatively uncontaminated site due to its environmental conditions being similar to those of the other sites (Bryan and Gibbs, 1983, Millward, 1995).

Three replicate samples were collected from intertidal areas within estuaries from the 12- and 34-site sets, while two replicates were obtained from the initial set. All study sites were located in the mid shore, where they were inundated during high spring or neap tides but exposed during low tides. Replicates were spaced approximately 0.5 m apart to capture small-scale sediment variability without oversampling microscale heterogeneity.

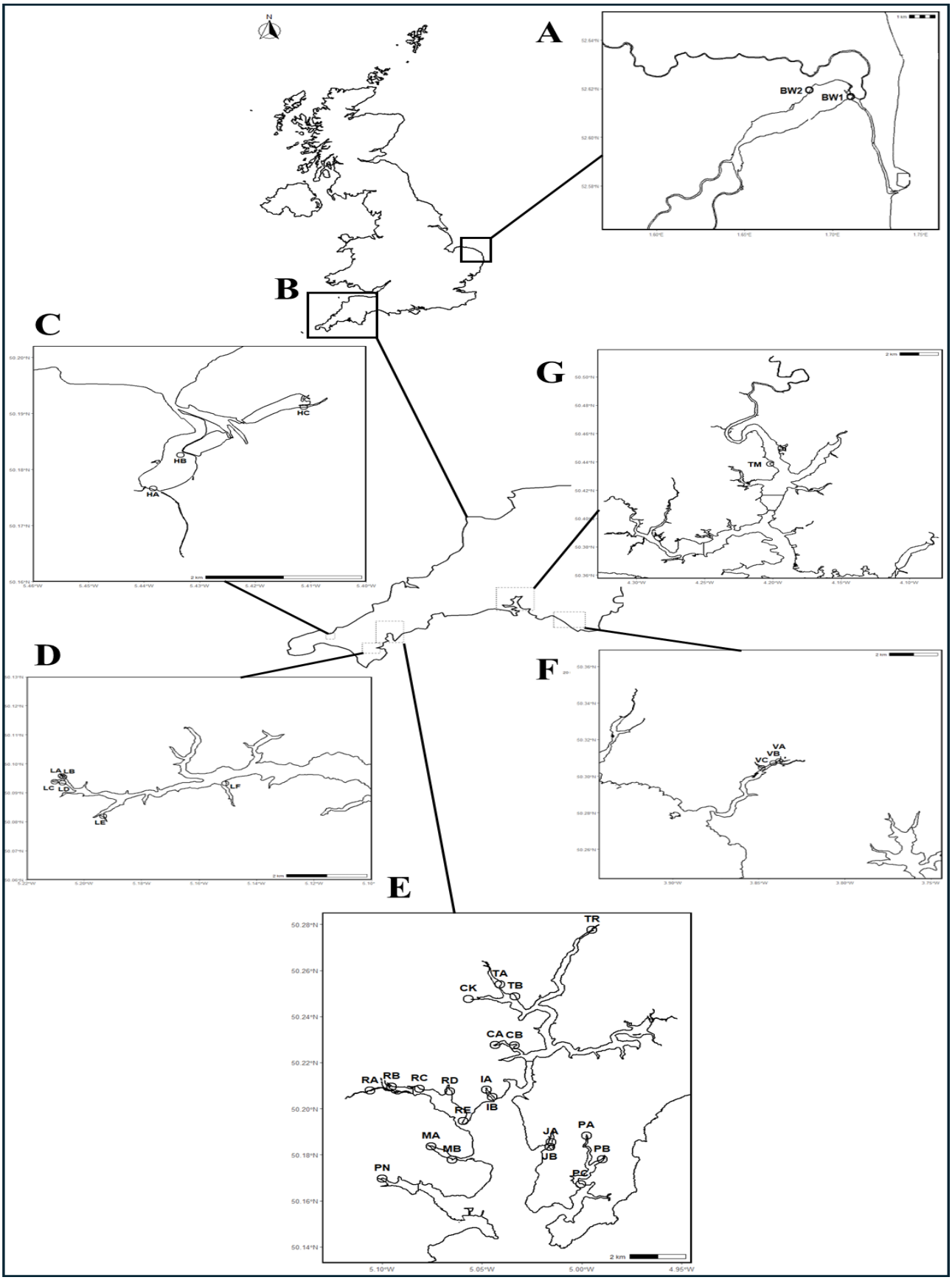


Fig. 2.1 Mid-shore study sites across **A)** Breydon Water, Norfolk and **B)** Southwest England.

Location abbreviations: **A)** Breydon Water (BW1, BW2); **C)** River Hayle Estuary (HA-HC); **D)** Helford River (LA-LF); **E)** Fal Tributaries: Pill Creek (IA, IB), Mylor Creek (MA, MB), Penryn (PN), Tresillian (TR), Percuil (PA-PC), St Just (JA, JB), Truro (TA, TB), Calenick (CK), Cowlands (CA, CB), Restronguet (RA-RE); **F)** Tamar River (TM); **G)** Avon River (VA-VC).

Table 2.1 Sampling sites with their corresponding rivers and British Grid References used in the study. Site acronyms refer to abbreviations used throughout the study, as listed alongside their full river names. Each site code corresponds to a specific location within the indicated river system.

River	Initial set	12-Sites	34-Sites	British Grid Reference
Breydon Water	**BW1**	**BW**		TG516081
Breydon Water	**BW2**			TG500084
River Avon A			**VA**	SX690471
River Avon B			**VB**	SX687470
River Avon C			**VC**	SX683467
River Tamar			**TM**	SX434624
Helford River A			**LA**	SW706266
Helford River B		**HR**	**LB**	SW707266
Helford River C			**LC**	SW704264
Helford River D			**LD**	SW706264
Helford River E			**LE**	SW716250
Helford River F			**LF**	SW747262
Pill Creek A		**PC**	**IA**	SW826386
Pill Creek B			**IB**	SW828383
Mylor Creek A		**MC**	**MA**	SW805360
Mylor Creek B			**MB**	SW812353
Penryn River			**PN**	SW787345
Tresillian River			**TR**	SW867462
Percuil River A		**PR**	**PA**	SW861363
Percuil River B			**PB**	SW866351
Percuil River C			**PC**	SW858340
St Just A			**JA**	SW848360
St Just B		**SJ**	**JB**	SW847358
Truro River A			**TA**	SW833437
Truro River B			**TB**	SW838431
Calenick Creek			**CK**	SW821431
Cowlands A		**CO**	**CA**	SW830408
Cowlands B			**CB**	SW837407
Restronguet A		**RA**	**RA**	SW784388
Restronguet B			**RB**	SW792389
Restronguet C		**RB**	**RC**	SW802388
Restronguet D		**RC**	**RD**	SW813386
Restronguet E			**RE**	SW817372
Hayle A		**HA**	**HA**	SW546363
Hayle B			**HB**	SW549370
Hayle C		**HB**	**HC**	SW566379

2.1.1 Sediment Sampling And Preservation

Mud samples were collected from top 2 cm of the mud flat at low tide and stored in Falcon tubes and frozen later at -80^{o}C to be used for the molecular work (see **Fig. 2.2**). Additionally, ≥1kg muds were sampled and stored in sealable sampling bags. However, due to the logistics of sampling in the SW, time taken between the first sample being collected and laboratory storage was at most 104 hrs. Sediments from Breydon Water were transported to the laboratory within 1 hr of collection. The remaining portion was refrigerated at 4^{o}C in the dark until processed to be used for the porewater extraction, Acid extractable metals (AEM), PH, salinity and grain size measurements.

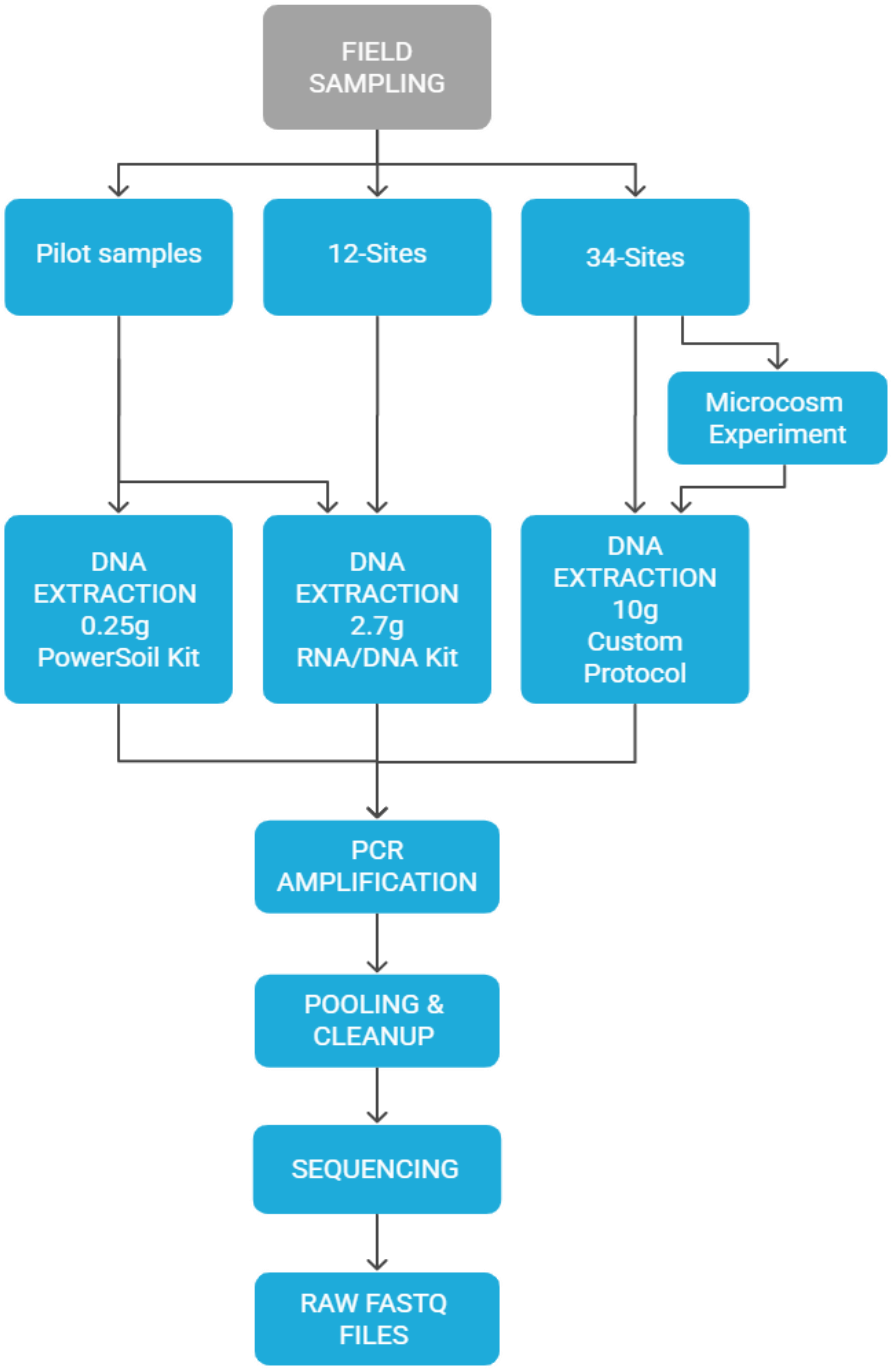


Fig. 2.2 Schematic overview of the sampling and molecular workflow. Sediments from multiple site sets and experiments were processed through DNA extraction, PCR amplification, library preparation, and sequencing to produce raw FASTQ files.

2.1.2 Microcosm Experiment

A microcosm experiment was set up to assess whether species' sensitivity to copper reflected their distribution along the pollution gradient. The experiment protocol followed the approach outlined by Austen and Somerfield (1997), who demonstrated the utility of sediment bioassays in examining microfaunal and meiofaunal responses to metal contamination gradients.

In March 2022, sediment samples were collected from the top 2 cm of mudflats at five estuarine sites among the 34 sites studied, all located in the mid-shore zone. These sites were selected to represent a range of porewater copper (PWCu) concentrations measured in (**Table 2.7 - Section 2.2.2**); Avon River (VC) with 13.37 µg/L, Percuil River (PA) with 6.54 µg/L, Helford River (LB) with 123 µg/L, Mylor Creek (MB) with 363.6 µg/L and Restronguet Creek (RD) with 887.3 µg/L see **Table 2.2** for site characteristics. Defaunated sediments were prepared from each of these five samples. The sediments were thoroughly mixed and homogenized, then subjected to three freeze-thaw cycles at -20 °C to eliminate existing fauna, providing a controlled substrate for the experiment. The experimental design included control samples and two treatments to investigate community responses to varying copper concentrations and interactions between communities from different pollution backgrounds. An overview of the experimental setups is provided in **Table 2.2**.

Table 2.2 A summary of the experimental setups, including the control and treatment groups with their respective sediment sources, defaunated sediments, replication and total number of samples.

Experimental Setup	Description	Sediment Source	Defaunated Sediment	Replicates	Total Samples
Control (Baseline)	Inoculum sediment only	VC, PA, LB, MB, RD	None	1 per site	5
Treatment 1	Cross-inoculation of inoculum sediments into defaunated sediments from all sites	VC, PA, LB, MB, RD	VC, PA, LB, MB, RD	3 per combination	75
Treatment 2	Mixed inoculum (VR) added to defaunated sediments, including mixed sediment	VR (VC + RD mixture)	VC, PA, MB, RD + (VC50+RD10)	3 per combination	15

VC (Avon River), PA (Percuil River), LB (Helford River), MB (Mylor Creek), RD (Restronguet Creek), VR (mixture of 10 g VC and 10 g RD inoculum sediments), VC50+RD10 (mixture of 50 g VC and 10 g RD defaunated sediments).

A) Control Samples (Baseline): Five containers were set up, each containing only inoculum sediment from one of the five sites (VC, PA, LB, MB, or RD). These controls served as baselines representing the initial community state before any experimental manipulation, allowing detection of changes due to the experimental setup itself. Each control was replicated once, resulting in a total of 5 samples.

B) Treatment 1; Cross-Inoculation Experiment: Inoculum sediments from each of the five sites were added to defaunated sediments from all five sites, resulting in 25 unique combinations (5 inoculum sediments × 5 defaunated sediments). Each combination was replicated three times, totalling 75 samples. This treatment aimed to investigate how communities from different pollution backgrounds respond when introduced into sediments with varying copper concentrations (**Table 2.3**).

C) Treatment 2; Mixed Community Interactions: A mixed inoculum (VR) was prepared by combining 10 g of sediment from the cleanest site (VC) with 10 g from the most polluted site

(RD). This mixed inoculum was added to defaunated sediments from four estuaries (VC, PA, MB, RD) and to a mixed defaunated sediment prepared by combining 50 g of clean defaunated sediment (VC) with 10 g of copper-polluted defaunated sediment (RD), referred to as VC50+RD10. This resulted in 5 combinations, each replicated three times, totalling 15 samples. This treatment was designed to study interactions between clean and polluted communities and to assess how communities from cleaner environments respond when introduced to partially contaminated sediments, such as the VC50+RD10 mixture (**Table 2.3**).

Table 2.3 Application of inoculum sediment additions to defaunated sediments, including control (baseline) and various treatment groups. Green-coded entries represent mixtures of two sediment sources, while black-coded entries denote control samples.

Treatment	Sample		Defaunated sediments				
			VC	**PA**	**LB**	**MB**	**RD**
1	**Inoculum sediments**	**VC**	VC	VC/PA	VC/LB	VC/MB	VC/RD
		PA	PA/VC	PA	PA/LB	PA/MB	PA/RD
		LB	LB/VC	LB/PA	LB	LB/MB	LB/RD
		MB	MB/VC	MB/PA	MB/LB	MB	MB/RD
		RD	RD/VC	RD/PA	RD/LB	RD/MB	RD
2		**VR**	VR/VC	VR/PA	VR/MB	VR/RD	VR/(VC50+RD10)

VC (Avon River), PA (Percuil River), LB (Helford River), MB (Mylor Creek), RD (Restronguet Creek), VR (mixture of 10 g VC and 10 g RD inoculum sediments), VC50+RD10 (mixture of 50 g VC and 10 g RD defaunated sediments).

Each microcosm consisted of a 600 ml plastic container containing 20 g of inoculum mud and 60 g of defaunated sediment, topped with 500 ml artificial seawater. The seawater was adjusted to the salinity of the collection site (20 S) and pH was measured (7.8 - 8). The mud mixtures were evenly distributed into 90 ventilated containers. they were gently aerated using a line diffuser system to ensured proper oxygen exchange to maintain aerobic conditions throughout the experiment. The samples were incubated in a dark environmental chamber, with the temperature gradually increased by 1-2°C per day until the experimental temperature of 20°C was reached, which was then maintained for the 60-day treatment period. Upon completion of

the experimental period, each microcosm was dismantled, and a 10 g aliquot of sediment was taken from each replicate for DNA extraction.

2.1.3 Porewater Extraction

Porewater was extracted within 48 hrs of arrival in the laboratory. This was done by centrifuging homogenised sediments at 3500 rpm for 10 mins (Simpson et al., 2000). Three pseudo-replicates were processed per site. Unfiltered composite samples were collected for salinity and pH measurements, as required (**Section 2.2.3**). Extracted porewater samples were filtered using 0.45 µm Fisherbrand™ Nylon Syringe filters. Prior to filtration, syringes and syringe filters were flushed, in sequence, with 5 - 10 mL of 10% HNO_3, 20 mL of ultrapure water (UPW; Elga Purelab Ultra, 18.2 MΩ cm) and 2 mL of sample. Porewater was filtered to <0.45 µm as in subsequent surveys. Filtered porewater was subsampled for trace metal analysis, acidified to 2% HNO_3 and refrigerated at 4^{o}C until use.

2.1.4 Measurement Of Porewater pH And Salinity

Porewater salinity and pH were determined. The pH was measured in the laboratory from porewater samples using a pH meter (Mettler Toledo SevenEasy S20). The instrument was calibrated at pH 4 and 7 using certified standards (NIST - National Institute of Standards and Technology). Salinity was measured in the laboratory using Fisherbrand™ Traceable™ Salinity Meter, which was manufacturer-certified and calibrated.

2.1.5 Determination Of Acid-Extractable Metal Concentrations

Acid-extractable metal concentrations were determined using subsamples of the same homogenised sediment used for the DNA extraction. Approximately 100 mg (dry weight equivalent) of wet sediment was extracted for 30 mins in 10 mL of 1 M HCl. Method blanks were processed by extracting only ultra-pure water (UPW) to account for potential contamination in the water and reagents. Sediment extracts were analysed for metals by Triple Quadrupole Inductively Coupled Plasma - Mass Spectrometry (ICP - MS- QQQ; Thermo Icap-TQ). Single-element calibration standards (PlasmaCAL) were prepared in a similar acid matrix. A multi-element standard solution (CLMS2A, SPEX CertiPrep) was prepared in a similar

matrix to assess precision of the analysis. And accuracy was assessed using certified lake water references, TM-27.3 and TMDA-64.2, both from Environment Canada.

2.1.6 Characterization of Sediment Particle Size Distribution

Median sediment grain size (D50) and the percentage of fines (LT63) were assessed using 5 g of homogenised and composite wet samples. A volume 25 mL of deionised water and 1 mL of 6.2 g/L sodium hexametaphosphate was added to the weighed sediment in a beaker (Kenny and Sotheran, 2013). The sediment suspension was vortexed overnight and, afterwards, washed through a 2 mm sieve. The fraction that passed through the sieve was stirred at 1000 rpm for 3 minutes to disperse clay particles and then analysed, in triplicates, by laser diffraction using the Malvern Mastersizer 2000. Results from the laser diffraction analyses were used to calculate the median grain size (50th percentile) and the percentage of fines (particles smaller than 63 µm). Additionally, the Folk and Ward (1957) inclusive graphic standard deviation (σ1) was estimated. This measure of dispersion is better suited for addressing deviations from a normal grain size distribution. The method categorises the degree of sediment sorting on a verbal scale (Folk and Ward, 1957) as follows:

$\sigma 1 < 0.35$ Very well sorted (VWS)

$0.35 \leq \sigma 1 < 0.50$ Well sorted (WS)

$0.50 \leq \sigma 1 < 1.00$ Moderately sorted (MS)

$1.00 \leq \sigma 1 < 2.00$ Poorly sorted (PS)

$2.00 \leq \sigma 1 \leq 4.00$ Very poorly sorted (VPS)

$\sigma 1 > 4.00$ Extremely poorly sorted (EPS)

2.1.7 DNA Extraction Methods

2.1.7.1 Sample Size and Whole-Sediment Extraction Tests

Nematodes were a dominant component of marine sediments, with densities of 0.5 to 5 million individuals per square metre in shallow marine environments (Schratzberger et al., 2019). Considering a sediment depth of 1 cm and an average wet bulk density of 1.51 g/cm³ for marine mud (Endler, 2009), this equated to approximately 33 to 331 nematodes per gram of sediment.

Whole-sediment sub-samples of just 0.25-0.5 g were nevertheless sufficient to yield large numbers of prokaryotic reads (Tian et al., 2015), but may underrepresent larger metazoans such as nematodes.

To assess whether amplifying DNA directly from whole sediment effectively captures a substantial number of amplicon sequence variants (ASVs), DNA was extracted from approximately 10 ml of uncontaminated sediment from Breydon Water, Norfolk (BW), along with sieved fractions from the same sample. As shown in **Table 3.5** (**Chapter 3**), average nematode ASV recovery was compared across four treatments (sieved worms and whole sediment at 0.25 g and 2.7 g). These comparisons helped evaluate which approach best captures nematode diversity.

2.1.7.2 Sieved Samples DNA Extraction

Mud samples from Breydon Water were collected and immediately preserved on site by adding DESS solution to reach a final volume of 50 mL. The DESS solution comprised 20% dimethyl sulphoxide (DMSO), 0.25 M disodium EDTA and was saturated with NaCl, following the method of Yoder et al. (2006). These preserved samples were refrigerated until processed. The sieving procedure for meiofauna samples adhered to the approach described by Warwick and Clarke (1998). Using ultrapure water, samples were thoroughly washed through 500 µm and 63 µm sieves to remove preservatives, salts, macrofauna and excess silt. Fine-mesh sieving was well suited to nematode studies but could add labour and risked losing delicate or fast-swimming taxa (Lallias et al., 2015; Kim et al., 2017).

To extract nematodes, the material retained on the 63 µm sieve was transferred into a 50 mL centrifuge tube using a 50% solution of Ludox® HS-40 (Sigma-Aldrich). The samples were agitated and centrifuged at 1,800 g for 10 minutes, following the protocol of Heip et al. (1985). High-density Ludox HS40 centrifugation rapidly separated organisms of differing densities from bulk sediment, improving retrieval efficiency (McIntyre & Warwick, 1984; Burgess, 2001). The supernatant containing the nematodes and other organisms was decanted onto a 63 µm sieve and the residue in the centrifuge tube was refilled with the 50% Ludox solution. This centrifugation and decantation process was repeated three times in total for each replicate sample.

After thorough washing with ultrapure water to remove any remaining Ludox, the nematodes and other organisms were filtered by pouring them onto glass fiber filters (47 mm GF/F Whatman, pore size 0.7 μm) using a vacuum pump. A portion weighing 0.25 g was collected and transferred into bead tubes for DNA extraction. These bead tubes were part of the DNeasy® PowerSoil® Pro kit (Qiagen, Hilden, Germany).

2.1.7.3 0.25 g Sediments

Two replicate samples were collected from two clean sites in Breydon Water, Norfolk. The DNA was extracted from 0.25 g of whole sediment and sieved material from each sample using the DNeasy® PowerSoil® Pro kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Hermans et al. (2018) suggest that the DNeasy kit is a powerful and universal DNA extraction method. The extraction process involved mechanical and chemical lysis through bead beating, followed by centrifugation. The supernatant was treated to remove inhibitors, such as humic acids, which could interfere with downstream applications. The DNA was then bound to a silica membrane in a spin column, washed to remove any remaining contaminants and eluted into a buffer for subsequent analysis, including PCR or sequencing.

2.1.7.4 1.4 - 2.7 g Sediments

Three replicate samples were collected from 12 intertidal locations in estuaries in SW England and two in Breydon water, Norfolk (details in Udochi, 2020). Both DNA and RNA were extracted from 1.4 - 2.7g of each sample by Charlotte Davies using the RNeasy PowerSoil Total RNA kit and DNA elution accessory kit (Qiagen, Hilden, Germany) according to an optimised version of the manufacturer's instructions with an added heat block step (45°C for 15 minutes) prior to the solution being added to the column. Notably, this extraction approach used a different kit than those employed for other sample sizes, which might have introduced minor variation in the diversity patterns observed.

2.1.7.5 10 g Sediments

The DNA was extracted from approximately 10 g of sediment from each of the three replicate samples collected at 34 intertidal locations across estuaries in Southwest England. The extraction followed a modified protocol based on the DNeasy® PowerSoil® Pro kit (Qiagen,

Hilden, Germany) adapted with permission from an unpublished protocol developed by Alastair Grant and Solomon Udochi (2021).

All reagents for DNA extraction, purification and concentration were prepared according to the protocol outlined in **Table 2.4**, with pH adjusted as required. Solutions were sterilized by autoclaving at 120 °C for 15 minutes, except for Buffer C1, which was filtered using sterile 0.2 μm PES filters to prevent degradation of SDS. Alcohols for Buffers C4 and C5 were added after sterilization to maintain their concentrations. Laboratory equipment; including 50 mL Falcon tubes, 1.5 - 2 mL Eppendorf tubes, 1.5 mL spin columns, 4 mm glass beads and hydrophilic magnetic beads (Sera-Mag Carboxylate-Modified Magnetic Particles, GE Healthcare Life Sciences) was sterilized prior to use. Prepared solutions were refrigerated and stored covered to prevent contamination.

Table 2.4 Protocol for the preparation of extraction and elution buffers.

Name	Contents	pH
Bead tube	5 mL of 4 mm Glass Beads in a 50 mL Falcon tube	N/A
Bead solution	181mM $NaPO_4$, 121 mM guanidinium isothiocyanate	8.94
C1	150mM NaCl, 4% SDS, 0.5 M Tris	10.94
C2	133 mM Ammonium acetate	7.37
C3	120 mM Aluminium ammonium sulphate dodecahydrate	3.29
C4	5 M GuHCl, 30 mM Tris, 9% isopropanol	4.24
C5	10 mM Tris, 100 mM NaCl, 50% EtOH	7.51
C6	10 mM Tris	8.71

2.1.7.5.1 10 g Sediments Isolation and Purification

The DNA extraction from the 10 g sediment samples involved several steps to ensure high yield and purity suitable for downstream applications. Each sediment sample was placed into a sterile 50 mL Falcon tube containing 5 mL of 4 mm glass beads. To facilitate cell lysis, 15 mL of bead solution and 1.2 mL of Buffer C1 were added to the tube. The mixture was then vortexed vigorously for 10 to 15 minutes to thoroughly homogenize the suspension and break open the cells through bead beating. After bead beating, the tube was centrifuged at 3,500 RPM

for 2 minutes to separate the supernatant, which contains the DNA and associated impurities, from the sediment particles and glass beads. The clear supernatant was carefully transferred to a new sterile tube.

Purification began by adding 5 mL of Buffer C2 to the supernatant, followed by a brief vortex and incubation at 4 °C for 10 minutes to precipitate contaminants. The tube was then centrifuged at 3,500 RPM for 4 minutes and the supernatant was transferred to another sterile tube. Next, 4 mL of Buffer C3 was added, mixed by vortexing for 10 seconds and incubated again at 4 °C for 10 minutes to facilitate the formation of aluminium hydroxide precipitates that aid in removing PCR inhibitors like humic substances. After a final centrifugation at 3,500 RPM for 4 minutes, the supernatant should have appeared clear that indicating the effective removal of inhibitors. This purified DNA solution was stored at -20 °C for short-term or -80 °C for long-term preservation, with samples kept on ice during handling to ensure minimal degradation.

The next phase involved concentrating and further purifying the DNA using magnetic beads. A 1% Tween-20 solution and 80% ethanol were prepared as needed. Magnetic beads were pre-cleaned by mixing them with the Tween-20 solution and then separated using a magnet to remove any residual impurities. For each replicate sample, 25 µL of the cleaned bead suspension was added to the DNA solution, followed by the addition of isopropanol (0.7× supernatant volume, ~19 mL) and Tween-20 (0.02%) to achieve the desired final concentrations. The mixture was incubated on ice or at 4 °C for 10 minutes to allow the DNA to precipitate onto the beads. Centrifugation at 1,500 RPM for 9 minutes separated the bead-DNA complex from the supernatant without damaging the beads. The beads were then washed twice with 500 µL of 80% ethanol and air-dried for 5 minutes to remove any remaining ethanol, which could inhibit PCR amplification.

The DNA was eluted from the magnetic beads by adding 300 µL of Buffer C6 and gently mixing. The supernatant containing the DNA was then transferred to a new tube. To ensure purity, a spin column clean-up was performed by adding 4 to 5x volume of Buffer C4, incubating and then passing the mixture through a spin column. The column was washed twice with 500 µL of Buffer C5 to remove any residual contaminants before eluting the purified DNA with 100 µL of Buffer C6. This final step ensured that the DNA was free from inhibitors and

suitable for PCR amplification with the method yielding significantly higher concentrations of DNA compared to conventional extraction techniques.

For the 10 g microcosm samples, an additional modification was introduced to further decrease PCR inhibitors, enhance extraction quality and increase the number of samples that could be processed simultaneously. Following the bead beating process, only 0.5 mL of the total supernatant volume of approximately 17 mL was removed. This adjustment enabled the remaining cleanup steps to be carried out in microcentrifuge tubes, simplifying the extraction process. The volumes for the inhibitor removal and precipitation steps were adjusted accordingly to accommodate the decreased supernatant volume, thereby improving overall DNA recovery and inhibitor removal.

2.1.8 PCR Amplification

Polymerase chain reaction (PCR was conducted in a volume of 20 µl, including 10 µl of PhusionTM Flash master mix (Thermo Scientific, UK), 1 µl of forward primer and 1 µl of reverse primer (final concentration of primers: 100 µM), 7 µl of ultrapure sterile water (MilliQ water) and 1 µl of DNA (with a concentration between trace amounts and 9.3 ng). The amplification of the DNA templates was carried out in either a Veriti™ HID 96-Well Thermal Cycler, 0.2mL system (Applied Biosystems, UK) or 0.2 ml PCR tubes, depending upon the number of samples be amplified.

To ensure the accuracy and reliability of the PCR and sequencing processes, microbial community standards (ZymoBIOMICS, Zymo Research, Irvine, CA, USA) were employed as positive controls and detailed results and analysis of the control DNA amplification and sequencing were provided in **Chapter 4**. Additionally, to identify primer pairs that provide reliable identification across a wide range of taxa, PCR amplification was carried out using thirteen barcoded primers. In brief, 11 primer pairs which successfully amplified PCR products of the correct size were used to target bacterial 16S, Archaeal 16S, fungal ITS, eukaryote ITS, eukaryote 18S and 28S, and metazoan COI (**Table 2.5**). Each includeed 8 base pair barcodes, length heterogeneity spacers and adapters for Illumina sequencing appended to their 5’ ends, following (Caporaso et al., 2012). PCR conditions had an initial denaturation at 98°C for 10 minutes, followed by 28 to 35 cycles of denaturation at 98°C for 30 seconds; annealing at a

primer specific temperature for 30 seconds and extension at 72°C for 30 seconds. This was followed by a final extension at 72°C for 5 minutes, before being held at 4°C. The PCR success varied between different sediments and annealing temperatures and cycle numbers were adjusted to optimise amplification yield and specificity. The final values used of annealing temperature and cycle numbers were given in **Chapter 3** (**Table 3.1).** the COI primer pairs; JB3, JB2 and mlCOI followed different PCR protocols which detailed in **Chapter 3.**

Table 2.5 Gene-specific primers, their approximate PCR product sizes, primer pairs, sequences and corresponding sources. PCR product sizes are listed without barcodes and Illumina adapters. IUPAC codes for ambiguous bases indicate that the primer was synthesised with an equimolar mixture of these bases at this position.

Gene specific primer	Approx PCR product size bp	Primer pair	Primer sequence	Source
16S	410	**515F**	GTGCCAGCMGCCGCGGTAA	Turner et al. (1999)
		926R	CCGYCAATTYMTTTRAGTTT	Quince et al. (2011)
	520	**SSU1ArF**	TCCGGTTGATCCYGCBRG	Bahram et al. (2019)
		SSU520R	GCTACGRRYGYTTTARRC	Bahram et al. (2019)
ITS	very variable	**ITS1f12**	GAACCWGCGGARGGATCA	Schmidt et al. (2013)
		ITS2	GCTGCGTTCTTCATCGATGC	White et al. (1990)
	750	**VRAIN2F**	CTTTGTACACACCGCCCGTCGCT	Vrain et al. (1992)
		VRAIN2R	TTTCACTCGCCGTTACTAAGGGAATC	Vrain et al. (1992)
18S	310	**1391f**	GTACACACCGCCCGTC	(Lane, 1991)
		EukBr	TGATCCTTCTGCAGGTTCACCTAC	(Medlin et al., 1988)
	460	**D512**	ATTCCAGCTCCAATAGCG	Zimmermann et al. (2011)
		D978R	GACTACGATGGTATCTAATC	Zimmermann et al. (2011)
	420	**TAReuk454FWD1**	CCAGCASCYGCGGTAATTCC	Stoeck et al. (2010)
		TAReukREV3	ACTTTCGTTCTTGATYRA	Stoeck et al. (2010)
	400	**G18S4**	GCTTGTCTCAAAGATTAA GCC	Blaxter et al. (1998)
		22R	GCCTGCTGCCTTCCTTGGA	Blaxter et al. (1998)
	500	**NEM**	GGGGAAGTATGGTTGCAAA	Sapkota and Nicolaisen (2015)
		18Sr2b	TACAAAGGGCAGGGACGTAAT	Porazinska et al. (2009)
28S	500	**DM568F**	TTGAAACACGGACCAAGGAG	Kounosu et al. (2019)
		RM3R	CRCCAGTTCTGCTTACCAAAA	Kounosu et al. (2019)
COI	370	**JB3adjusted**	TGGGCATCCTGAGGTTTAT	Tytgat et al. (2019)
		JB5	AGCACCTAAACTTAAAACATAATGAAAATG	Derycke et al. (2005)
	370	**JB2**	ATGTTTTGATTTTACCWGCWTTYGGTGT	Derycke et al. (2007)
		JB8	CCCCTCTAGTCTWCTATTTCTTAATAC	Derycke et al. (2007)
	310	**mlCOIintF**	GGWACWGGWTGAACWGTWTAYCCYCC	Leray et al. (2013)
		jgHCO2198	TAIACYTCIGGRTGICCRAARAAYCA	Geller et al. (2013)

2.1.9 DNA and PCR Products Visualisation

The amplified DNA fragments were checked by electrophoresis using a 1 % (W/V) agarose TBE gel visualised with SafeWhite (NBS Biological Ltd, UK) and size was assessed using Thermo Scientific™ GeneRuler 100 bp DNA Ladder (Fisher Scientific, UK). Additionally, this assessment was essential for verifying DNA quality, ensuring that the extracted DNAis suitable for downstream applications. Additionally, it enabled the estimation of molecular weight by comparing the migration distances of the samples to a DNA ladder, thereby confirming that the amplified products corresponded to the expected sizes (Sambrook, 1989, Lee et al., 2012).PCR Product Purification

The PCR products were purified using Aline Biosciences PCRClean DX kit (Aline Biosciences, Woburn, USA) following the manufacturer protocol except that the ratio of bead suspension to PCR product which was 1.8:1 1st run, 1:1 2nd and 3rd runs and 0.7:1 at 4th run. This ratio was altered in the light of experience to improve removal of free primer adaptors and adaptor dimers, which were not fully removed when the ratio in the manufacturer's protocol (1.8:1) was used (Quail et al., 2009). The PCR products quantification followed methods described in **section 2.2.8**.

2.1.10 Quantification

To quantify the DNA and PCR products, samples were also analysed by using a StepOneTM Real-Time PCR system (Applied Biosystems, UK) as a fluorimeter with SYBR Green I nucleic acid gel stain supplied by Sigma-Aldrich Company limited according to the manufacturer's protocol. The maximum excitation wavelength of SYBR Green was 497 nm. A 10,000x solution in dimethyl sulfoxide (DMSO) working solution was prepared by diluting the SYBR Green reagent with the manufacturer's buffer at a ratio of 1:200. A volume of 190uL of working solution was added to 10uL of the two DNA standards provided as part of kit, while 199uL of working solution was added to 1uL of each DNA sample. The PCR microplate was gently vortexed and then left to incubate at room temperature for 2 minutes before fluorescence was measured and DNA concentrations calculated using a two-point calibration, with newly prepared standards for each set of samples. Additionally, extraction yield and DNA molecular weight were assessed using gel electrophoresis as detailed in **section 2.5.4**.

2.1.11 Sequencing

The PCR products from multiple samples were pooled in equimolar amounts, with slightly lower amounts used for samples with weak amplification, to create a single composite sample. This initial pool of the pilot samples was sent to the Earlham Institute in Norwich, UK, for sequencing on a single lane of the Illumina MiSeq, using a pre-made library for 300 bp paired-end sequencing, followed by NovaSeq 6000 flow cell with 250 bp paired-end for the 12 and 34 sets then NextSeq 1000 with 300 bp paired-end for the experiment set of samples. Different sequencing platforms were used due to updates in available technology and platform-specific optimisation.

2.1.12 Analysis Of Sequences and Data Visualisation

Sequencing data from the first run were initially analysed using Mothur software (version 1.45.3; www.mothur.org) (Schloss et al., 2016) and USEARCH (version 11.0.667; www.drive5.com/usearch) (Edgar, 2010) to convert the raw sequencing data into a sites x ASVs table and identify the closest matching sequences in the SILVA, UNITE and MIDORI databases. Initial decisions on which primers to use to amplify all samples were made on the basis of these results. Subsequently all data was re-evaluated with updated taxonomic tools and methods via the LotuS2 pipeline (Özkurt et al., 2022) (**Fig. 2.3**), unless indicated otherwise the ASV abundances and taxonomic annotations presented here were obtained using LotuS2.

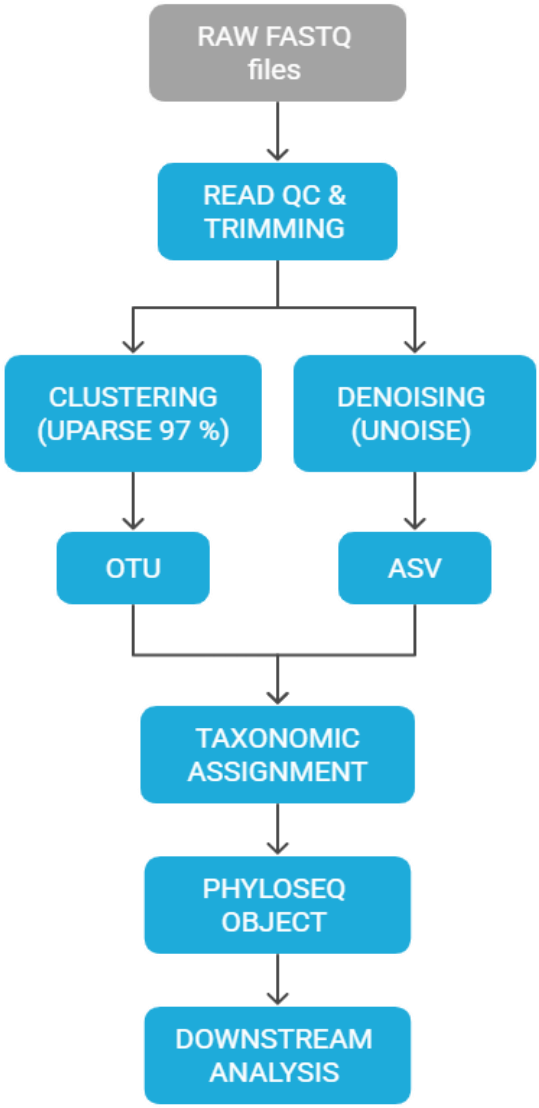


Fig. 2.3 Schematic overview of the LotuS2 bioinformatics workflow. Raw reads were quality filtered and processed via either OTU clustering or ASV denoising, followed by taxonomic assignment and data structuring for ecological analysis.

The pipeline LotuS2 integrateed pre-processing, sequence clustering, taxonomic assignment and phylogenetic tree construction. Pre-processing involved trimming primer sequences and sample-specific barcodes, alongside stringent quality filtering to retain high-quality reads. Operational Taxonomic Unit (OTU) construction was performed using UPARSE at 97% clustering and ASVs were generated using the UNOISE algorithm, which denoised sequences to correct errors and recover all true biological sequences in the dataset. All bacterial 16S and metazoan ITS, 18S, 28S and COI reads were denoised to amplicon sequence variants (ASVs)

to preserve single-nucleotide resolution, which improves accuracy and ecological sensitivity (Callahan et al., 2017). For Archaea, clustering was performed at 97% OTUs, as reference coverage for archaeal 16S remained sparse, collapsing near-identical reads reduces spurious singletons and aligns with current archaeal practice (Grant et al., 2023, Regueira-Iglesias et al., 2023). The Lowest Common Ancestor (LCA) method was employed to improve taxonomic classification accuracy. These approaches were selected based on their demonstrated accuracy and reproducibility in ecological studies (Özkurt et al., 2022). The resulting sites x ASV tables, taxonomic annotations and phylogenetic trees were provided as R Phyloseq objects or text files by Professor Alastair Grant, School of Environmental Sciences, UEA.Taxonomic assignment for bacteria and Archaea was conducted using KSGP version 1.0 (Grant et al., 2023) with the GTDB taxonomy hierarchy, while Eukaryote assignments utilised Eukaryome version 1.7 (www.eukaryome.org) (Tedersoo et al., 2024), which specifically covered the 18S, ITS and 28S markers. Cytochrome oxidase I (COI) assignments employed MIDORI-Longest (www.reference-midori.info) (Leray et al., 2018). Targeted primers, sequences and their PCR product sizes for each amplicon were presented in **Table 2.5**.

Data handling, graph plotting and statistical analyses were carried out using R (R Core Team, 2024). The relative abundance of taxonomic groups were calculated and visualised with the Phyloseq v1.48.0 (McMurdie and Holmes, 2013) and Vegan v2.6.4 (Oksanen et al., 2022) packages, while all univariate and multivariate indices were computed using the Vegan package. Rarefaction analysis was performed to standardise sampling effort and compare species richness across samples Sanders (1968), using the Phyloseq package. In addition, sample-based accumulation curves first described by (Arrhenius, 1921), were generated with the vegan function specaccum to examine how cumulative ASV richness increased as additional sites were added (Sanders, 1968, Béguinot, 2016), thereby testing whether the 12- and 34-site surveys had approached sampling sufficiency. The K-dominance curves, generated using the Vegan package were employed to assess species dominance patterns and community structure across samples (Clarke, 1990a).

To further evaluate the effectiveness of different primers and sample treatments in detecting nematode species diversity, a uniform rarefaction depth of 100 reads was selected to approximate the number of individuals typically identified in morphological analyses, thereby facilitating direct comparisons between molecular and morphological diversity estimates

(Hurlbert, 1971, Gotelli and Colwell, 2001). Box plots of ASV richness at this depth were generated for each primer across diverse sample types, illustrating total species richness detected. The K-dominance curves, which depict the cumulative relative abundance of species ranked from most to least abundant, were also applied to compare species dominance patterns and community structures between samples (Sanders, 1968, Warwick, 1986). These approaches are widely applied in marine ecology to assess biodiversity and understand the influence of environmental factors on community composition (Warwick, 1986, Clarke, 1990b).

Principal Coordinate Analysis (PCoA) by Gower (1966) and Non-metric Multidimensional Scaling (NMDS) by Kruskal (1964) are ordination methods used to visualize patterns in complex datasets by reducing their dimensions. According to the two authors, PCoA transforms dissimilarity matrices into orthogonal axes to preserve distances between data points and NMDS focuses on preserving the rank order of distances, making it suitable for ecological data that may not meet metric assumptions. They were performed using the ordinate function from the Phyloseq package, was used to reduce dimensionality and visualise patterns in the dataset, offering an alternative means of exploring variation in community structure. Principal Components Analysis (PCA), originally introduced by Pearson (1901) and refined by Hotelling (1933) is a dimensionality reduction technique widely used to uncover patterns in multivariate datasets. These methods, combined with cluster analysis, have proven valuable for distinguishing microbial communities across diverse environments (Ramette, 2007).

The Bray-Curtis coefficient is a widely used metric for analysing microbial communities, quantifying compositional dissimilarity based on relative abundance without considering evolutionary relationships (Bray and Curtis, 1957). This measure indicates 0% similarity for communities with no shared species and 100% similarity for those with identical compositions (Clarke et al., 2014), making it particularly effective for detecting changes in community composition due to environmental factors like pollution. Joint absences do not influence calculations, ensuring that Bray-Curtis similarity between two samples remains unaffected by ASVs absent in both. Its sensitivity to relative abundances and exclusion of joint absences make it well-suited for assessing ecological dissimilarities, offering a robust framework for analysing compositional shifts driven by environmental gradients or disturbances (Ricotta and Podani, 2017). Conversely, UniFrac functions as a phylogenetic measure that evaluates

community similarity based on evolutionary relationships (Lozupone and Knight, 2005). For this analysis, Weighted UniFrac was employed, which accounts for both the phylogenetic distance and the relative abundance of taxa in each community. This approach was particularly suitable for comparing microbial communities based on phylogenetic relationships, regardless of differences in abundance (Lozupone and Knight, 2005). The distance matrix was calculated with the Phyloseq and Vegan packages. Originally proposed by Jaccard (1901), the Jaccard coefficient quantifies similarity between two samples on the basis of shared presences while ignoring joint absences. It is defined as the size of the intersection divided by the size of the union of the sample sets, yielding values from 0 (no shared taxa) to 1 (identical composition). Because it relies solely on presence-absence data, Jaccard is often preferred in metabarcoding studies where read abundances can be distorted by variable gene-copy number and PCR bias (Elbrecht and Leese, 2015).

To explore the influence of environmental variables on community composition, BIOENV analysis was performed using the Vegan package identifying key environmental drivers of microbial variation (Clarke and Ainsworth, 1993). Hierarchical clustering dendrograms created with the hclust function in R (Murtagh and Legendre, 2014) were used to evaluate clustering patterns among sites based on community composition (Sokal and Rohlf, 1962). Dendrograms, first described by Sokal and Rohlf (1962), provided a visual representation of site dissimilarities. The hierarchical clustering dendrogram illustrated the structure in site relationships, demonstrating clear ecological gradients from clean to heavily polluted sites (Pang et al., 2023). Bubble plots generated with ggplot2 (Clarke and Ainsworth, 1993) and ggtree (Yu et al., 2017) were used to visualise the relationships between ecological patterns and environmental variables (Somerfield et al., 1994a). The SIMPER analysis (Similarity Percentage) (Clarke and Ainsworth, 1993), conducted with the Vegan package, identified which ASV/OTUs contributed most to differences between sample groups, providing insights into copper-tolerant and copper-sensitive taxa. This approach helps determine key taxa driving ecological variation and understand how environmental factors, such as pollution, affect community composition (Terlizzi et al., 2005, Piola and Johnston, 2008). The ANOSIM (Analysis of Similarity) (Clarke, 1993) was used to quantify replicate clustering within sites and overall site differentiation using Bray-Curtis dissimilarity matrices. The R statistic measured group separation, with values near 1 indicating strong differentiation and values near 0 signifying overlap and statistical significance was assessed through 999 permutations.

Analysis of Similarity (ANOSIM) (Clarke, 1993) was used to quantify replicate clustering within sites and overall site differentiation using Bray-Curtis dissimilarity matrices. The R statistic measures group separation, with values near 1 indicating strong differentiation and values near 0 signifying overlap and statistical significance was assessed through 999 permutations. Additionally, An Analysis of Variance (ANOVA) and Post-hoc comparisons using Tukey's HSD test conducted in R, following Levene's Test for homogeneity of variances using the car package (Fox and Weisberg, 2019) to compare ASV/OTUs richness across sites and identify significant differences in diversity metrics (St and Wold, 1989). Phylogenetic trees were visualised using the ggtree package providing insights into the evolutionary relationships among ASV/OTUs. Finally, removal of rare ASV/OTUs was implemented using custom scripts in R to evaluate the impact of low-abundance taxa on clarifying community structure and reducing noise.

2.2 Results And Discussion

2.2.1 Environmental Variables Across Study Sites

Understanding the environmental variables is crucial for interpreting metal bioavailability and its potential impact on benthic organisms and microbial communities. Key physicochemical parameters measured included porewater pH, salinity, sediment grain size, total organic carbon (TOC) and dissolved organic carbon (DOC). These factors influenced metal speciation, mobility and bioavailability in estuarine environments (Bryan and Langston, 1992, Chapman and Wang, 2001, Luoma and Rainbow, 2008). The ecological relevance of these patterns was explored in later chapters in relation to taxonomic composition and diversity.

All sites and corresponding results discussed in this section were from the 34-site dataset unless explicitly stated otherwise. To further investigate how metal contamination and other variables shape site differences, PCA was performed on both the 12-site and 34-site datasets. In the 12-site dataset (**Fig. 2.4 A**), heavily polluted sites (HA, HB, RA, RB, RC) clustered on the left, aligning with elevated copper-related variables (e.g., eqpCu, AEMCu), whereas less impacted sites (e.g., BW, HR, MC, PC, PR, CO) grouped on the right, reflecting comparatively lower metal loads and underscoring copper's dominant influence. In the 34-site dataset (**Fig. 2.4 B**), heavily polluted sites (HA, HB, RA, RB) aligned with vectors for porewater copper (PWCu)

and AEMCu, while less contaminated sites associated more strongly with sediment grain-size variables (D50, LT63), indicated that both metal contamination and sediment characteristics drive site separation. Correlations between site clusters and diversity levels were evident at heavily polluted sites, although patterns differed among taxonomic groups, as shown in **Chapters 4, 5**, and **6**.

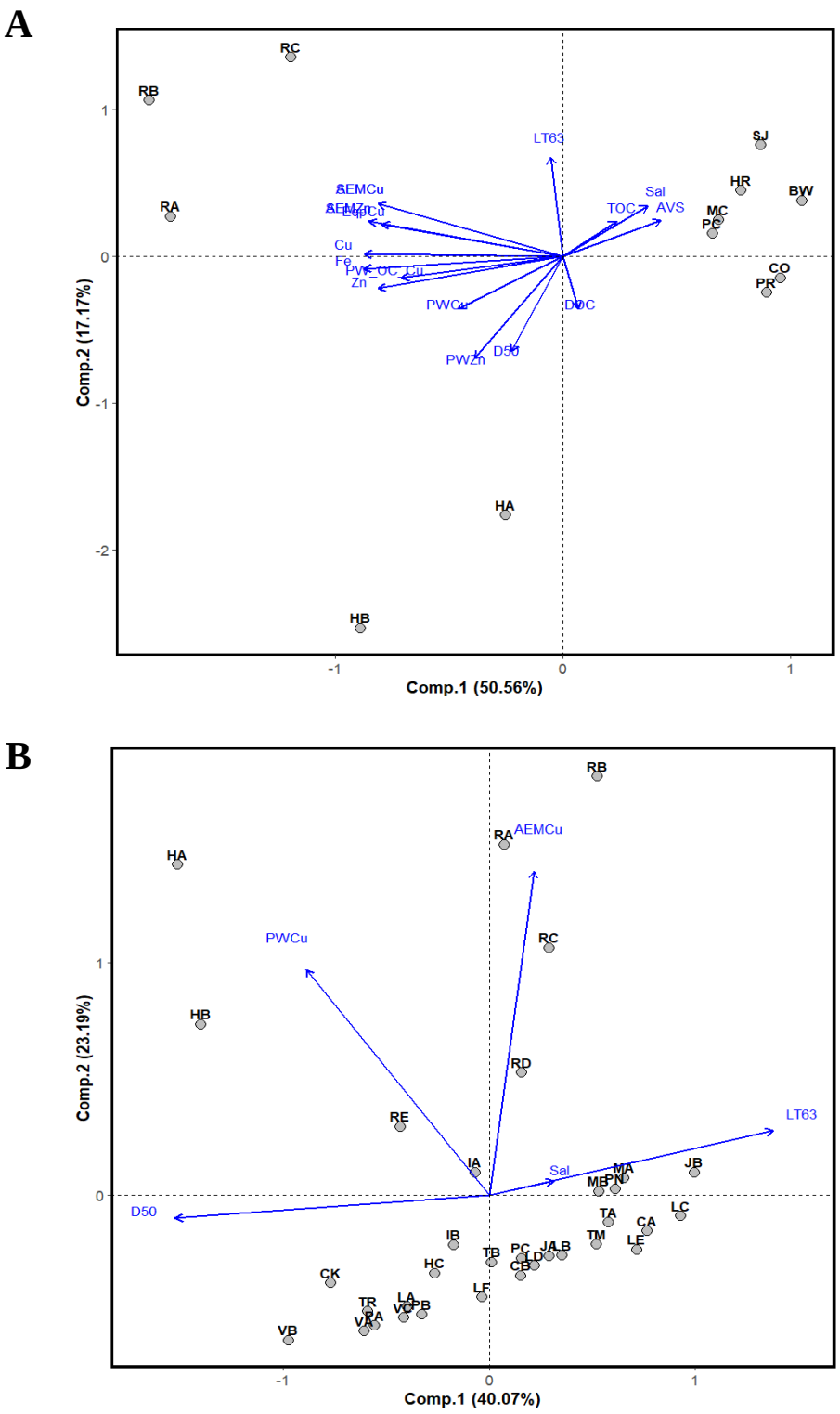


Fig. 2.4 Principal Components Analysis (PCA) ordination of standardised environmental variables averaged across the sampling surveys. **A**) 12 sites. **B**) 34 sites. Arrangement from left to right along the Comp. 1 axis represents changes in environmental gradients across the study sites. Environmental variables are represented by blue arrows and sites are shown as grey-filled circles. Site codes as in **Table 2.1.** PWCu (Porewater Cu), AEMCu (Acid-Extractable Cu), D50 (Median Grain Size), LT63 (<63 µm Fines).

2.2.2 Porewater pH and Salinity

Salinity measurements across the 34 study sites exhibited variation, highlighting the influence of freshwater inputs and tidal mixing in the estuaries were presented in **Tables 2.6 and 2.7**. The River Avon sites (VA, VB and VC) demonstrated lower salinities ranging from 0.9 to 4.9 S, characteristic of upper estuarine conditions. In contrast, sites in the Helford River and Percuil River showed higher salinities, approaching marine conditions with values between 15.2 and 38 S.

Porewater pH values varied from 6.3 at RA (Restronguet Creek) to 7.82 at LF (Helford River), with lower pH levels generally observed at more contaminated sites. These measurements were similar to those previously reported in the Fal Estuary (Bryan and Gibbs, 1983, Perryman, 1996) and were comparable to expected seawater values. Evidence suggested that the Carnon River, which flows into the Fal Estuary, can be as acidic as pH 4.65 influencing metal behaviour in estuarine environments.

Table 2.6 12-site set Key physicochemical characteristics of surface (top 2 cm) sediments sampled in Autumn 2017. Reprinted from (Udochi, 2020).

Site	Total Cu (µg/g)	Total Zn (µg/g)	Fe_2O_3 (%)	SEM Cu (µmol/g)	SEM Zn (µmol/g)	AEM Cu (µg/g)	AEM Zn (µg/g)	<63 µm (%)	AVS (µmol/g)	PW Cu (µg/L)	PW Zn (µg/L)	TOC (%)	DOC (mg/L)	EqP Cu (µmol/g OC)	PW/OC Cu (µg/mg OC)	Sal (S)	D50 (µm)	Salinity (S)	pH
BW	8	103	4.5	0.14	0.59	8.7	38.4	84.3	< 0.5	1.8	20.0	1.83	10.5	7.60	0.14	36.6	13.1	36.6	7.76
PR	118	232	5.9	1.01	1.56	64.0	102.1	72.9	< 0.5	3.2	72.0	10.04	11.0	9.82	0.26	19.8	31.4	19.8	7.65
CO	183	355	5.39	1.24	2.50	78.8	163.8	73.4	1.5	1.9	111.8	5.15	12.2	-5.01	0.14	35	32	35	7.58
HR	215	491	5.66	1.60	3.05	101.5	199.5	84.8	1.7	3.5	62.1	4.65	4.9	-3.10	0.60	21	21.3	21	7.67
PC	645	668	5.81	3.81	5.44	242.0	355.9	77.9	1.2	2.6	134.3	4.28	9.2	60.38	0.25	35.9	22.9	35.9	8.11
SJ	554	621	5.87	3.98	6.13	253.0	400.9	80.8	4.3	1.6	81.4	4.5	8.9	-6.12	0.16	39.5	22.8	39.5	7.76
MC	792	801	6.13	5.32	6.94	337.9	454.2	74.7	4.2	3.4	57.2	6.46	13.3	16.60	0.24	31.2	27.8	31.2	7.76
HA	862	1261	7.31	5.18	8.13	328.7	531.6	57.0	< 0.5	223.9	270	1.42	30.9	336.53	6.92	21.3	53.8	21.3	7.69
RC	2413	2367	9.48	22.70	23.37	1441.4	1528.4	84.9	0.5	23.6	98.9	3.38	8.5	626.56	2.49	32	23.1	32	7.64
RA	2978	2467	10.06	27.72	21.73	1760.3	1420.9	77.4	< 0.5	130.3	211.2	6.53	7.4	399.34	15.60	6.6	25.5	6.6	7.78
RB	3183	3015	10.45	26.98	25.87	1713.2	1691.9	82.8	< 0.5	62	191.4	2.99	9.3	910.57	6.00	31.3	24.2	31.3	7.70
HB	2565	3753	9.98	4.69	11.76	297.6	769.2	69.5	< 0.5	31.5	913.6	1.64	5.9	173.38	4.57	26.8	3194.2	26.8	7.69

Total Cu, Total Zn, Fe_2O_3, SEMCu/SEMZn (Simultaneously Extracted Metals), AEMCu/AEMZn (Acid-Extractable Metals), LT63 (<63 µm fines), AVS (Acid Volatile Sulphide), PWCu/PWZn (Porewater Metals), TOC (Total Organic Carbon), DOC (Dissolved Organic Carbon), PW_OC_Cu (Dissolved Organic Carbon normalised PWCu), EqPCu (Equilibrium Partitioning Cu), D50 (Median Grain Size), Sal (Salinity), and pH. Site codes as in **Table 2.1.**

Table 2.7 34-site set Key physicochemical characteristics of surface (top 2 cm) sediments sampled in March 2022 for this study.

Site	PW Cu (µg/L)	AEM Cu (µg/g)	D50 (µm)	LT63 (%)	Salinity (S)	PH
VA	9.06	4.4	37.9	61.1	0.9	7.13
VB	9.53	7.8	51.4	59.5	4.1	7.12
VC	13.37	10.8	32.5	70.6	4.9	7.08
TM	7.82	123.0	17.7	86.8	20	7.05
LA	9.22	73.3	32.5	70.2	15.3	6.93
LB	6.01	123.2	20.6	85.2	21.4	6.46
LC	6.54	123.7	15.2	91.1	15.5	6.87
LD	5.95	125.5	20.6	82.4	22.2	7.37
LE	7.33	74.2	11.2	87.4	15.2	6.92
LF	4.84	37.1	27.9	76.2	35.4	7.82
IA	73.18	303.3	24	78.2	30.7	7.45
IB	6.03	311.0	27.9	73.1	26	7.06
MA	16.28	344.7	15.2	87.1	28.4	6.73
MB	5.56	363.6	17.7	84.0	36	6.87
PN	9.01	328.9	17.7	87.4	27.5	6.91
TR	12.92	68.4	37.9	65.3	11.6	6.88
PA	6.54	24.1	37.9	62.9	24.9	7.03
PB	5.89	30.5	32.5	63.8	32.8	6.92
PC	4.59	167.2	24	77.3	38	7.22
JA	9.52	162.5	15.2	75.0	35.7	7.33
JB	8.66	282.9	15.2	89.8	37.4	6.8
TA	8.38	196.2	17.7	87.6	20.9	7.39
TB	5.46	195.7	24	78.4	23.6	7.39
CK	14.4	209.4	44.1	59.3	3.27	7.24
CA	7.67	126.2	13	87.4	31.7	6.78
CB	10.74	103.5	17.7	78.7	19	6.95
RA	55.92	1797.7	20.6	76.5	9.8	6.26
RB	94.95	1754.3	20.6	88.6	13.6	6.66
RC	21.78	1426.5	20.6	80.8	28.8	6.91
RD	25.23	887.3	24	80.0	32.4	6.8
RE	17.32	806.3	37.9	63.4	32.7	6.69
HA	430.71	365.4	44.1	65.4	29.5	6.92
HB	263.64	359.4	51.4	57.1	28.9	6.68
HC	10.69	196.5	27.9	67.3	23.2	7.47

PWCu (Porewater Cu), AEMCu (Acid-Extractable Cu), D50 (Median Grain Size), LT63 (<63 µm Fines), Sal (Salinity), pH (pH). Site codes as in **Table 2.1.**

2.7.2 Sediment Grain Size Distribution

Sediment grain size, characterized by the median particle diameter (D50) and the percentage of fine particles <63 μm (LT63), is a critical factor influencing metal adsorption and retention in sediments (Simpson et al., 2011). The analysis revealed that fine particles dominated most sites, with the percentage of fines ranging from 57.1% at HB (Hayle Estuary) to 91.1% at LC (Helford River) (**Tables 2.6 and 2.7)**. Median grain sizes varied between 11.2 μm at LE (Helford River) and 51.4 μm at HB (Hayle Estuary), corroborating previous surveys (Greenwood, 2001, Shipp and Grant, 2006). The sediments were generally poorly sorted, indicating a wide range of particle sizes, which is typical of estuarine environments where varying hydrodynamic conditions facilitated the accumulation of both fine and coarse materials (Kenny and Sotheran, 2013).

2.2.3 Acid-Extractable Metal and Porewater Copper Concentrations

Acid-extractable metal concentrations (AEM), determined using 1 M HCl, served as proxies for the bioavailable fraction of metals in sediments (Bryan and Langston, 1992). The findings revealed significantly elevated acid-extractable copper (Cu) concentrations at historically contaminated sites such as Restronguet Creek and Hayle Estuary (**Tables 2.6** and **2.7**). For instance, Cu levels reached up to 1,797.7 μg/g at RA and 1,754.3 μg/g at RB in Restronguet Creek, markedly higher than those observed at reference sites like the River Avon (e.g., 4.4 μg/g at VA). Notably, some acid-extractable copper concentrations in the supposedly "clean" Helford River (e.g., LB, LC) were found to be higher than levels recorded in PA, PB and TR within the Fal system. These concentrations were also comparable to those in the Tamar, which is generally regarded as moderately contaminated based on total metal concentrations. These patterns underscored the complexity of metal distribution and the necessity of site-specific assessments to contextualise contamination levels. Such findings further supported observed ecological effects linked to metal contamination in these estuaries (Greenwood, 2001, Rainbow, 2020).

Porewater metal concentrations (PWCu) were critical for assessing the immediate exposure of benthic organisms to metals (Burton, 2018). The highest porewater Cu concentrations were observed at sites with elevated total and acid-extractable metal levels. Porewater Cu concentrations exhibited significant spatial variability across sites. The highest concentrations

were observed at HA (430.7 µg/L) in the Hayle region and RB (95 µg/L) in Restronguet Creek. While sedimentary Cu concentrations were higher in Restronguet Creek (e.g., RB with AEMCu = 1754.3 µg/L) compared to the Hayle (AEMCu = 365.4 µg/L), porewater Cu concentrations were disproportionately higher in the Hayle Estuary. This disparity was likely due to higher metal release rates into the porewater under specific chemical conditions prevalent in the sandy sediments of HA, where metals were notably more bioavailable than elsewhere in the Fal and Hayle estuaries (Grant, 2010). The increased partitioning of metals into the porewater at this site was likely facilitated by factors such as relatively low sediment total organic carbon (TOC) content and coarser grain size (Udochi, 2020). These concentrations exceeded environmental quality standards set by the European Commission (Nugent and Rhinard, 2015), suggesting potential ecological risks to aquatic organisms.

Multiple studies have assessed porewater Cu concentrations in the Fal and Hayle estuaries, revealing consistent trends despite slight variations in reported values. Collectively, these studies demonstrated elevated porewater Cu levels in the Hayle Estuary compared to the Fal Estuary. Greenwood (2001) reported Cu concentrations up to 495.7 µg/L in the River Hayle and 68.2 µg/L in Restronguet Creek. In a subsequent survey using different analytical techniques, Greenwood documented even higher concentrations, reaching 3,378.2 µg/L in the River Hayle and 769.6 µg/L in Copperhouse Pool, while noting 62.2 µg/L in Restronguet Creek. Similarly, Ogilvie and Grant (2008) observed Cu levels up to 783 µg/L in the River Hayle, 463 µg/L in Copperhouse Pool and 27.65 µg/L in Restronguet Creek. Earlier work by Bryan and Gibbs (1983) reported porewater Cu concentrations up to 83 µg/L in the surface sediments of Restronguet Creek. These findings aligned with the current study, confirming higher porewater Cu concentrations in the Hayle Estuary relative to the Fal Estuary. Discrepancies in absolute concentrations across studies may be attributed to differences in analytical methodologies, temporal variations, or environmental factors influencing metal mobility (Udochi, 2020).

In summary, the Fal and Hayle estuaries were heavily contaminated with metals due to historical mining activities. Sediment metal concentrations have remained stable over several decades, establishing a persistent contamination gradient across the study sites. The sediment physicochemical characteristics vary markedly across this gradient, resulting in significant differences in metal bioavailability. These varying attributes, coupled with the absence of other

interfering pollutants, made the Fal and Hayle estuaries ideal sites for investigating the effects of metal contamination on in situ benthic communities.

2.3 Molecular Data Processing

2.3.1 DNA Extraction and Quality Control

Effective molecular analysis relied heavily on the quality and concentration of extracted DNA (Hermans et al., 2018). In this study, DNA extraction protocols were meticulously tailored to accommodate the diverse nature of sediment samples, ensuring optimal concentrations for subsequent PCR amplification. This foundational step is crucial, as the quality of DNA directly affected the success of downstream molecular applications.

2.3.2 Effect of DNA Dilution on PCR Efficiency

The DNA was successfully extracted from sediment samples of varying sizes (0.25 g, 2.7 g and 10 g), as well as from sieved meiofauna samples. Several dilution factors were tested during protocol optimisation to improve amplification efficiency and reduce potential inhibitor effects. The DNA yields varied across samples, with the highest yields obtained from the experimental 10 g sediment extractions up to 405 ng/µL. Dilution factors for PCR amplification were tested accordingly to optimize amplification efficiency and account for potential PCR inhibitors presented in the sediment samples (**Table 2.8).**

Table 2.8 DNA concentration ranges and dilution factors for PCR optimization.

Group	Sample size	Elution volume ul	DNA concentration range µg/l	Dilution for PCR	Amplicon concentration range in PCR reactions µg/l
Pilot	10g sieved	50	0-4	1/10th	0-0.4
Pilot	0.25 g	50	0-70	1/6th	0-7
12-sites	2.7 g	50	0-79	1/6th	0-7.9
34-sites	10 g	100	112-405	1/50th	2.24-8.1
Microcosm	10 g	100	6-28 *	1/3rd	2-9.3

***Not all DNA samples from the experiment were measured for concentration; only 18 random ones were checked for quality control.**

The tailored dilution approach was instrumental in enhancing PCR efficiency across diverse sample types. When low quantities of target DNA fragments were expected, along with high concentrations of co-extracted inhibitors or non-target DNA, diluting the DNA extraction can reduce inhibitor effects, improving amplification efficiency and the overall quality of downstream processes (Deiner et al., 2015, Percze et al., 2024). By adjusting DNA concentrations to fall within the optimal range for PCR amplification, the likelihood of successful and reproducible genetic analyses was significantly increased. For instance, samples with excessively high DNA concentrations, such as those from the 34-sites group, were diluted 1/50th to prevent PCR inhibition, ensuring accurate amplification of target genetic markers. Conversely, samples with lower initial DNA concentrations, including the pilot (Sieved) and Microcosm groups, were diluted accordingly to maintain sufficient DNA availability for PCR without introducing inhibitors (**Table 2.8)**.

The variability in DNA yields underscored the importance of optimizing extraction protocols for environmental samples. PCR inhibition can be caused by either the presence of chemical inhibitors or the ratio of total DNA to that of the targeted DNA (Deiner et al., 2015).

Additionally, PCR inhibition can alter the detection of targeted species extracted from samples (Jiang et al., 2005).

Evaluating and optimizing DNA extraction and amplification techniques is crucial to obtain consistent and reliable data, facilitating accurate identification across a broad spectrum of organisms (Hermans et al., 2018).

2.3.3 Primer Selection and PCR Amplification

After optimizing DNA and PCR protocols, thirteen primer pairs targeting different genetic markers were initially tested to maximize taxonomic coverage. Primers yielded successful amplifications and were selected for further analysis. Some primers showed limited amplification or specificity issues, highlighting the importance of primer selection in metabarcoding studies (Taberlet et al., 2012). More details about Primer Selection results and discussion were provided in **Chapter 3**.

2.3.4 Relationship Between DNA Concentration and Sediment Characteristics

The observed variability in DNA concentrations can be attributed to several factors inhered to the sediment samples. Sites with higher contamination levels, particularly those impacted by historical mining activities, often exhibited lower DNA yields. This reduction was likely due to the toxic effects of heavy metals on microbial communities, thereby decreasing the overall biomass, increasing DNA degradation and reducing DNA availability (Johnson, 1986, Gilbert et al., 2006, Rainbow, 2020). Sediment physicochemical properties, such as total organic carbon (TOC) content and grain size, played a crucial role in influencing DNA extraction efficiency. High TOC and fine-grained sediments hinder DNA recovery by binding DNA to organic material and small particles. In contrast, sandy sediments, which typically have lower TOC, facilitated higher DNA bioavailability, making extraction more efficient. This relationship is key to understanding microbial community structures and their response to environmental factors like metal contamination (Pearman et al., 2020, Niu et al., 2022).

2.3.5 Correlation Between Metal Concentrations and Community Composition

Environmental and molecular data were further integrated to explore broader ecological patterns. Estuarine and coastal environments, particularly in temperate regions, were often affected by industrial pollution, including metal contamination. Multivariate analyses such as non-metric multidimensional scaling (NMDS) and clustered analysis using Bray-Curtis dissimilarity were applied to these ecosystems. These methods demonstrated that microbial and meiofaunal communities clustered based on metal contamination levels and sediment characteristics, emphasizing the influence of pollution on community composition and ecological function in these marine ecosystems (Grant, 2010). These patterns were explored in detail in Chapters 4, 5, and 6, where community composition was compared across contamination gradients using ordination and clustering methods.

Statistical tests highlighted metal concentrations, particularly porewater Cu and Zn, as key predictors of microbial community composition, with higher concentrations resulting in distinct communities compared to less contaminated sites. Grant (2010) discussed these effects within estuarine and coastal ecosystems, while (Udochi, 2020) specifically emphasized the influence of porewater Cu and Zn on community shifts, further confirming the significant role these metals played in shaping microbial diversity across varying contamination levels.

2.3.6 Conclusion

The methodological approach applied in DNA extraction and PCR optimization helped overcome the challenges posed by the heterogeneous nature of estuarine sediments. By implementing tailored dilution strategies, this study improved the consistency of molecular analyses and laid the foundation for assessing microbial diversity in relation to environmental contamination. Sampling effort was supported by accumulation curves presented in **Chapters 4** and **5**, confirming adequate sequence recovery. In addition, a bacterial positive control was used to validate sequencing performance (**Chapter 4**). These approaches collectively support the reliability of the generated data and its relevance to community-level responses in the Fal and Hayle estuaries.

Chapter 3:

Selection and Optimization of 16S, ITS, 18S, 28S and COI Primers for Metabarcoding Studies

3.1 Introduction

3.1.1 Overview of Metabarcoding and Primer Functionality

Metabarcoding is an amplicon-based, high-throughput sequencing approach in which universal primers amplify short, taxonomically informative gene regions from bulk-organism or environmental DNA, the resulting reads are then clustered or denoised and taxonomically assigned with reference databases (Taberlet et al., 2012, Fonseca, 2018). By systematically assessing how these primers capture a broad range of prokaryotic and eukaryotic taxa, this chapter provides insights into which markers most effectively profile marine sediment communities while highlighting design trade-offs in sensitivity, coverage and taxonomic resolution. Achieving robust metabarcoding results typically requires primers that capture a wide taxonomic scope yet amplify DNA regions that are sufficiently variable to distinguish different taxa (Zhang et al., 2020). Technical biases may include primer-template mismatches, taxon-specific amplification efficiency, copy-number variation, and polymerase or GC-content effects, all of which can skew diversity estimates and relative abundances (Fonseca, 2018). To minimise these effects, careful optimisation of PCR conditions and primer selection was essential. Clean amplicons of the correct size must be generated to ensure accurate downstream sequencing (Goldberg et al., 2016, Ruppert et al., 2019).

Bacterial and archaeal communities are commonly surveyed by targeting the 16S rRNA gene, which contains both conserved regions (used for broad-range primer binding) and hypervariable segments (V1-V9) that enable differentiation among species (Wade, 2002, Cole et al., 2005). For Fungal communities the most frequently used primers were the highly variable

ITS region for species-level identification, though incomplete reference data can limit precision (Badotti et al., 2017). For eukaryotes, the 18S SSU and 28S LSU genes were often chosen for phylogenetic and community-level assessments (Sonnenberg et al., 2007, Nyati et al., 2013, Latz et al., 2022, Zimmermann et al., 2024). Metazoan studies typically employed the mitochondrial cytochrome-c oxidase subunit I (COI) gene, which offers high resolution for animal taxa but may suffer from amplification biases and limited reference coverage for some groups (Timm et al., 2022, Antil et al., 2023). The regions targeted by these primers were illustrated in **Fig. 3.1.** Employing multiple loci allowed simultaneous profiling of prokaryotic and eukaryotic communities and provided cross-validation of ambiguous assignments (Taberlet et al., 2012). The emphasis placed on individual loci reflects whether the focus is on abundant bacteria, fungi or larger metazoans (Bik et al., 2012).

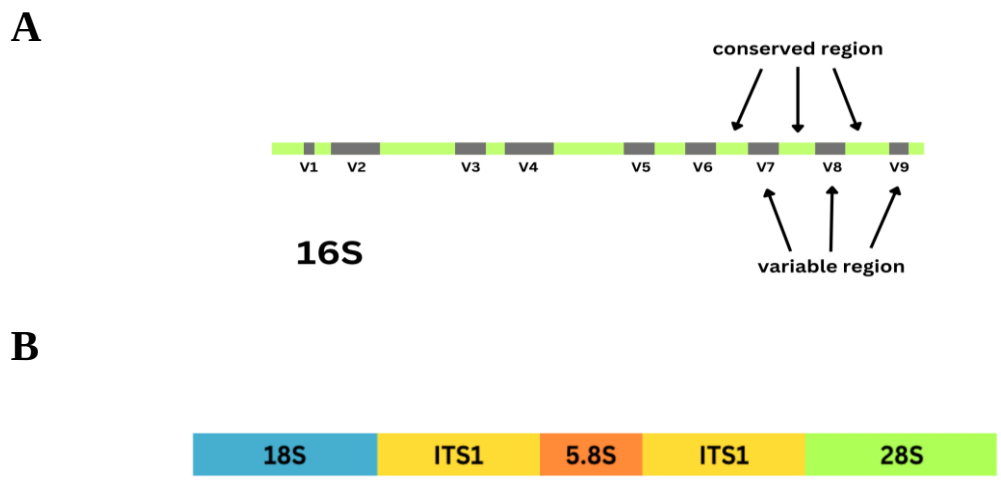


Fig. 3.1 Primer-targeted regions for gene amplification. **A)** The 16S rRNA gene in prokaryotes, highlighting variable and conserved regions modified from Cox et al. (2013). **B)** The rRNA operon in eukaryotes, including the 18S, ITS1, 5.8S, ITS2 and 28S gene regions.

Given the different primer binding requirements across taxonomic groups (e.g., conserved vs. variable regions, amplicon length), the choice and optimisation of primers for metabarcoding can profoundly shape DNA-based survey outcomes. By carefully selecting and fine-tuning primers, researchers can maximise diversity detection and thus more confidently interpret how microbial or metazoan communities respond to ecological drivers, including pollution. In this

study, emphasis was placed on evaluating and optimising primers for coverage and diversity and relative abundance rather than absolute abundance, aiming to capture the widest possible spectrum of taxa in marine sediments.

3.1.2 Selection of Primers for Evaluation

This study targeted multiple taxonomic groups, including Bacteria, Archaea, Fungi, single-celled eukaryotes and Metazoa, with a particular emphasis on nematodes. This coverage of a wide range of taxonomic groups facilitated a more comprehensive assessment of ecosystem biodiversity than approaches based on morphological identification (Tytgat et al., 2019, Fonseca et al., 2022). Meiofauna, operationally defined as metazoans passing through a 1 mm mesh but retained on a 63 µm mesh (Heip et al., 1985), have traditionally been analysed separately in existing literature (Somerfield et al., 1994b). This separation complicated direct comparisons with the current approach, although the literature making use of meiofauna in environmental monitoring has focused on nematodes which, with the exception of juveniles, were entirely within the meiofaunal size range (Somerfield et al., 1994a, Millward and Grant, 2000, Pawlowski et al., 2024), the differing extraction protocols, DNA yields and sequencing depths can confound diversity estimates across studies (Bik et al., 2012, Creer et al., 2016). In evaluating the primers employed for PCR amplification, each primer pair was examined based on its original description and bench-marked for (i) taxonomic coverage, (ii) amplification efficiency, (iii) known primer bias and (iv) complementarity with other markers.

Ideally, PCR primers should be free of taxonomic biases within their target group, however, for most eukaryotic taxa, relative read abundance cannot reliably represent actual organismal abundance due to multicopy genes, variable biomass, and amplification bias. This assumption may hold more closely for unicellular groups like bacteria, but not for metazoans or other multicellular taxa. Following amplification, the PCR products were sequenced to determine the taxonomic coverage of each primer pair, which was then compared with the intended target range and with community profiles obtained from alternative primer sets and RNA-Seq data. Applying these criteria provided a transparent framework for selecting primer combinations that maximised diversity detection while limiting bias (Pawlowski et al., 2014, Elbrecht and Leese, 2015).

The 16S rRNA 515F/926R primer pair, originally designed by Turner et al. (1999) to target small subunit rRNA sequences for investigating phylogenetic relationships among cyanobacteria and plastids, has been extensively utilized in subsequent studies to examine wider bacterial diversity (Caporaso et al., 2011). The forward primer (515F) incorporated a modification described by Parada et al. (2016) to improve coverage of marine microbiomes.

Bahram et al. (2019) developed the 16S rRNA SSU1ArF/SSU520R primer pair to provide more comprehensive coverage of the major Archaeal lineages. Bahram's rationale was to improve the identification and quantification of Archaea by designing primers that efficiently amplified archaeal 16S rRNA genes, thus overcoming the limitations of previously available primers which often failed to cover a broad spectrum of archaeal lineages. Subsequently, Martin-Pozas et al. (2023) utilized the SSU1ArF/SSU520R primer pair in their study to simultaneously analyze bacterial, archaeal and eukaryotic communities within various environmental samples. They found that the primer efficiently amplified archaeal 16S rRNA genes, providing representation of archaeal taxa across diverse environmental conditions. This primer successfully captured major archaeal groups, including both common and rare lineages such as Euryarchaeota and Thaumarchaeota, demonstrating its effectiveness for broad-scale environmental and ecological studies. Martin-Pozas et al. (2023) concluded that the SSU1ArF/SSU520R primer showed better performance than previous universal primers by providing more inclusive coverage and improving the detection of archaeal diversity, making it a valuable primer set for comprehensive microbial community assessments.

The ITS region of the rRNA gene was targeted using the ITS1f12/ITS2 primer pair, first described by Schmidt et al. (2013) and White et al. (1990), was designed to amplify the fungal ITS1 region within the ribosomal RNA operon. This primer is highly specific to fungi, especially basidiomycetes and was commonly used in fungal community studies and metabarcoding. Multiple studies have confirmed the effectiveness of ITS1 primers for fungal identification and diversity studies. Rajkowska et al. (2023) and Szulc et al. (2017) successfully employed ITS primers ITS1F12/ITS2 in their study of fungal diversity, facilitating the identification of various fungal genera and contributing to the understanding of microbial diversity in different environments. These primers were also used by the Earth Microbiome Project, demonstrating their reliability and applicability in comprehensive environmental studies (Smith et al., 2018, Chrismas et al., 2023) Similarly, Harnelly et al. (2022) demonstrated

the utility of ITS1f/ITS2 primers for species-level identification of macroscopic fungi, emphasizing the advantages of DNA barcoding over traditional morphological methods. These findings underscored the efficacy of ITS primers in fungal community analyses and supported their continued use in microbial diversity studies. However, Walters et al. (2015) and Tedersoo et al. (2022) demonstrated that modified fungal-specific versions of the ITS1f/ITS2 primers, which extended the sequencing primers into the amplicon region, resulted in approximately doubled yields of reads clustering against the UNITE fungal database. These modifications enhanced specificity and detection of diverse fungal taxa, especially in low-diversity samples. However, they also reduced the detection of Ascomycetes, a major group, potentially leading to an underrepresentation of fungal diversity.

The ITS region was targeted using the VRAIN2F/VRAIN2R primers, introduced by (Vrain et al., 1992), were developed to amplify the ITS region of nematodes. This method has been employed in phylogenetic studies to distinguish nematode species and has been compared with other primers for its accuracy and specificity in nematode taxonomy studies (Derycke et al., 2010, Nguyen et al., 2021, Hajihassani et al., 2023).

For eukaryote detection, Wang et al. (2014) emphasized the importance of evaluating and optimizing primers to ensure effective coverage of various eukaryotic groups, highlighting that proper primer selection is essential for accurately detecting and assessing eukaryotic diversity in environmental samples. The 18S rRNA gene (V1–V2 regions) was amplified using the G18S4/22R primer pair has been applied in metabarcoding marine nematode communities. Originally designed by Blaxter et al. (1998), these primers have been used in DNA barcoding to assess nematode diversity. In a study on marine nematode communities, Tytgat et al. (2019) they were combined with COI markers amplified using the JB3/JB5 primer pair to enhance species-level resolution. This combination allowed for a comprehensive assessment of biodiversity and the impact of environmental stressors on nematode populations across various ecosystems. Together, they offered strong specificity and coverage for nematode barcoding, supporting detailed biodiversity assessments in marine environments (Derycke et al., 2005, Tytgat et al., 2019).

The 18S rRNA gene was targeted using the TAReuk454FWD1/TAReukREV3 primer pair, originally described by (Stoeck et al., 2010) and developed for the Tara Oceans project (Delage

et al., 2023). It has been widely used to target the V4 region of the 18S rRNA gene for metabarcoding studies of eukaryotic microorganisms across a broad range of environments, including soil, marine and freshwater ecosystems. Fonseca et al. (2022) employed this primer set in parallel with other ribosomal 16S rRNA and mitochondrial COI markers to examine microbial and metazoan diversity within the Antarctic benthic environment.

The 18S rRNA gene (V4 region) was amplified using the D512/D978 primer pair have primarily been used for identifying diatoms and other eukaryotic microorganisms in environmental samples. Introduced by Zimmermann et al. (2011), these primers have been widely adopted for barcoding diatoms, offering a good balance of taxonomic resolution and coverage in molecular studies such as (Luddington et al., 2012) and (Rivera-Garcia et al., 2018). Another study by Kim et al. (2017) highlighted the strong performance of the D512/D978 primers in amplifying diverse eukaryotic groups, including Metazoans, Protists and Fungi, though smaller eukaryotes <63 μm were excluded. However, it also revealed a high proportion of unclassified eukaryotes, suggesting potential gaps in the available reference databases.

Originally developed by (Lane, 1991) and (Medlin et al., 1988), the 18S rRNA gene (V9 region) was targeted using the 1391F/EukBr primer pair for examining microbial eukaryotes in marine environments (Stoeck et al., 2010). It is highly conserved forward primer (1391f) that could yield a broader range of higher-level taxa than exclusively eukaryote-specific sets, albeit with some risk of amplifying non-eukaryotic sequences (Amaral-Zettler et al., 2009, Stoeck et al., 2010). Comparisons with V4-specific primers suggested that while 1391f/EukBr provided broad taxonomic coverage, the V4 region may be more suitable for lineages requiring finer resolution (Stoeck et al., 2010, Liu et al., 2019).

The 18S rRNA gene (V6–V8 regions) was amplified using the Nemf/18Sr2b primer pair, focusing specifically on nematodes and other metazoans. It was designed by Sapkota and Nicolaisen (2015) to exclude plant and fungal DNA, enabling more focused amplification of nematode DNA in environmental samples. According to the same source, this primer set has been widely applied in studies involving nematode communities across various habitats, including soil and has shown a high efficiency in metabarcoding without requiring nematode enrichment before PCR. By reducing biases introduced by enrichment, as noted by the same

authors, it allowed for the recovery of sequences from a wide range of nematode taxa, making it particularly effective for soil and root-associated nematodes, capturing a broad spectrum of species without the need for labour-intensive enrichment processes. The primer has been compared to other primer sets, such as JB3/JB5 and demonstrated superior performance in nematode detection, generating a higher proportion of nematode-specific reads (Sapkota and Nicolaisen, 2015, Sikder et al., 2020).

The D4-D5 region of the 28S rRNA gene is a variable region that offers notable taxonomic advantages. Unlike the more conserved 18S rRNA gene, the 28S rRNA evolves at a faster rate than 18S rRNA, enabling better species discrimination and precise phylogenetic assessments, making it valuable for biodiversity studies and environmental DNA analysis (Machida and Knowlton, 2012). The DM568F/RM3R primer pair targeted this region and was specifically designed for eukaryotic microbial analysis, including nematode detection. A 2019 study by (Kounosu et al.) demonstrated its effectiveness in amplifying nematode species while avoiding bacterial contamination, providing high taxonomic resolution and it is particularly useful for distinguishing nematode genera and species in complex environmental samples.

3.1.3 Optimizing DNA Extraction and Determining Appropriate Sample Size

Marine benthic micro- and meiofaunal communities within seabed sediments underpin key ecosystem processes and are recognised as sensitive bio-indicators of environmental change (Giere, 2009). Traditional morphological identification methods for aquatic macroinvertebrates can be labour-intensive and subject to inconsistencies, whereas DNA metabarcoding offers a more rapid, standardised alternative that often detects a greater number of unique taxa (Hajibabaei et al., 2011). However, morphological and molecular approaches frequently capture different aspects of biodiversity, indicating that a combination of both methods provided complementary insights for a more comprehensive assessment of aquatic ecosystems (Emmons et al., 2023, Schuijt et al., 2024).

Nevertheless, variation in DNA extraction methods and sample size selection can affect the accuracy and reproducibility of metabarcoding studies. Gielings et al. (2021) reviewed the application of metabarcoding for marine meiofauna and highlighted inconsistencies in sample collection, choice of genetic markers (with 18S rRNA often selected over COI), primer

selection, sequencing platforms and bioinformatic workflows, all of which impaired cross-study comparisons. The present study focused on utilising multiple genetic markers to improve taxon coverage, optimising DNA extraction methods and determining appropriate sample sizes for metabarcoding marine benthic meiofauna.

3.2 Methods

3.2.1 Primer Design for Illumina Metabarcoding

The design strategy adopted from (Fadrosh et al., 2014) involved utilizing forward and reversed primers consisting of an Illumina adaptor; a 8-base barcode, 0-7 bases of length heterogeneity spacers and a gene-specific primer as in **Fig. 3.2**. The heterogeneity spacer prevented all the gene-specific primers being read in phase during sequencing. This method enabled multiplexing of multiple PCR products, allowing for the accurate and reliable amplification of targeted genetic markers across numerous samples (Hamady et al., 2008, Kozich et al., 2013). For ease of identification, primers will be identified either by the name of the forward primer or by an abbreviation of this name, particularly when it is very long.

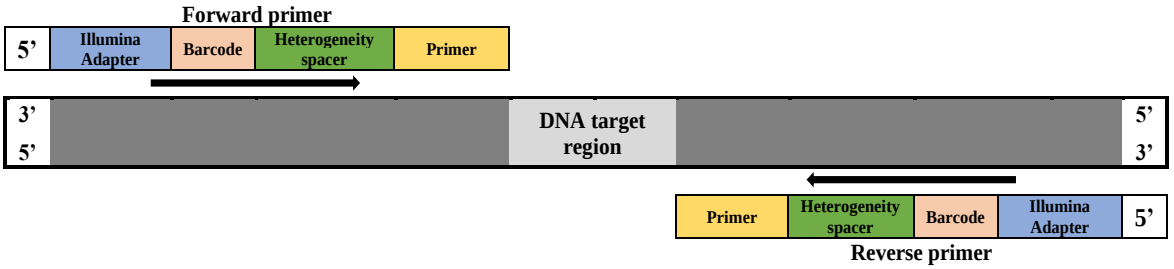


Fig. 3.2 Schematic diagram of the primer design for DNA metabarcoding, including Illumina adapters, barcodes, heterogeneity spacers and primers. The DNA target region is bordered by forward and reverse primers, each containing the necessary components for sequencing and identification.

3.2.2 PCR Amplification and sequencing

Polymerase chain reactions (PCRs) were conducted using barcoded primers on a total of 1 to 245 DNA samples, encompassing all possible sets, including Pilot samples, 12-site samples, 34-site samples and experimental sets, as detailed in **Chapter 2**. It was conducted in a volume of 20 µl, including 10 µl of PhusionTM Flash master mix (Thermo Scientific, UK), 1 µl of forward primer and 1 µl of reverse primer (final concentration of primers: 100 µM), 7 µl of ultrapure sterile water (MilliQ water) and 1 µl of DNA (with a concentration between trace amounts and 9.3 ng). The amplification of the DNA templates was carried out in either a Veriti™ HID 96-Well Thermal Cycler, 0.2mL system (Applied Biosystems, UK) or 0.2 ml PCR tubes, depending upon the number of samples be amplified.

The amplification consisted of an initial denaturation at 98°C for 10 minutes, followed by 28 to 35 cycles of denaturation at 98°C for 30 seconds; annealing at a primer specific temperature for 30 seconds and extension at 72°C for 30 seconds. This was followed by a final extension at 72°C for 5 minutes, before being held at 4°C. The PCR success varied between different sediments and annealing temperatures and cycle numbers were adjusted to optimise amplification yield and specificity. The final values used of annealing temperature and cycle numbers were given in **Table 3.1**.

Table 3.1 Final PCR conditions for the thirteen primers, listed by their abbreviated names. The 1st set includes 0.25g and 2.7g whole sediment samples, along with 10g sieved extractions, while the 2nd and 3rd sets utilized 10g whole sediment samples.

Primer	Final PCR conditions			
	1st set		2nd/3rd sets	
	Annealing Temp °C	Cycles	Annealing Temp °C	Cycles
515F	67	35x	67	35x
ARF	50	35x	50	35x
ITS	52	35x	52	35x
VRAIN	67	35x	-	-
E1391	68 Hot Start fusion Master Mix	35x	-	-
D512	51	35x	-	-
TAR	55	33x	54	35x
G18S	63	28x Then 33x for weak ones	63	33x
NEM	61	32x Then 35x for weak ones	-	-
DM568	68	35x	65	35x
JB3	1st PCR Non coded 50 2nd PCR coded 65	1st PCR Non coded-33x 2nd PCR coded 40x	1st PCR Non coded 50 2nd PCR coded 60	1st PCR Non coded-10x 2nd PCR coded 30x
JB2	50-62	35 and mlCO protocol*	-	-
mlCOI	50-62	35 and mlCO protocol*	-	-

The primer pairs 515F and ITS were well established within our laboratory, demonstrating consistent success in amplification and subsequent sequencing. Because these were already known to work well, the full set of 12-site dataset's samples were amplified with barcoded versions at the initial run. However, the other primer pairs, despite their potential, required further testing and optimization to achieve similar levels of efficacy and reliability.

For the COI JB3 primer pair, it was observed that while the non-barcoded version successfully amplified the target COI region, the coded primers failed to produce any amplification despite optimization efforts. To address this issue, a two-step PCR protocol was implemented. In the first round, non-coded primers were used to amplify the COI region from genomic DNA. The resulting PCR products then served as templates for a second round of PCR using the coded primers (conditions detailed in **Table 3.1**). This strategy allowed the coded primers to bind more efficiently to the simpler PCR products rather than the complex genomic DNA, effectively incorporating the codes into the final amplicons and resolving the initial amplification problem. The COI primer pairs JB2 and mlCOI followed a PCR protocol that involved an initial phase of 16 cycles with denaturation at 95°C for 10 seconds, annealing at 62°C for 30 seconds (decreasing by 1°C per cycle) and extension at 72°C for 60 seconds. This was followed by 25 additional cycles with a constant annealing temperature of 46°C, as described by (Leray et al., 2013). All primers were supplied by SigmaAldrich Company limited. The purification, quantification of PCR products and sequence analysis were conducted following the protocols outlined in **Chapter 2**. The PCR products from multiple samples were pooled in equimolar amounts to create a single composite sample. This initial pool of the pilot samples was sent to the Earlham Institute in Norwich, UK, for sequencing on a single lane of the Illumina MiSeq, using a pre-made library for 300 bp paired-end sequencing, followed by NovaSeq 6000 flow cell with 250 bp paired-end for the 12 and 34 sets then NextSeq 1000 with 300 bp paired-end for the experiment set of samples. The ASV abundance was initially assessed using Mothur software (version 1.45.3; www.mothur.org) (Schloss et al., 2016) and USEARCH (version 11.0.667; www.drive5.com/usearch) (Edgar, 2010). The data were subsequently re-evaluated with updated taxonomic tools and methods via the LotuS2 pipeline (Özkurt et al., 2022). Taxonomic assignment for bacteria and Archaea was conducted using KSGP version 1.0 (Grant et al., 2023) with the GTDB taxonomy hierarchy, while Eukaryote assignments utilised Eukaryome version 1.7 (www.eukaryome.org) (Tedersoo et al., 2024), which specifically covered the 18S, ITS and 28S markers. COI assignments employed

MIDORI-Longest (www.reference-midori.info) (Leray et al., 2018). Detailed methods were described in **Chapter 2**. Numbers of sequencing reads for each amplicon were shown in **Table 2.5**.

3.3 Results:

3.3.1 PCR Amplification

Gel electrophoresis was employed to visualise the PCR products, and confirm amplification success of the target DNA regions. This technique facilitated the assessment of product size and purity, confirming the presence of specific amplicons and identifying any non-specific amplification (Arunachalam et al., 2021). The success of PCR amplifications is summarised in **Table 3.1**. Non-specific products that were shorter than the target can be removed by the magnetic bead clean-up process which described in **Chapter 2**. Additionally, larger non-specific fragments can be eliminated during the bioinformatics step using size filtering, which excluded sequences that fall outside the expected length range.

Amplification success using PCR was verified by gel electrophoresis and was summarised in **Table 3.2**. Most markers amplified cleanly after minor optimisation: the ARF, VRAIN, E1391, D512, TAR, G18S and DM568 primers consistently produced single products of the expected size, and 515F primer was usable once large non-specific fragments were removed during magnetic-bead clean-up. The ITS primer displayed greater size heterogeneity, typical of this locus, but a dominant target band became evident once cycling conditions were refined. The JB3 primer required a two-step PCR; this protocol worked well for the 12- and 34-site sets but proved unreliable for the experimental set, so those data were excluded. Primers JB2 and mlCOI did not amplify under any tested conditions and were therefore omitted from downstream sequencing. Overall, the optimised primer panel provided robust amplification coverage for bacteria, archaea and the majority of eukaryotic targets while transparently flagging loci that remained problematic.

Table 3.2 Overview of amplification results for the 13 primer pairs, summarising amplification quality and success across all sequencing runs.

Primer	Amplification Summary
515F	Non-specific large products observed; successful amplification in most samples
ARF	Single clear product consistently observed
ITS	High variability; improved results with main band visible in later runs
VRAIN	Single clear product in initial set; not tested further
E1391	Clear product observed with hot start mix
D512	Single clear product consistently observed
TAR	Clear product; few samples showed short non-specific bands
G18S	Effective amplification; distinct bands visible even in weak samples
NEM	Single clear product; no further testing in later sets
DM568	Consistently produced correct-sized product
JB3	Initially poor specificity; improved with 2-step PCR in 12- and 34-site sets; inconsistent in experimental set
JB2	No amplification detected
mlCOI	No amplification detected

3.3.2 Primer Performance and Taxonomic Classification

The relative abundance of microbial communities amplified by two prokaryotic primers; 515F (**Fig. 3.3 A)** and ARF (**Fig. 3.3 B**), with bacteria and Archaea being dominant at the domain level as targeted, followed by smaller proportions of Archaea and Eukaryota and a minor portion unclassified by the LCA approach used.

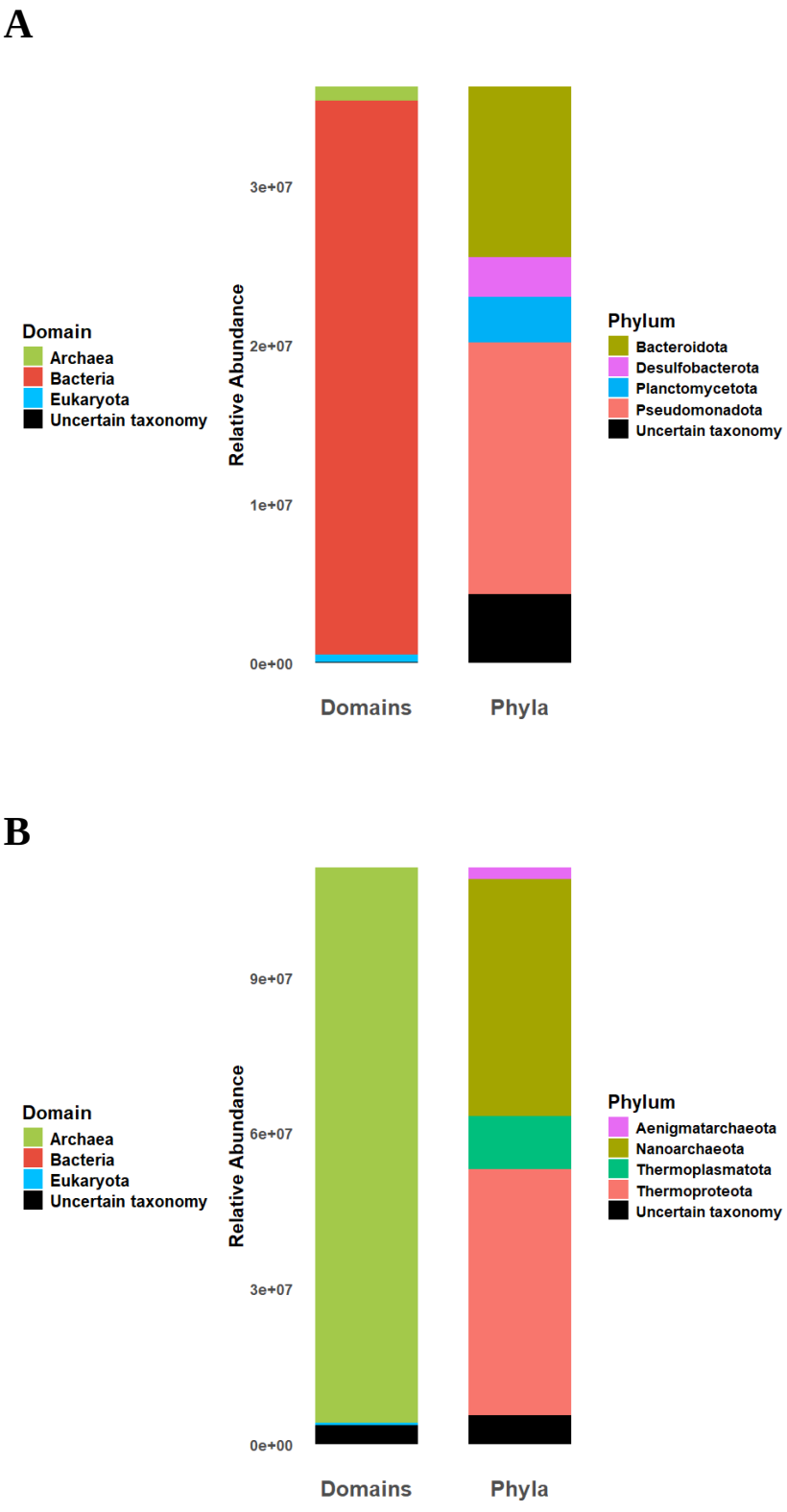


Fig. 3.3 Relative abundance of domains and the most abundant phyla for; **A)** 515F and **B)** ARF prokaryotic primers. Data derived from for all datasets and processed using LotuS2/KSPG. "Uncertain taxonomy" refers to assignments that remain unresolved at the domain or phylum level.

Prokaryote databases such as SILVA and GTDB classified organisms using a taxonomic hierarchy in which the two domains of Bacteria and Archaea were divided into phyla with no intermediated layers. In contrast, eukaryote databases used one or more taxonomic levels between domain and phylum (Burki et al., 2020, Vaulot et al., 2022, Tedersoo et al., 2024). Taxonomic composition of sequences for the eukaryotic 18S, ITS and 28S primers was analysed at various taxonomic levels using the standardized Eukaryome database. This categorized organisms into distinct groups starting at the "kingdom" level and ensured consistency in data interpretation across taxonomic ranks, facilitating reliable and comparable analyses (Tedersoo et al., 2024). For a summary of the most abundant “kingdoms”, refered to **Table 3.3**.

Table 3.3 Top 10 Eukaryotic Kingdoms, as ranked and defined in the Eukaryome database (version 1.7, www.eukaryome.org), according to the number of detected Amplicon Sequence Variants (ASVs).

Metazoa	Viridiplantae	Fungi	Stramenopila	Alveolata
Rhodoplantae	Rhizaria	Amoebozoa	Euglenozoa	Other kingdoms

The dominant kingdoms detected by each primer were illustrated in **Fig. 3.4**, showedvariation in community composition based on relative abundance. Primers that mainly targeted metazoans, such as DM568, G18S, NEM and TAR, recovered Metazoa as the dominant group, whereas Stramenopile prevailed in the fungal primers ITS and VRAINStramenopila. Primer JB3 predominantly detected Eukaryota, identified as Eukaryota_2759 in the MIDORI-Longest database, as the dominant kingdom. For a substantial number of sequences amplified with the ITS, NEM and VRAIN primers, taxonomy was not resolved at the kingdom level. The top phyla detected for each primer as follows: Bacillariophyta (diatoms) dominated in D512 and ITS, Rozellomycota (fungi) in VRAIN, Mollusca in E1391, Annelida in TAR, Arthropoda in DM568 and Nematoda in G18S, JB3 and NEM (**Fig. 3.5)**. Nematoda were also present in relatively large proportions in DM568, TAR and ITS, though they did not dominate. Additionally, for a substantial number of sequences amplified with JB3 and VRAIN primers, taxonomy was not resolved at the phylum level, indicating sequences that could not be confidently classified (Sapkota and Nicolaisen, 2015, Macheriotou et al., 2019, Barrenechea

Angeles et al., 2024). This could reflect gaps in reference databases, the detection of uncharacterised taxa, or the presence of sequencing-PCR artefacts. This variation across primers highlighted their distinct taxonomic specificity. Each primer set targets a subset of eukaryotes and none served as comprehensive "universal" Eukaryote or metazoan primers, a point revisited in **Chapter 6** when comparing metabarcoding with metatranscriptomic data. The number of reads per sample varied substantially, ranging from a minimum of 1.3K reads for the VRAIN to a maximum of 726.4K reads for the TAR primer (Error! Reference source n ot found.**)**.

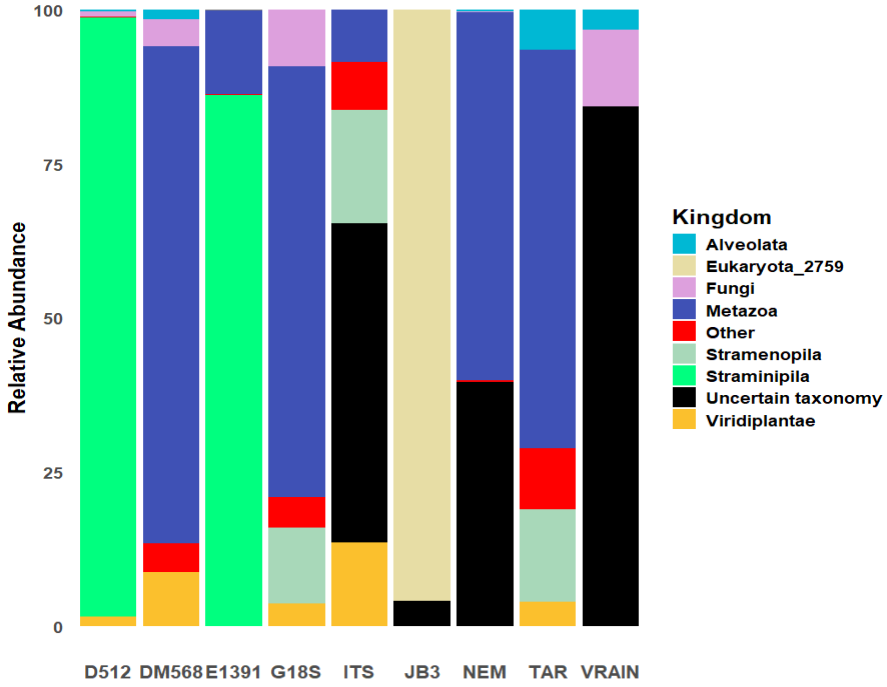


Fig. 3.4 Relative abundance of the five most abundant kingdoms. Data were processed using the LotuS2 pipeline using the MIDORI-Longest database for COI JB3 and the Eukaryome database for all others. The D512, E1391 and Vrain datasets were obtained from pilot samples, while G18S, ITS and TAR data were derived from all datasets. DM568 (34-sites and exp), NEM (pilot and 12-sites) and JB3 (pilot, 12 and 34 sites). "Uncertain taxonomy" refers to assignments that remain unresolved at the kingdom level.

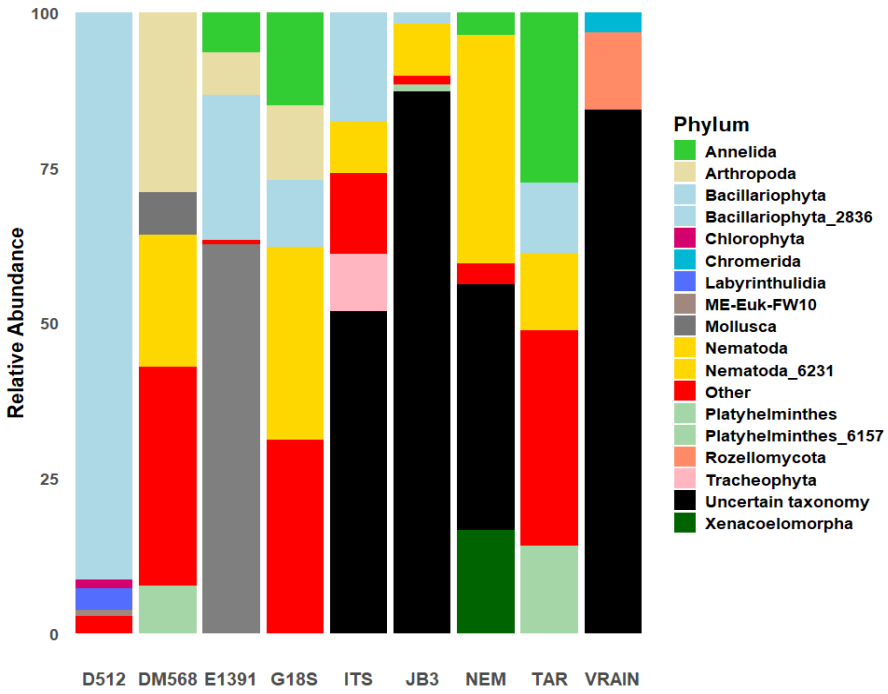


Fig. 3.5 Relative abundance of the five most abundant phyla. Data were processed using the LotuS2 pipeline using the MIDORI-Longest database for COI JB3 and the Eukaryome database for all others. The D512, E1391 and Vrain datasets were obtained from pilot samples, while G18S, ITS and TAR data were derived from all datasets. DM568 (34-sites and exp), NEM (pilot and 12-sites) and JB3 (pilot, 12 and 34 sites). "Uncertain taxonomy" refers to assignments that remain unresolved at the phylum level.

Table 3.4 Overview of sequencing performance for each primer pair from pilot samples. All data processed using the LotuS2 pipeline.

Primer pair	Average Reads/sample	Average ASVs/sample	Sequencing platform
515F	36.1K	3982	MiSeq
ARF	493.9K	963	NovaSeq
ITS	40.5K	884	MiSeq
VRAIN	1.3K	8	MiSeq
G18S	698.3K	3426	NovaSeq
TAR	726.4K	1037	NovaSeq
D512	90.9K	256	MiSeq
NEM	32.3K	32	MiSeq
E1391	56.3K	21	MiSeq
DM568	8.9K	55	MiSeq
JB3	233.3K	32	MiSeq

The 515F primer generated an average of 36.1K reads and 3982 ASVs per sample. Although primer 515F was specifically designed to target bacterial 16S rRNA sequences, amplification resulted in 96.2% bacterial reads, with 2.4% classified as Archaea and 1.3% as Eukaryotes. The Pseudomonadota phylum, also known as Proteobacteria, made up the largest proportion of reads, followed by Bacteroidota and Planctomycetota (**Fig. 3.6)**. Further details on the additional taxonomic analysis were provided in **Chapter 4**.

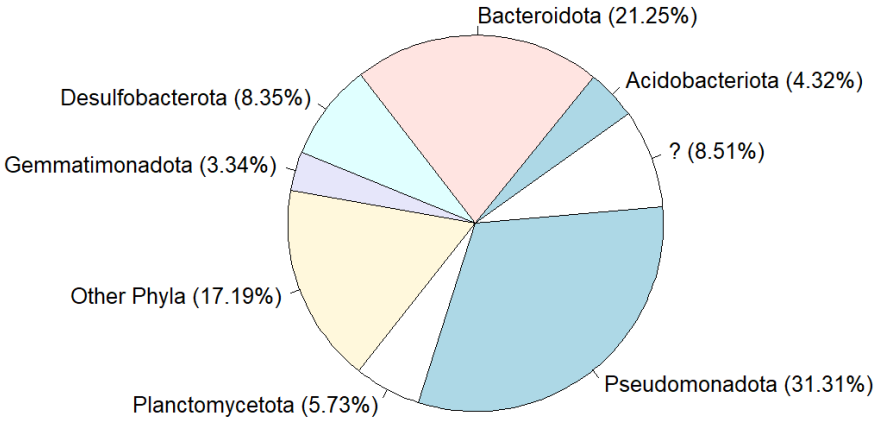


Fig. 3.6 Distribution of reads across phyla for all datasets, obtained using the **515F** primer pair and based on the total number of reads. Taxonomic assignments were made using the KSPG database, so follow the GTDB taxonomy. For clarity, Phyla contributing under 2.2% each are grouped as "other phyla," totalling 17.19% of abundance. '?' indicates unresolved taxonomy at this level by LCA classification using KSGP matches.

The primer pair ARF was originally designed to target 16S sequences from Archaea (Bahram et al., 2019) and generated an average of 493.9K reads and 963 OTUs per sample. At the domain level, 96.3% of ARF reads were classified as Archaea, 3.1% remained unidentified to any taxonomic group and 0.6% were assigned to other domains. The dominant phyla identified were Thermoproteota, Nanoarchaeota and Thermoplasmatota, accounting for 41.5%, 39.9% and 8.9% of the total taxonomic composition, respectively **(Fig. 3.7)**. This analysis excluded a substantial number of reads without matches in the KSPG database, which likely represented PCR artefacts. Further details on these exclusions, along with additional taxonomic analysis, were provided in **Chapter 5**.

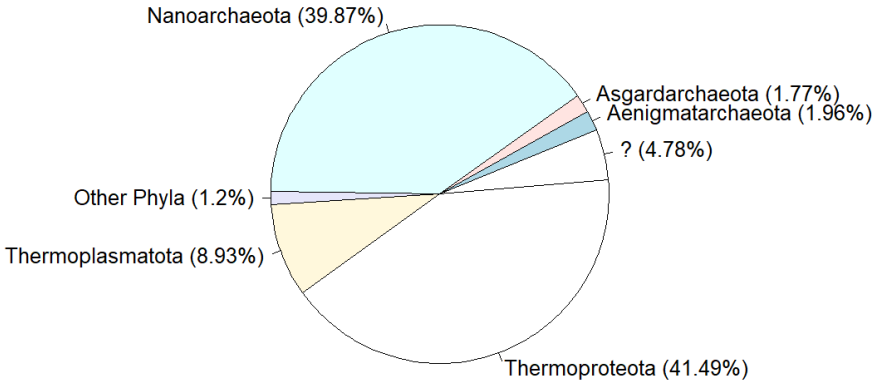
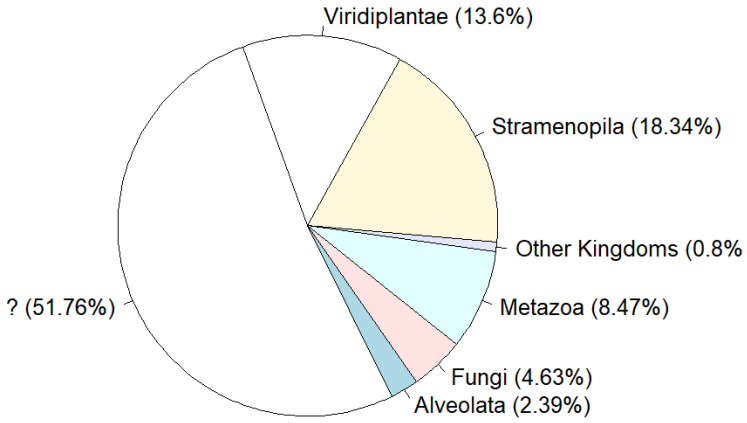


Fig. 3.7 Distribution of reads across phyla for all datasets, based on the total number of reads derived from the **ARF** primer. Taxonomic assignments were made using the KSPG database, so follow the GTDB taxonomy. For clarity, Phyla contributing under 0.7% each are grouped as "other phyla," totalling 1.2% of abundance. '?' indicates unresolved taxonomy at this level by LCA classification using KSGP matches.

The Internal Transcribed Spacer (ITS) primer pair generated an average of 40.5K reads and 884 ASVs per sample. The primer pair was originally designed to target ITS sequences from fungi (Schmidt et al., 2013). However, at the level of kingdom only 4.6% of reads were assigned to fungi (**Fig. 3.8 A)**. The other main taxonomic groups identified were Stramenopila 18.3%, Viridiplantae 13.6% and Metazoa 8.5%. Around half of the reads did not have hits in the ITS section of Eukaryome database. When classified at phylum level, 8.2% of the reads were attributed to Nematoda (**Fig. 3.8 B)**.

A



B

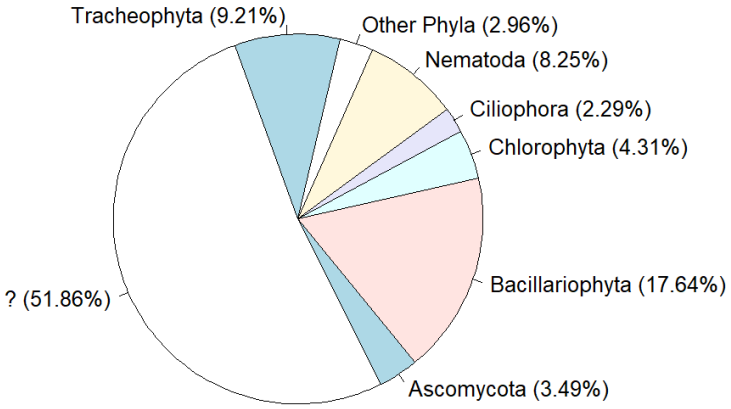


Fig. 3.8 Distribution of reads across **A)** Kingdoms and **B)** phyla for all datasets, based on the total number of reads derived from the **ITS** primer. Taxonomic assignments were made using the Eukaryome database. For clarity, Phyla contributing under 1.4 % each are grouped as "other phyla," totalling 3% of abundance. '?' indicates unresolved taxonomy at this level by LCA classification using Eukaryome matches.

The VRAIN primer pair was originally designed to target ITS sequences from Eukaryotes (Vrain et al., 1992). It generated between 683 and 1917 reads for two samples, but these represented only 7 and 9 ASVs respectively which were mainly composed of fungi 13.2% and Alveolata 5.3%. 81% of the reads were unclassified (**Fig. 3.9 A** and **B)**.

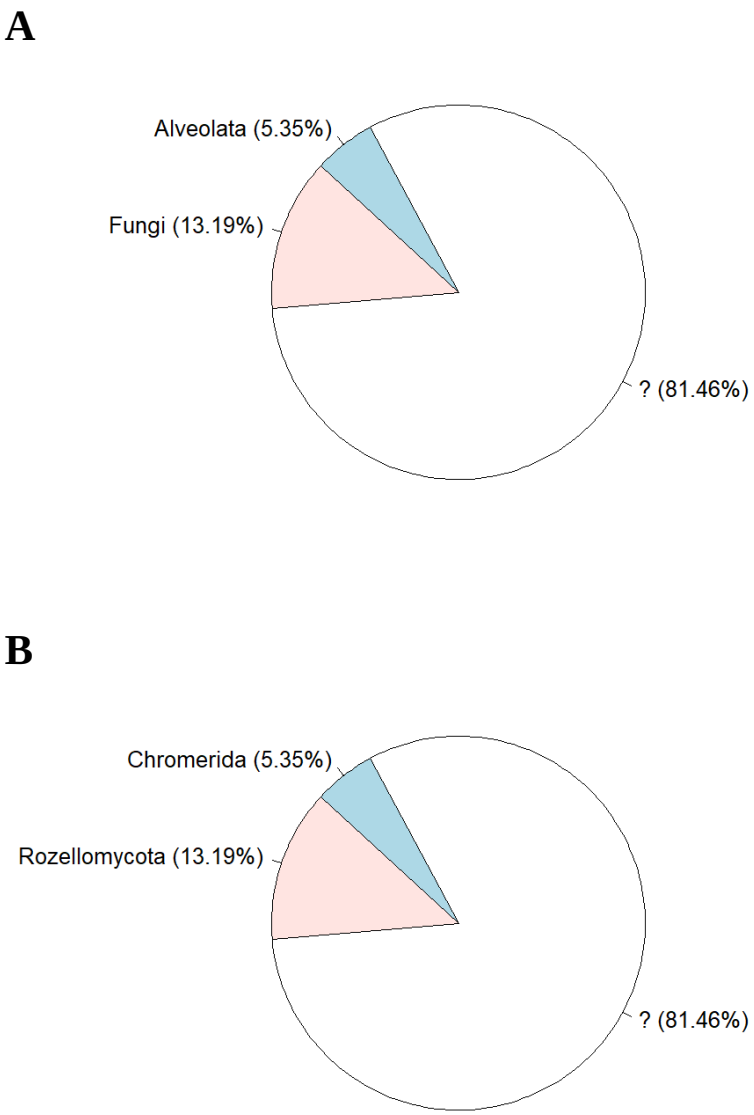


Fig. 3.9 Distribution of reads across **A)** Kingdoms and **B)** phyla for the pilot dataset, based on the total number of reads derived from the **VRAIN** primer. Taxonomic assignments were made using the Eukaryome database. '?' indicates that taxonomy is not resolved at this level by LCA classification based on matches in Eukaryome.

On average, the G18S primer pair yielded 698.3K reads and 3,426 ASVs per sample. Metazoa accounted for 70% of the reads (**Fig. 3.10 A)**. The proportion of ASVs in each kingdom (**Fig. 3.10 B)** showed some variation compared to reads, with Metazoa making up 54.7% of ASVs and Stramenopila at 9.6%, despite constituting 12.4% of the reads. Fungi exhibited similar proportions in both reads and ASVs. Rhizaria contribute 9% of the ASVs, although they

represented only 2.4% of the reads, suggesting higher diversity in this group. At the phylum level, Nematoda were the most abundant, comprising 31.1% of reads (**Fig. 3.11 A)**. However, the proportion of ASVs in each phylum (**Fig. 3.11 B)** differed, with Nematoda representing only 13.6% of ASVs, indicating lower species diversity despite their abundance. Arthropoda, conversely, made up 12% of reads but accounted for 22.2% of ASVs, reflecting greater taxonomic diversity. Platyhelminthes showed similar proportions in both reads and ASVs, while Cercozoa accounted for 8.1% of ASVs but only 2.2% of reads. These comparisons highlighted differences in taxonomic diversity that were not solely based on abundance.

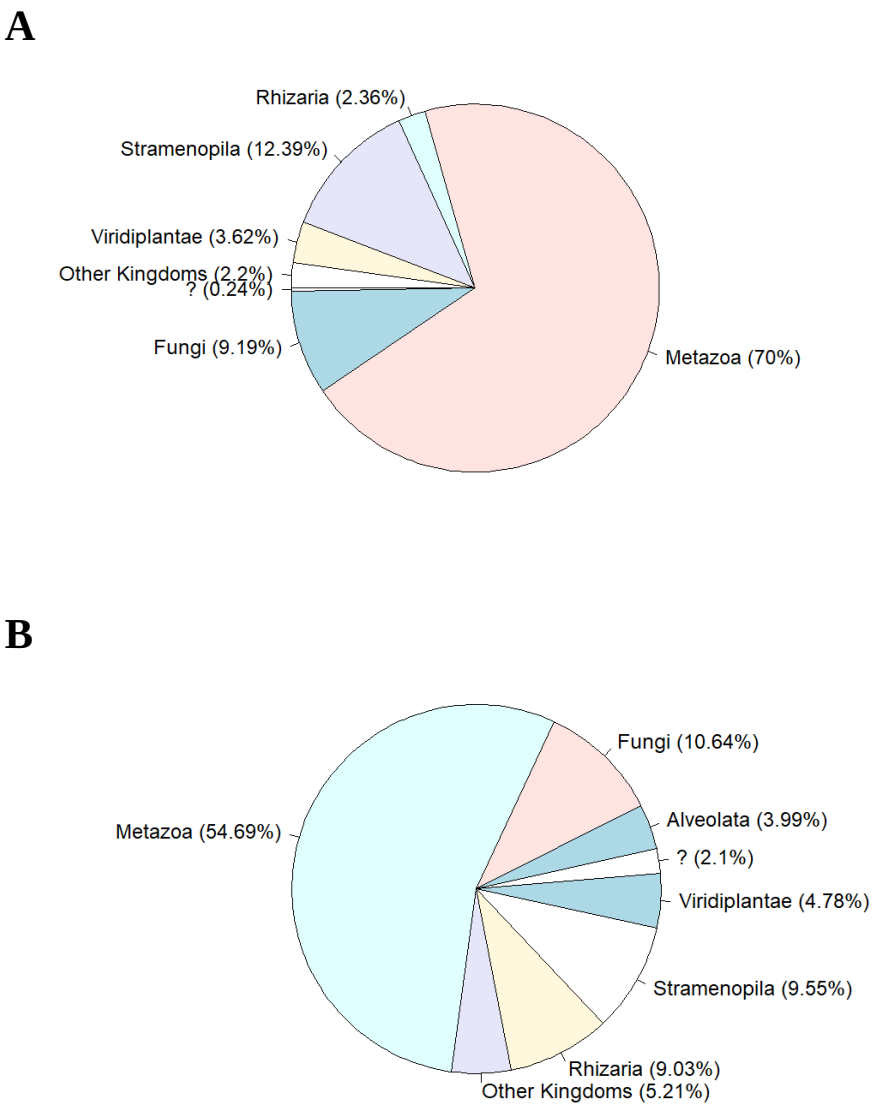


Fig. 3.10 Distribution of reads and ASVs across kingdoms for the G18S primer pair across all datasets. **A)** shows the proportion of reads, while **B)** displays the proportion of ASVs. Taxonomic assignments were made using the Eukaryome database. The '?' symbol denotes unresolved taxonomy at this level, as classified by LCA using Eukaryome matches.

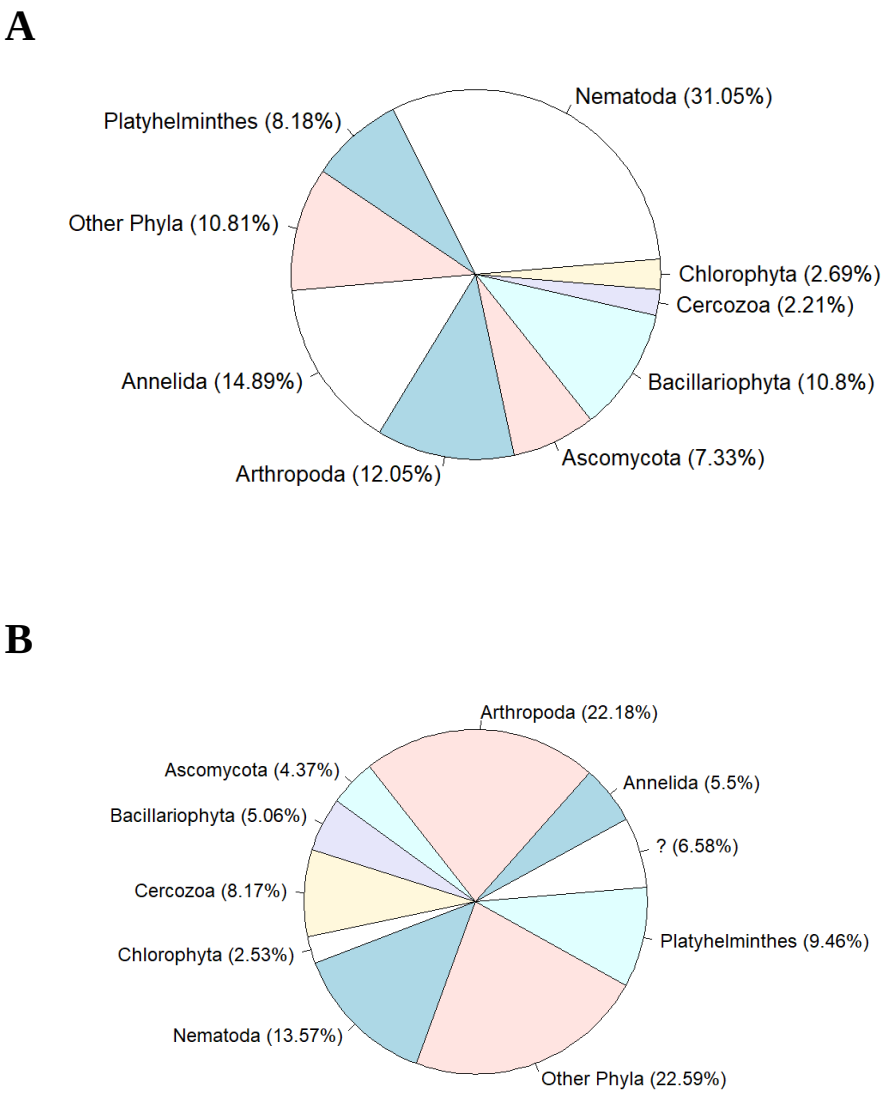


Fig. 3.11 Distribution of reads and ASVs across phyla for the G18S primer pair across all datasets. **A)** shows the proportion of reads, while B) displays the proportion of ASVs. Taxonomic assignments were made using the Eukaryome database. Phyla contributing less than 2% of the total in each chart are grouped as "other phyla." The '?' symbol denotes unresolved taxonomy at this level, as classified by LCA using Eukaryome matches.

The TAR primer pair generated an average of 726.4K reads and 1,037 ASVs per sample, amplifying a wide range of taxonomic groups. At the kingdom level, Metazoa accounted for 64.6% of the reads (**Fig. 3.12 A)**. The proportion of ASVs in each kingdom (**Fig. 3.12 B)** showed notable differences, with Metazoa making up only 10.3% of ASVs. Stramenopila and Fungi exhibited roughly similar proportions in both reads and ASVs, with Stramenopila at

12.4% of reads and 14.4% of ASVs. Rhizaria contributed 17.1% of ASVs, despite representing only 2.4% of the reads, while other kingdoms account for 2.84% of ASVs but only 2.2% of reads. Among the top phyla, Annelida dominated with 27.4% of the reads, followed by Nematoda at 12.4%, Platyhelminthes at 14.0%, Bacillariophyta at 11.4% and Arthropoda at 6.1% (**Fig. 3.13 A)**. The proportion of ASVs across phyla (**Fig. 3.13 B)** showed a different distribution, with 'Other Phyla' representing the largest share at 27.0%, followed by Cercozoa at 11.2%. Bacillariophyta and Labyrinthulidia each comprised 5.1% of ASVs. These differences reflected a higher diversity within certain groups despite lower read abundance, particularly for Cercozoa and the broad 'Other Phyla' category. Unresolved classifications ('?') accounted for a substantial portion of the ASVs.

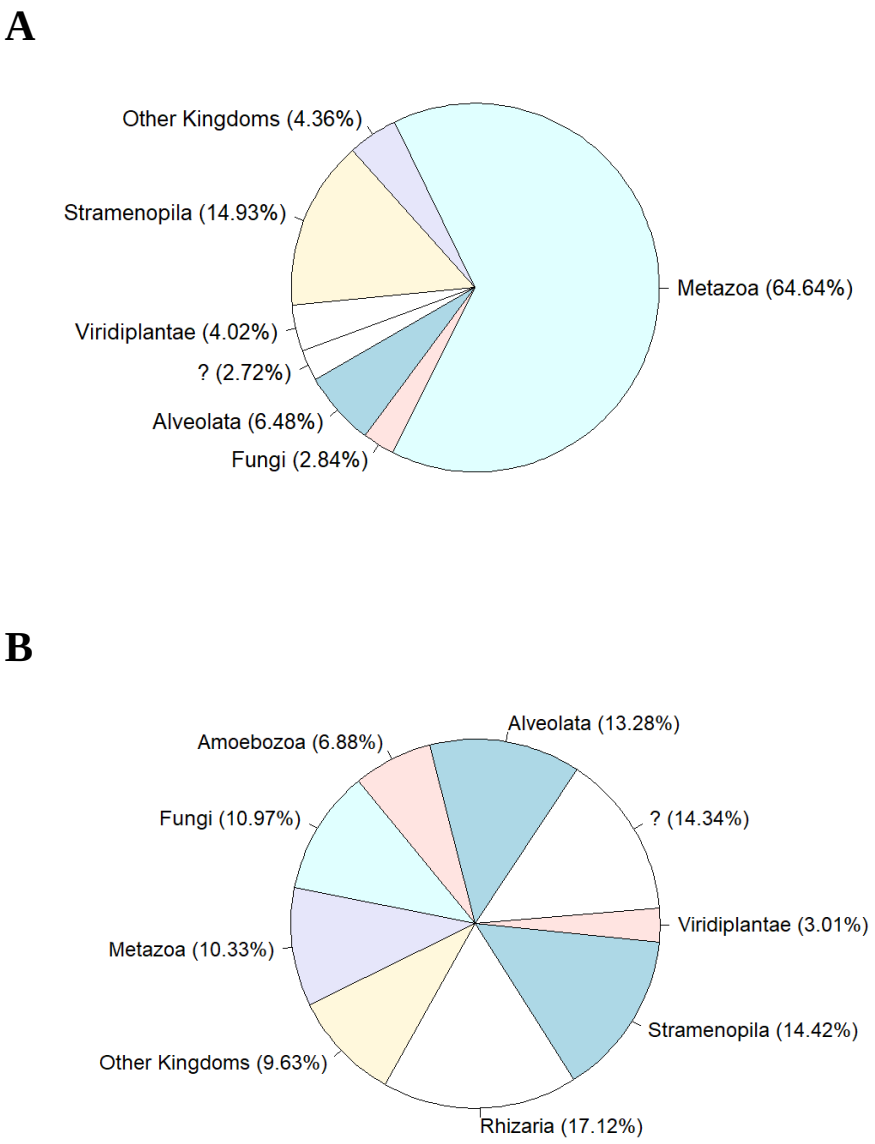
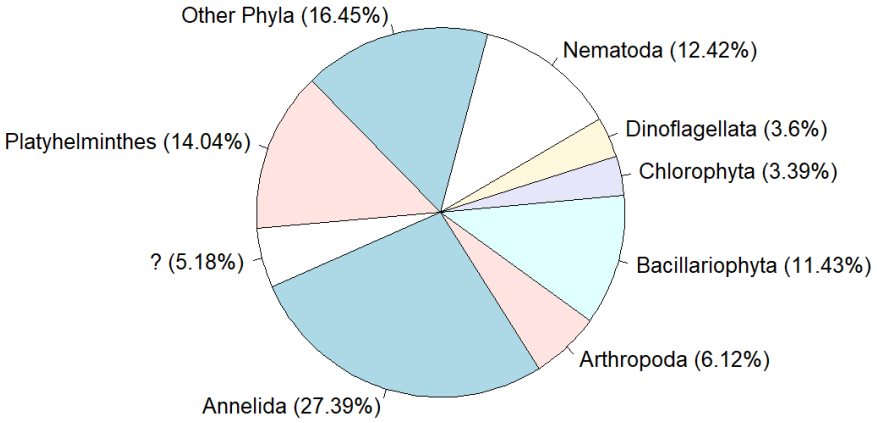


Fig. 3.12 Distribution of reads and ASVs across kingdoms for the **TAR** primer pair across all datasets**. A)** shows the proportion of reads, while **B)** displays the proportion of ASVs. Taxonomic assignments were made using the Eukaryome database. Kingdoms contributing less than 2% of the total in each chart are grouped under 'Other kingdoms'. The '?' symbol denotes unresolved taxonomy at this level, as classified by LCA using Eukaryome matches.

A



B

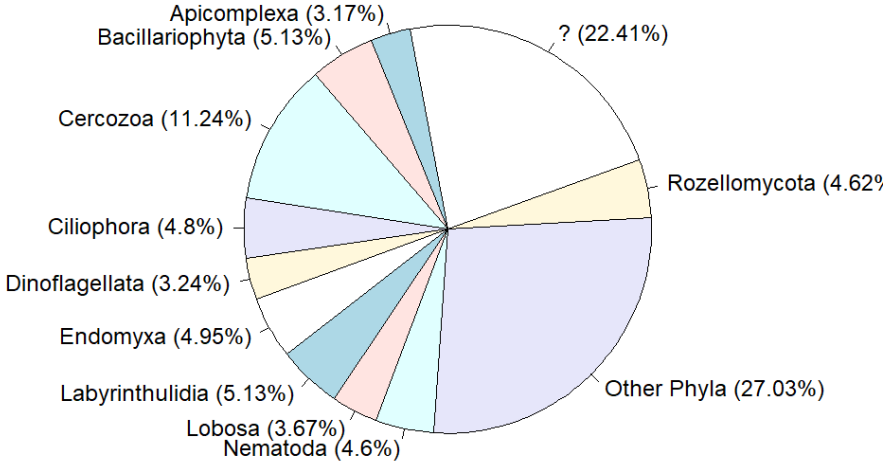


Fig. 3.13 Distribution of reads and ASVs across phyla for the **TAR** primer pair across all datasets**. A)** shows the proportion of reads, while **B)** displays the proportion of ASVs. Taxonomic assignments were made using the Eukaryome database. Phyla contributing less than 2% of the total in each chart are grouped under 'Other Phyla'. The '?' symbol denotes unresolved taxonomy at this level, as classified by LCA using Eukaryome matches.

The NEM primer data generated an average of 302K reads and 640 ASVs across the four pilot samples and showed a distinct pattern in taxonomic coverage at both the kingdom and phylum levels. At the kingdom level, Metazoa dominated with 59.6% of the reads (**Fig. 3.14 A).**

Unresolved classifications ('?') accounted for 39.7% of the reads, while other kingdoms combined represented only 0.7%. The proportion of ASVs in each kingdom (**Fig. 3.14 B)** showed Metazoa comprising the largest share at 80.8% of ASVs. Unclassified taxa ('?') make up 13.1%, with Stramenopila and Alveolata contributing 2.1% and 1.1% of ASVs, respectively, while other kingdoms constituted 2.9% of ASVs.

At the phylum level, Nematoda accounted for 36.8% of the reads and 24.8% of ASVs, followed by Xenacoelomorpha (16.6% of reads; 4.9% of ASVs), Annelida (3.6% of reads; 21.3% of ASVs) and Arthropoda (1.2% of reads; 11.0% of ASVs) (**Fig. 3.15 A** and **B)**. Unresolved classifications ('?') represented 39.7% of reads and 14.7% of ASVs. These differences highlighted higher diversity within certain phyla despite lower read abundance, particularly for Nematoda and Annelida, and underscored the broad range of taxa detected by the NEM primer.

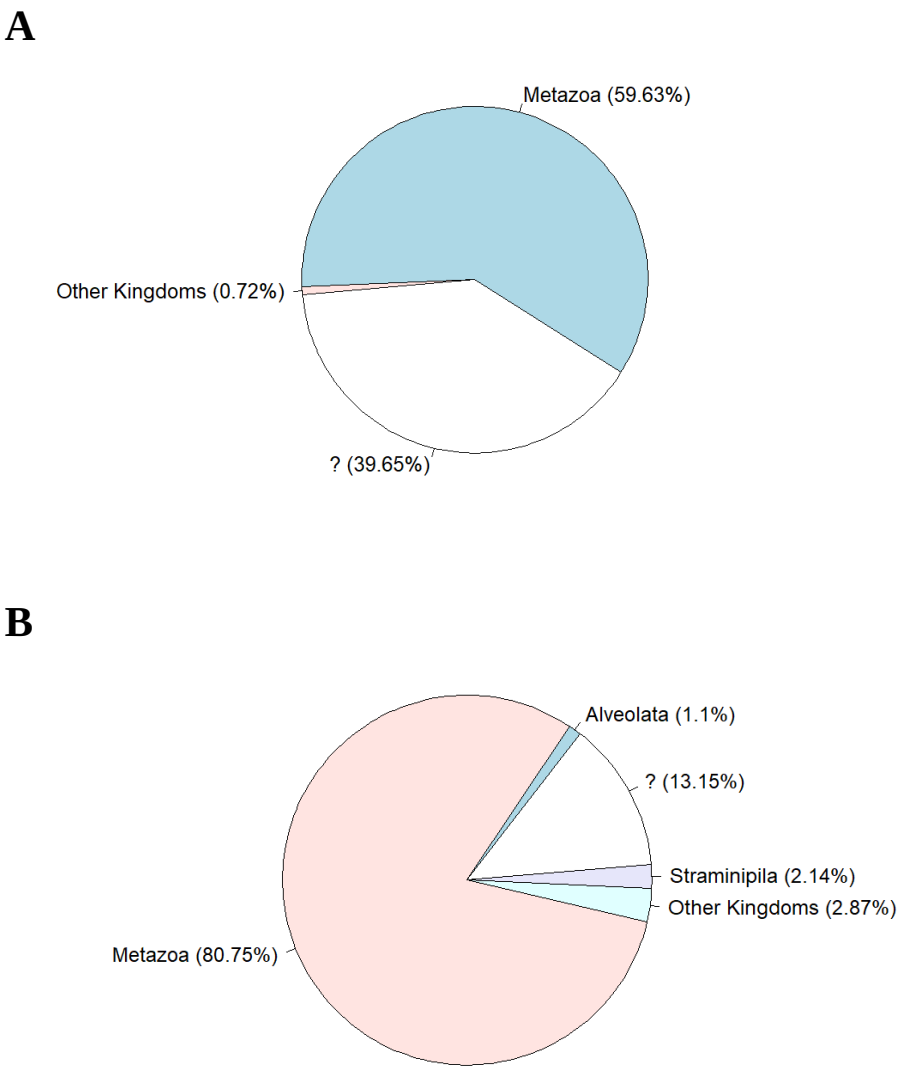


Fig. 3.14 Distribution of reads and ASVs across kingdoms for the **NEM** primer pair across all datasets. **A)** shows the proportion of reads, while **B)** displays the proportion of ASVs. Taxonomic assignments were made using the Eukaryome database. Kingdoms below 0.03% in reads and below 1% in ASVs are grouped as 'Other Kingdoms'. The '?' symbol denotes unresolved taxonomy at this level, as classified by LCA using Eukaryome "Best hit" matches.

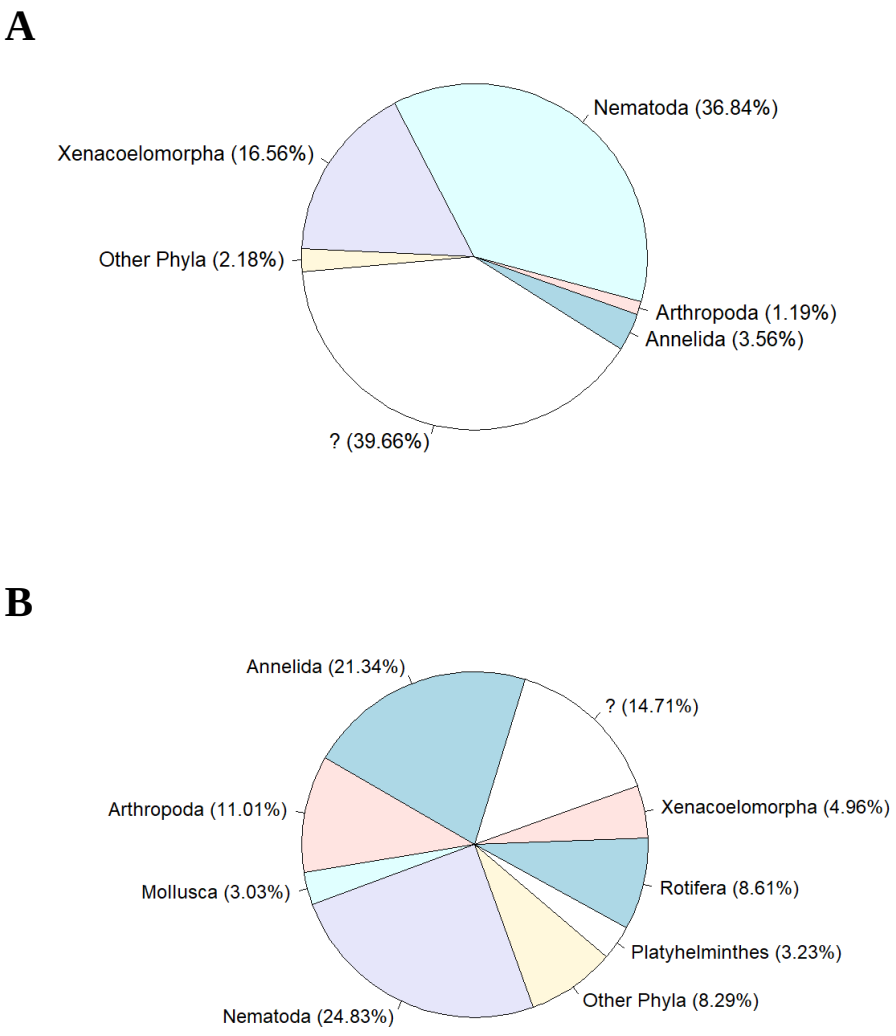
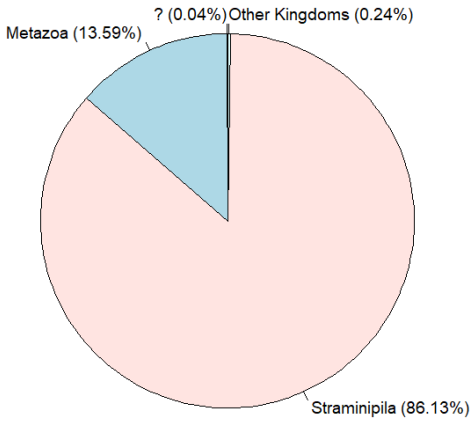


Fig. 3.15 Distribution of reads and ASVs across phyla for the NEM primer pair across all datasets. **A)** shows the proportion of reads, while **B)** displays the proportion of ASVs. Taxonomic assignments were made using the Eukaryome database. phyla below 1% in reads and 3% in ASVs are grouped as 'Other Phyla'. The '?' symbol denotes unresolved taxonomy at this level, as classified by LCA using Eukaryome "Best hit" matches

The E1391 primer was originally designed to target 18S sequences from Eukaryotes (Liu et al., 2019). It generated 56k reads for a single sample, but these reads corresponded to only 21 ASVs, predominantly Stramenopila and Metazoans. The main phyla amplified were Phaeophyta (brown algae) 62.6%, Bacillariophyta (diatoms) 23.5% and Arthropoda 6.8% (**Fig. 3.16**).

A



B

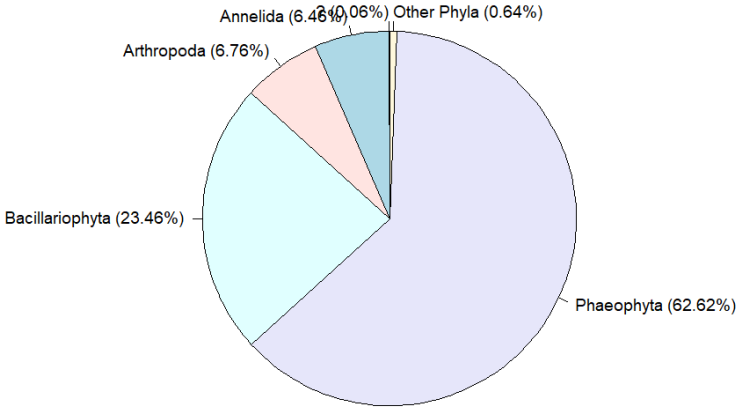


Fig. 3.16 Distribution of reads across **A)** Kingdoms and **B)** phyla for the pilot dataset, based on the total number of reads derived from the **E1391** primer. Taxonomic assignments were made using the Eukaryome database. For clarity, Phyla contributing under 1% each are grouped as "other phyla," totalling 0.6% of abundance. '?' indicates unresolved taxonomy at this level by LCA classification using Eukaryome matches.

The D512F primer generated an average of 90.9K reads across four samples, identified a relatively lower number of ASVs (256) than the other 18S primers, which may indicate a limited taxonomic range or reduced efficacy in diversity capture. When classified at phylum

level, small proportions of the reads were attributed to Labyrinthulida 3.4%, Chlorophyta 1.5%. However, the majority of the phyla proportion amplified by D512F were stramenopiles (97%), of which diatoms made up 91.3% (**Fig. 3.17 A** and **B**).

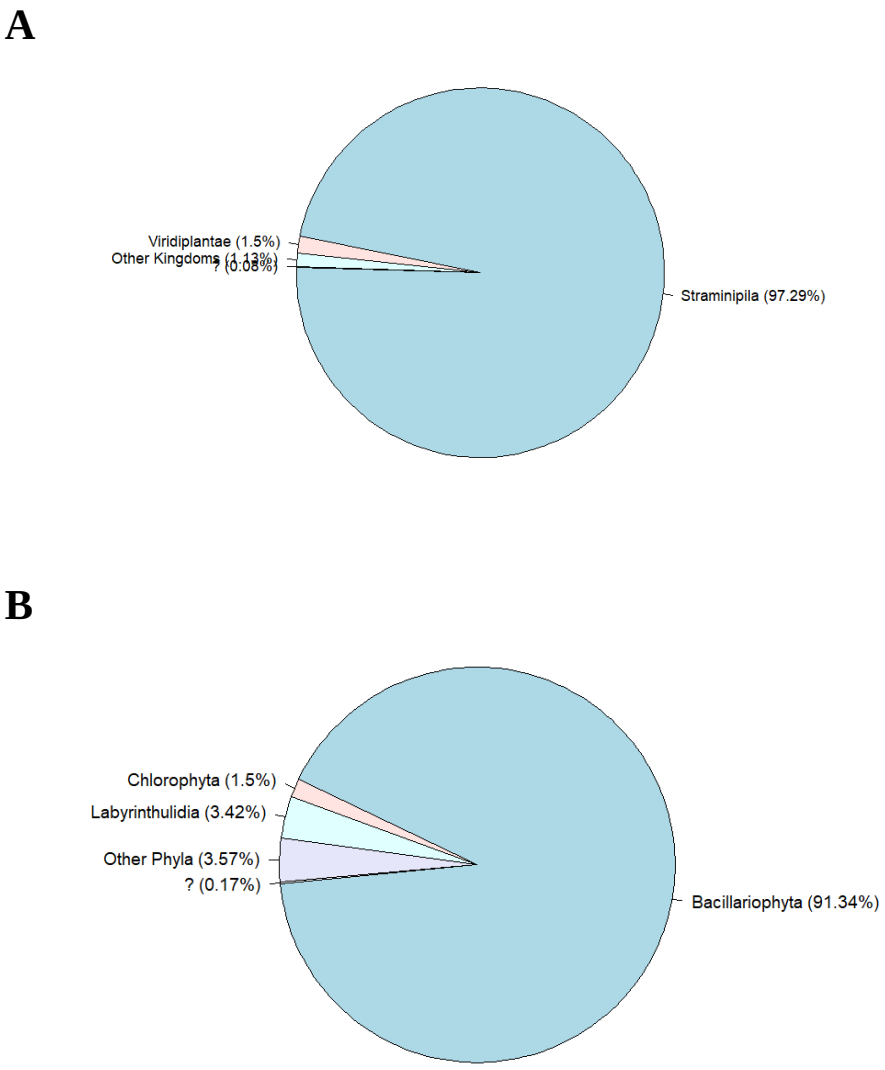


Fig. 3.17 Distribution of reads across **A)** Kingdoms and **B)** phyla for the pilot dataset, based on the total number of reads derived from the **D512F** primer. Taxonomic assignments were made using the Eukaryome database. For clarity, Phyla contributing under 1% each are grouped as 'other phyla', totalling 3.6% of abundance. '?' indicates unresolved taxonomy at this level by LCA classification using Eukaryome matches.

Although the DM568 primer yielded lower sequencing depth and diversity than 18S eukaryotic primers, it provided substantial taxonomic insights. Generating an average of 8.9K reads and 55 ASVs per sample, the DM568 primer captured important data on taxonomic distribution across kingdoms and phyla. This lower sequencing yield was largely due to the limited availability of eukaryotic 28S sequences in databases like SILVA and Eukaryome compared to the more comprehensive 18S datasets. Nonetheless, a substantial proportion of reads were assigned to nematodes, demonstrating the primer's effectiveness in detecting specific taxa.

At the kingdom-level distribution, Metazoa overwhelmingly dominated both reads and ASV proportions, comprising 93.1% of reads (**Fig. 3.18 A)** with smaller contributions from Alveolata, Other Kingdoms, Rhizaria and Viridiplantae, which together represented the remaining 6.9%. Looking at the ASV proportions (**Fig. 3.18 B),** Metazoa maintained the largest share at 61.8%, followed by Alveolata at 10.6% and Rhizaria at 9.8%. Fungi, Viridiplantae, Other Kingdoms and Stramenopila appeared in smaller proportions, indicating broader taxonomic diversity within ASVs than in read abundance (**Fig. 3.19 A**).

Focusing on phylum-level data, Nematoda emerged as the dominant group in read distribution, making up 33.9% of reads, followed by Annelida at 19.8% and Platyhelminthes at 15.5% (**Fig. 3.19 A**). Arthropoda contributed 14.1%, while 'Other Phyla' accounted for 6.4%, with unresolved classifications ('?') adding another 5.2%. This distribution highlighted the prominence of a few key phyla, particularly Nematoda, within the detected community structure. For ASV proportions among phyla (**Fig. 3.19 B)**, Nematoda again led with 23.6%, closely followed by 'Other Phyla' at 26.0%, which reflected high diversity within taxa outside the major groups. Arthropoda and Platyhelminthes made up 14.6% and 12.2% of ASVs, respectively. Additional minor groups, including Cercozoa, Chytridiomycota and Ciliophora, each contributed around 3-7% of ASVs, showcasing a diverse taxonomic representation.

In summary, despite the lower sequencing depth, the DM568 primer effectively captured key taxa, especially nematodes, offering valuable insights into community structure and diversity. Further taxonomic characterisation of 28S rRNA-based ASVs was provided in **Chapter 6**.

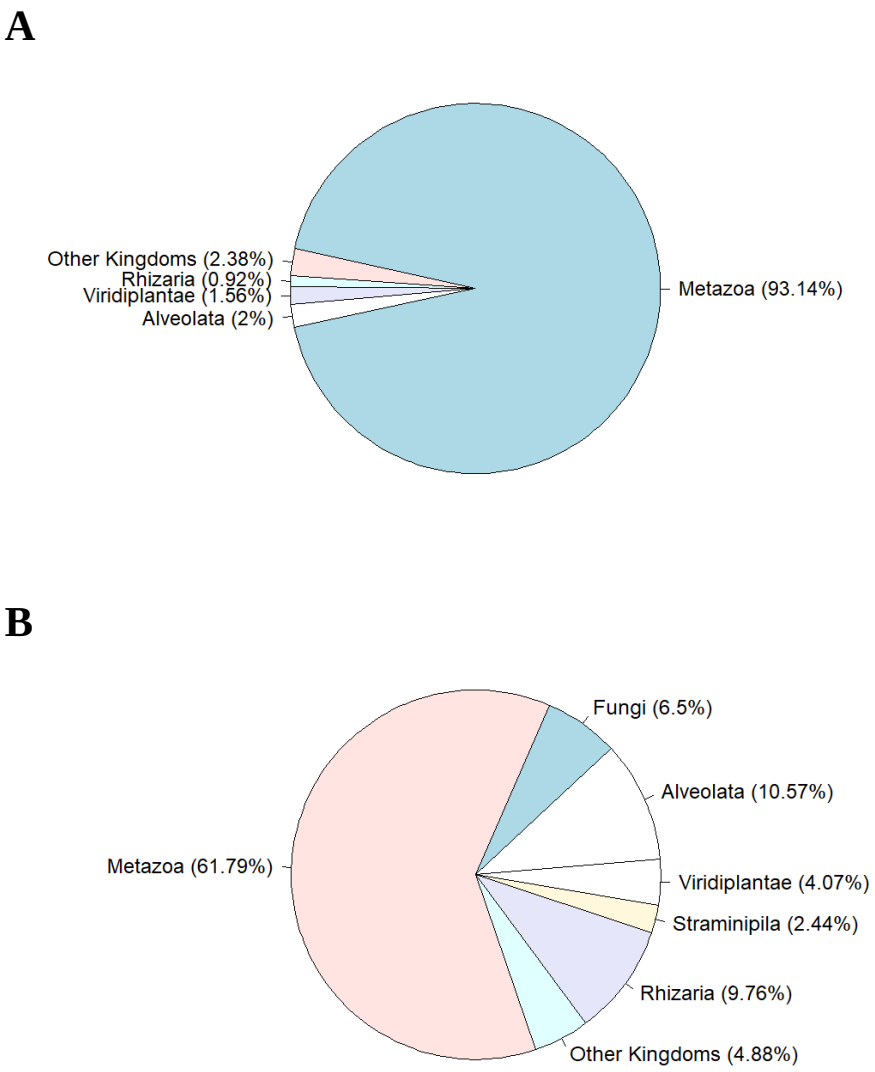


Fig. 3.18 Distribution of reads and ASVs across kingdoms for the **DM568** primer pair across all datasets. **A)** shows the proportion of reads, while **B)** displays the proportion of ASVs. Taxonomic assignments were made using the Eukaryome database. For clarity, Kingdoms contributing less than 0.8% of total reads and less than 2% of ASVs are grouped as 'Other Kingdoms' in each chart. '?' indicates unresolved taxonomy at this level by LCA classification using Eukaryome matches.

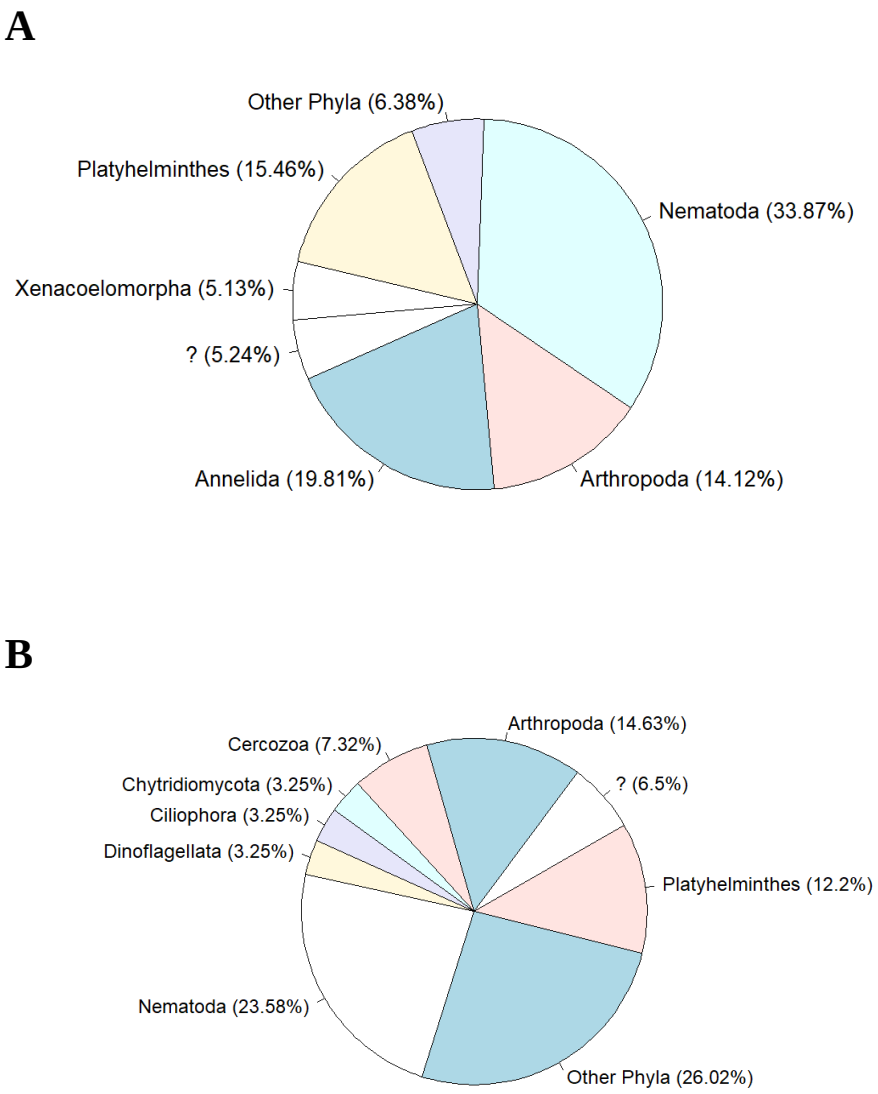


Fig. 3.19 Distribution of reads and ASVs across phyla for the DM568 primer pair across all datasets. A) shows the proportion of reads, while B) displays the proportion of ASVs. Taxonomic assignments were made using the Eukaryome database. For clarity, phyla contributing less than 3% of total reads and less than 3% of ASVs are grouped as 'Other Phyla' in each chart. '?' indicates unresolved taxonomy at this level by LCA classification using Eukaryome matches.

Although the JB3 primer generated a substantial volume of sequence data, averaging 233.3K reads per sample, only 32 ASVs were identified. Nearly all 99.9% of the reads were assigned to the Eukaryota kingdom, indicating a highly targeted amplification in this domain. For read distribution, Eukaryota accounted for 99.9% of reads, with only 0.1% remaining unclassified

('?'). This reflected the JB3 primer's targeted performance towards eukaryotic sequences. The ASV proportions aligned closely, showed that 92.6% of ASVs belong to Eukaryota, with 7.5% remaining unresolved ('?').

In terms of read distribution across phyla, the JB3 sequences included some reads from nematodes 6.9%, Ascomycota 3.5%, Arthropoda 0.8% and other minor phyla 0.4%, but the vast majority 88.4% were unclassified at the phylum level (**Fig. 3.20 A)**. The ASV proportions reflected a similar pattern, with 67.0% of ASVs unclassified ('?'). Among the classified ASVs, Nematoda held the largest share at 8.5%, followed by Rhodophyta 7.5%, Ascomycota 6.4% and Oomycota 6.4%, with minor contributions from Arthropoda 3.2% and Platyhelminthes 1.1% (**Fig. 3.20 B)**. Despite the high read yield, the limited diversity within the ASVs highlighted the JB3 primer's specificity and suggested a narrower amplification range compared to other primers, which may restrict broader taxonomic representation.

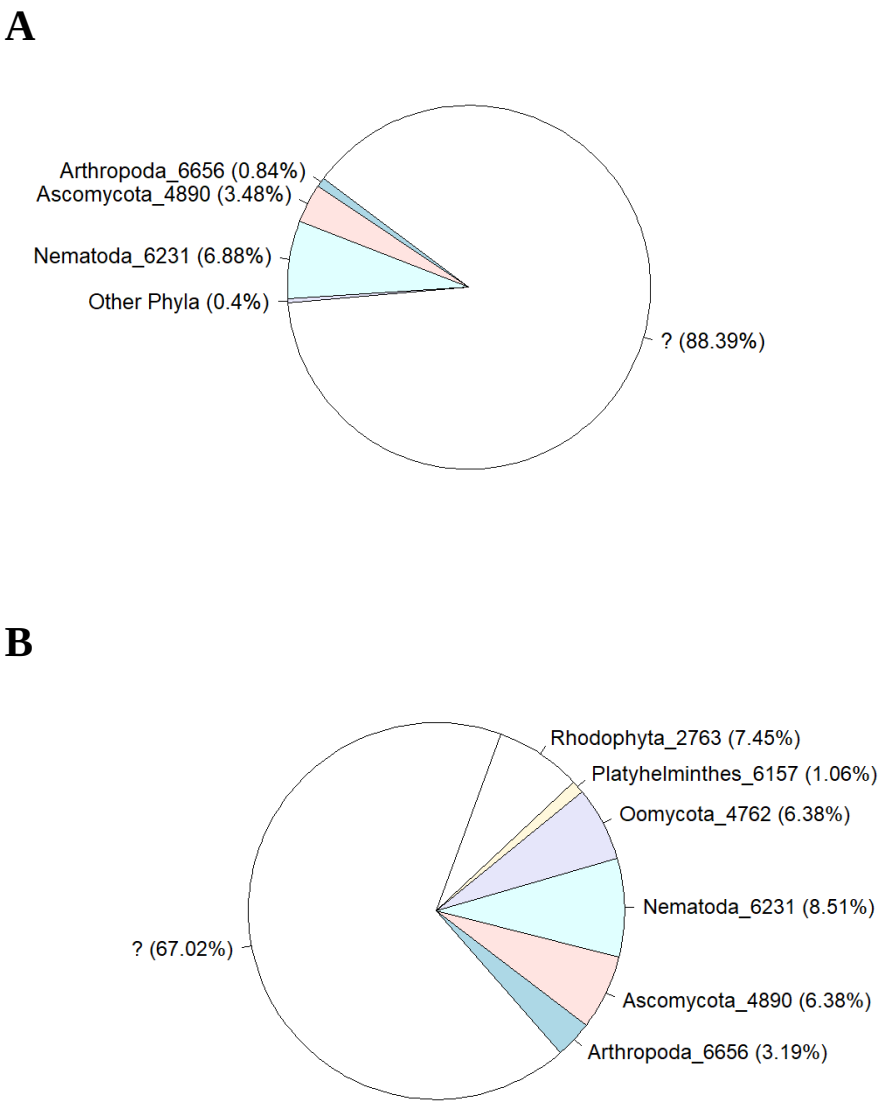


Fig. 3.20 Distribution of **A**) reads and **B**) ASVs across phyla for the **JB3** primer pair across all datasets. Taxonomic assignments were made using the Eukaryome database. Taxonomic assignments were made using MIDORI-Longest COI database. For clarity, phyla contributing less than 0.4% of total reads and with no threshold applied to ASVs are grouped as 'Other Kingdoms' or 'Other Phyla' in each chart. '?' indicates unresolved taxonomy at this level by LCA classification using MIDORI matches.

3.3.3 Performance Comparison of Primers and Microscopy for Nematode Detection

Building on the extraction and amplification procedures detailed in **Chapter 2**, he effectiveness of each primer set in recovering nematode diversity was evaluated by directly comparing

metabarcoding outputs with traditional morphology-based counts. A direct comparison of molecular and morphological methods was undertaken to evaluate their effectiveness in nematode identification (**Table 3.5**). Morphologically, an average of 21 species was recorded per 100 sieved individuals (Udochi, 2020). By contrast, the metabarcoding approaches uncovered substantially higher ASV counts, particularly with the G18S and NEM primers. The G18S primer retained the highest number of ASVs in both sieved and whole-sediment treatments, yielding 359 ASVs for sieved samples and 317-483 ASVs for 0.25 g and 2.7 g sediment extractions, respectively. The NEM primer also returned elevated ASV numbers relative to microscopy. The TAR primer detected moderate ASV numbers, whereas JB3 identified the fewest ASVs across all treatments. The 2.7 g whole-sediment extraction yielded 483 ASVs, whereas the 0.25 g extraction yielded 317 ASVs, however, these counts were generated using different extraction protocols (**Chapter 2**).

Table 3.5 Average **nematode** ASVs detected using different primers and sample treatments in this study, compared with species counts from morphological identification by Udochi (2020). All samples were collected from BW, Norfolk in spring 2020, except for the 0.25 g-TAR primer data, which were obtained from the moderately polluted site Mylor Creek (MC), southwest England in spring 2022. DNA sequencing was performed using NovaSeq.

Identification Method	0.25 g Whole sediment	2.7 g Whole sediment*	Sieved nematodes
Microscopic identification (Sp.)	-	-	21
G18S (ASV)	317	483	359
TAR (ASV)	50	37	56
NEM (ASV)	86	118	108
JB3 (ASV)	11	16	19

* The 2.7 g whole sediment samples were extracted using an RNA-specific kit, as detailed in **Chapter 2**. Assessment of Molecular Identification Methods for Metazoan Diversity.

To test if different primers and sample treatments influence had an effect on diversity detection, the average ASV counts was analysed for 0.25 g and 2.7 g whole-sediment extractions, as well as nematodes separated by sieving (**Fig. 3.21**). Overall, the G18S primer consistently recovered the highest average ASV counts across all sample types. TAR recovered over 1000 ASVs from both 0.25 g and 2.7 g extractions, compared to nearly 4000 ASVs recovered by G18S. NEM

yielded 499 ASVs from the 2.7 g sample. JB3 produced the fewest ASVs across all treatments, with values below 160.

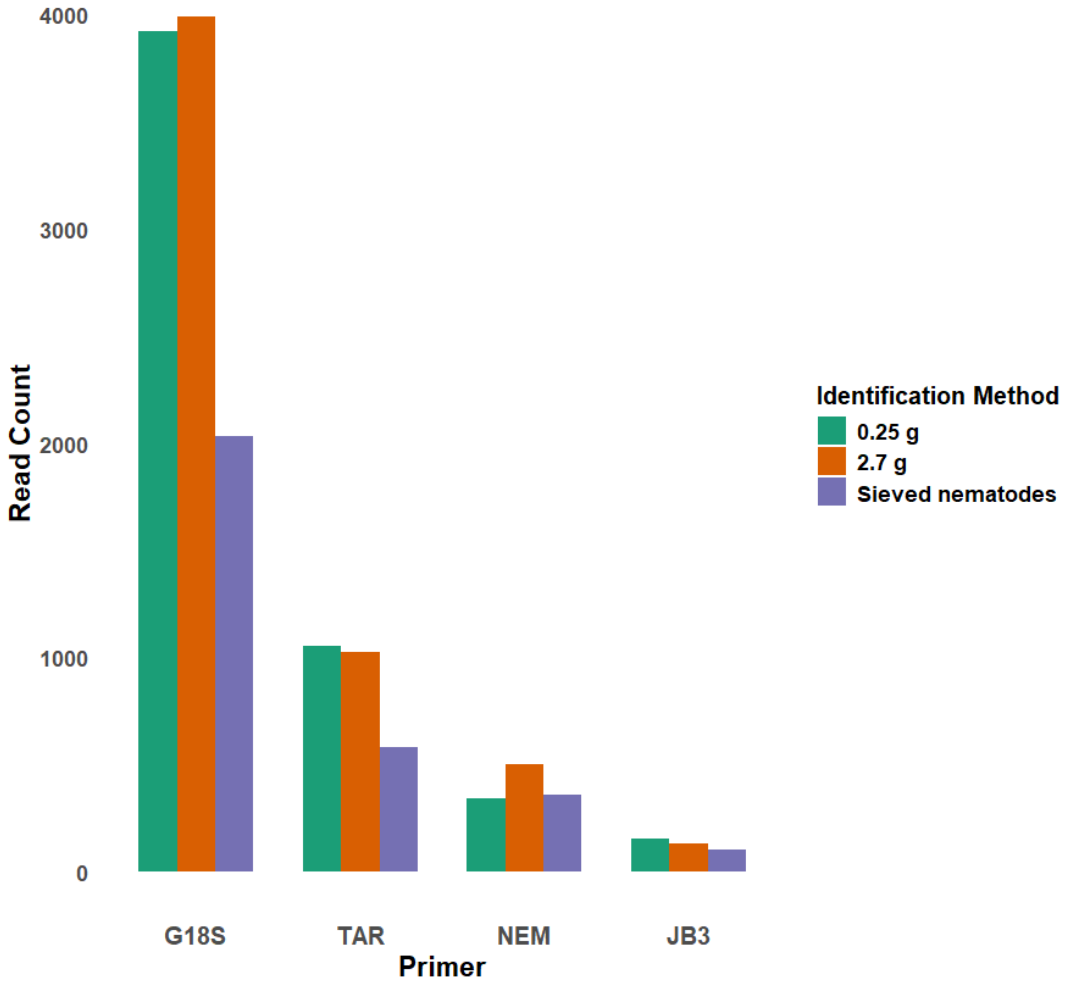


Fig. 3.21 Average **metazoan** ASVs detected using different primers and sample treatments. All samples were collected from BW, Norfolk in spring 2020, except for the 0.25 g-TAR primer data, which were obtained from the moderately polluted site Mylor Creek (MC), southwest England in spring 2022. DNA extractions and sequencing were performed using NovaSeq. The 2.7 g whole sediment samples were extracted using an RNA-specific kit, as detailed in **Chapter 2**.

3.3.3.1 Comparison of Species Richness and Community Structure

Rarefaction of the metabarcoding data to a standardized sample size of 100 reads allowed for direct comparison of molecular diversity estimates with those obtained from morphological methods. The resulting box plots of ASV richness revealed differences in the total number of species detected by each primer across various sample treatments. K-dominance curves facilitated comparisons of the relative abundances of dominant taxa between samples

The analysis of ASV richness and species dominance patterns revealed pronounced influences of primer choice and sample treatment on the detection of nematode diversity. Primer selection emerged as the most critical factor, with the biggest differences in nematode diversity observed between primers. The G18S primer consistently exhibited the highest ASV richness across all sample types (**Fig. 3.22**), detecting a level of diversity comparable to that obtained through morphological identification. The TAR primer also performed well, yielding slightly fewer species than G18S but still providing a substantial representation of the nematode community.

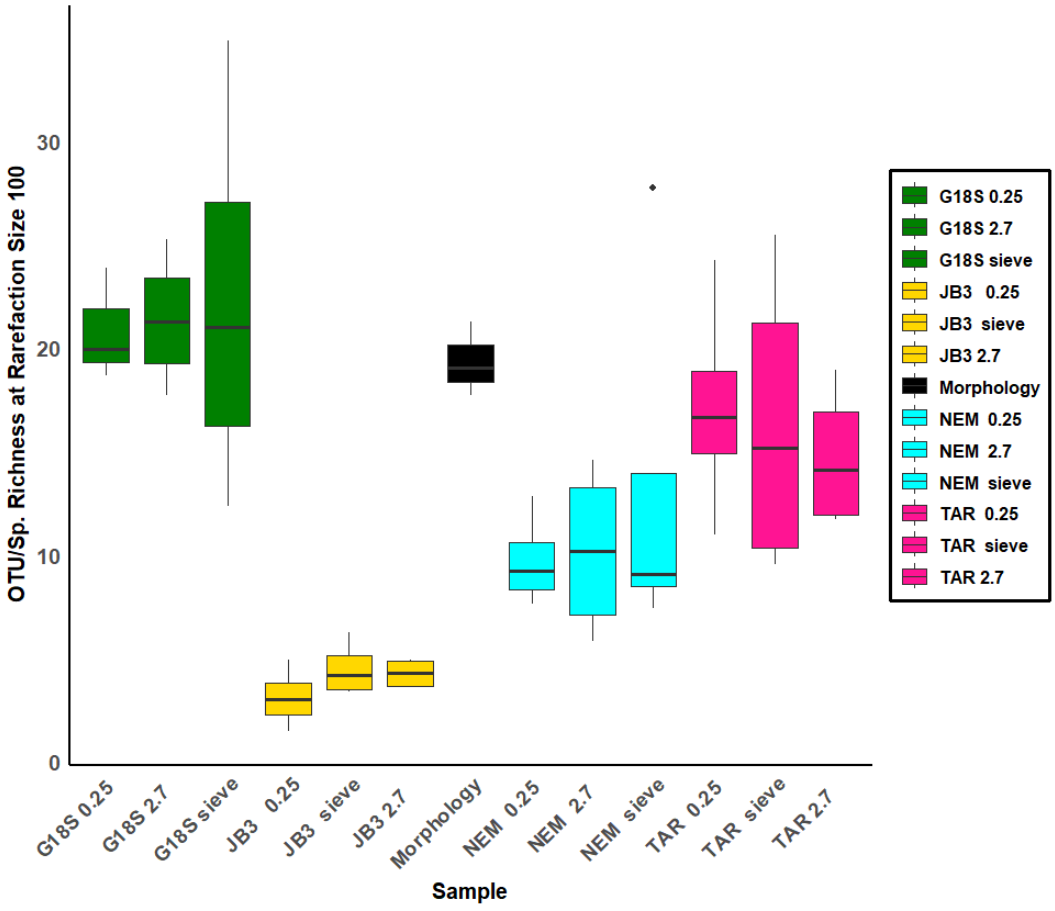


Fig. 3.22 Combined box and whisker plot showing nematode ASV richness detected by primers G18S, JB3, NEM and TAR, together with morphological identification, across all samples.

K-dominance plots demonstrated that G18S and TAR primers produced similar, although not identical, patterns in species dominance (**Fig. 3.23 A** and **B**). Both primers detected communities with a balanced distribution of abundant and rare species, indicating their effectiveness in capturing overall community structure. In contrast, the JB3 and NEMF primers detected appreciably lower diversity, often with a single ASV dominating the abundance data. The k-dominance curves for JB3 and NEMF showed steeper slopes, reflecting dominance by a few species and reduced detection of rare taxa. JB3 consistently underperformed across all sample sizes assessed and NEMF also showed limited effectiveness in capturing the full spectrum of nematode diversity.

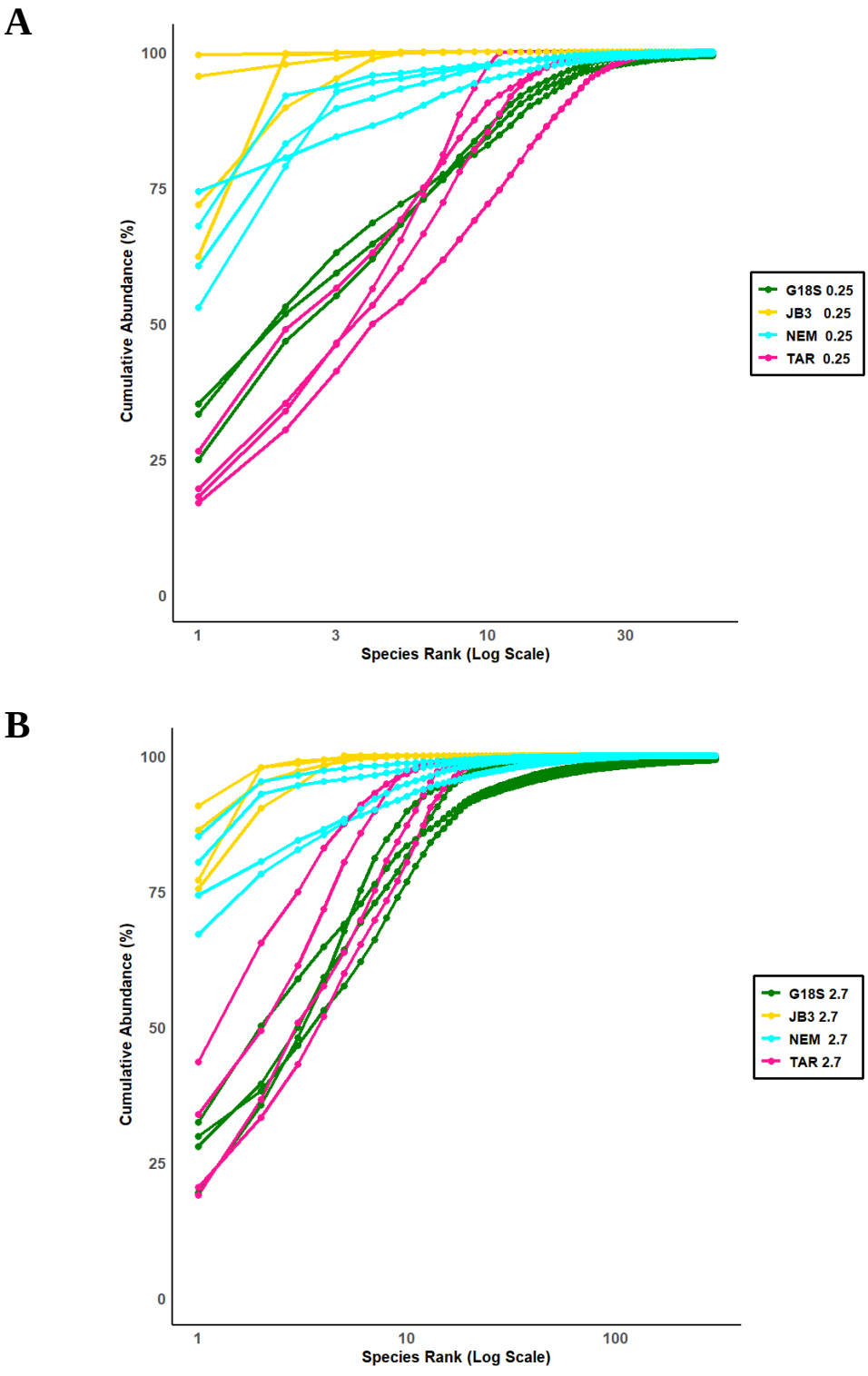


Fig. 3.23 Combined box-and-whisker and K-dominance plots illustrating nematode ASV richness and species-dominance patterns detected by various primers and treatments across different sample types. **A)** and **B)** present K-dominance curves for primers G18S, JB3, NEM and TAR on 0.25 g and 2.7 g samples, respectively. (continued in **Fig. 3.24**)

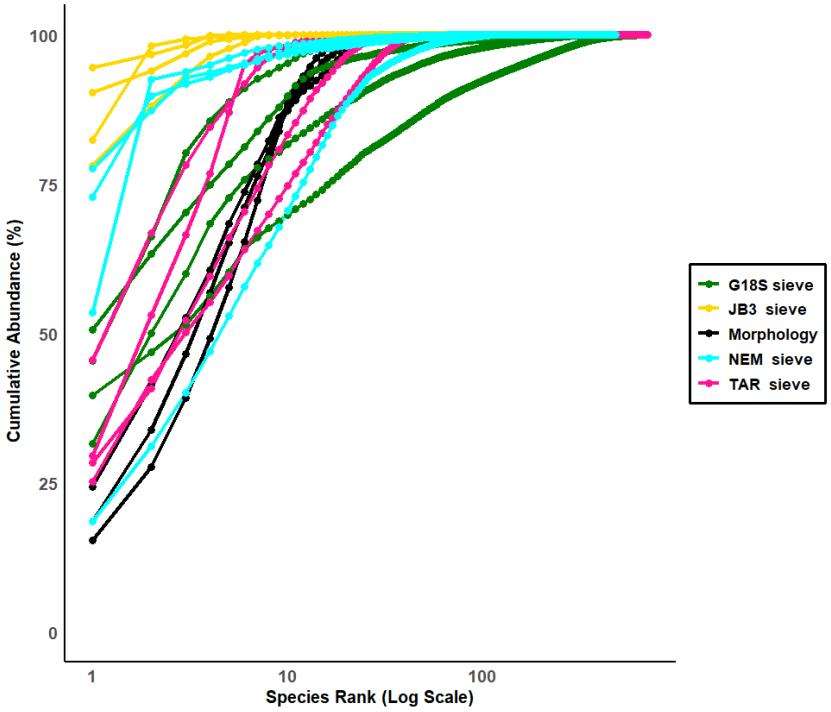


Fig. 3.24 (continuation of **Fig. 3.23**) shows K-dominance curves for the same primer sets applied to sieved-mud samples.

To assess how sample volume influences ASV richness, diversity patterns were examined across multiple sediment weights. For G18S (**Fig. 3.25 A**), the 0.25 g, sieved, and 2.7 g treatments display overlapping ASV-richness medians, however, as the 2.7 g samples were extracted with a different protocol, these values are presented descriptively rather than as a direct test of sample-size effects. In the Percuil River data (**Fig. 3.25 B**), 10 g extractions gave the highest ASV counts. The G18S recovered the greatest richness and DM568 also yielded high, consistent values across replicates.

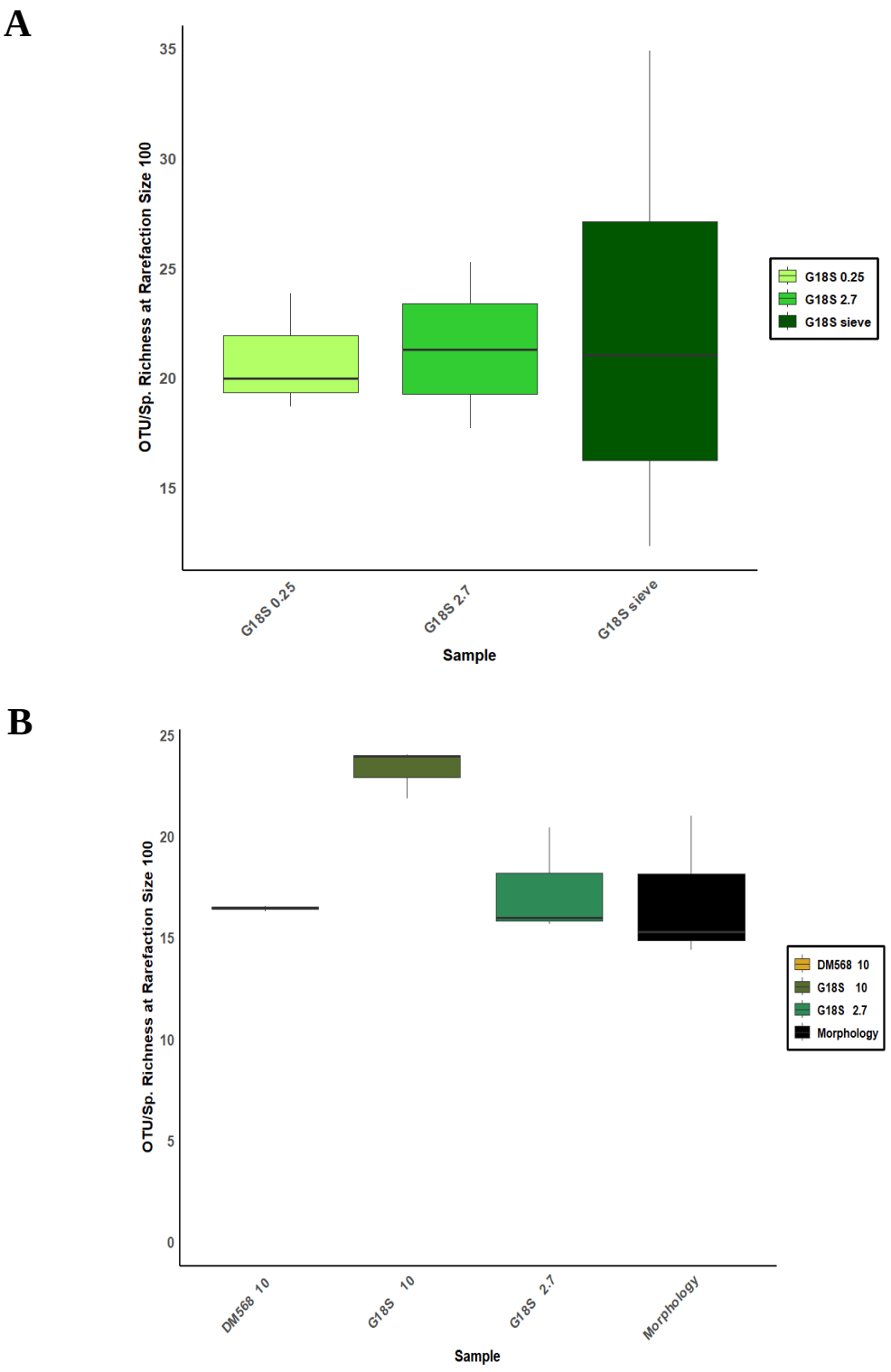


Fig. 3.25 Combined boxes and whiskers and K-dominance plots illustrating the nematode ASV richness and species-dominance patterns detected by various primers and treatments across different sample types. **A)** compares ASV richness among treatments (0.25 g, 2.7 g, sieved) using the G18S primer, while **B)** compares ASV richness among treatments (10 g, 2.7 g and sieved) in sediments sampled from the Percuil River, SW-UK, using G18S and DM568 primers alongside morphological analysis.

A comparison of k-dominance curves is presented in **Fig. 3.26 A** for the 0.25 g, 2.7 g, and sieved sediment treatments using the G18S primer. The sieved samples trace the flattest, most gradual curve, though replicate spread is evident, whereas the 0.25 g and 2.7 g curves are steeper and almost overlapping, indicating a stronger influence of a few abundant ASVs at these smaller volumes. **Fig. 3.26 B** extended the analysis to the larger 10 g extractions. The G18S 10 g method displayed the flattest curve. The DM568 10 g method also performed relatively well, with a moderately flat curve, though it tended to emphasize dominant taxa compared to G18S 10 g. The G18S 2.7 g method showed a steeper curve.

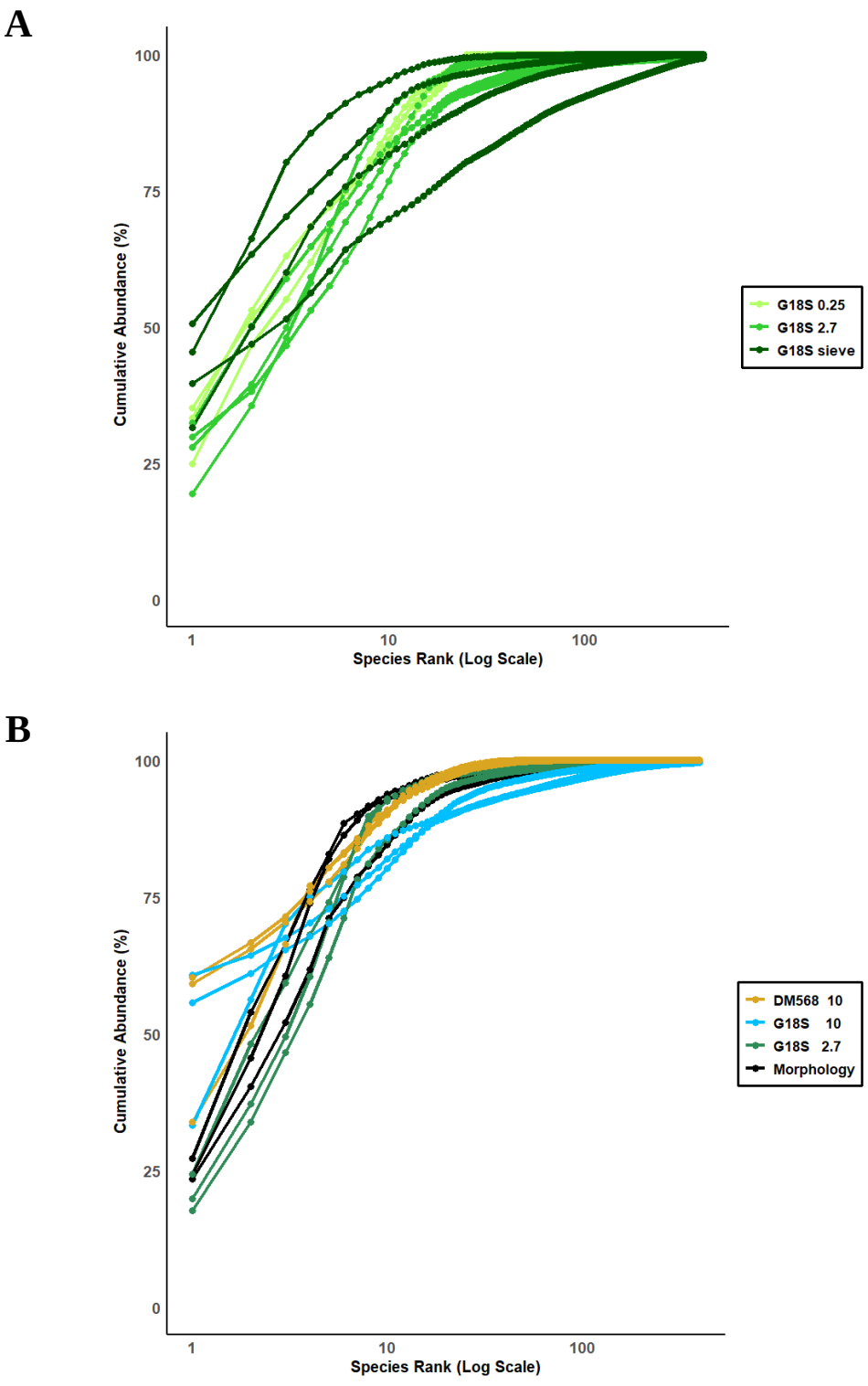


Fig. 3.26 Combined boxes and whiskers and K-dominance plots illustrating the nematode ASV richness and species-dominance patterns detected by various primers and treatments across different sample types. **A)** displays the K-dominance curves corresponding to the ASV-richness, while **B)** shows the related K-dominance curves including morphology data**.**

3.4 Discussion

An important consideration was whether any single primer set could comprehensively capture the entire diversity of marine sediment communities or whether multiple markers were required to cover both unicellular and metazoan taxa (Tytgat et al., 2019, Fonseca et al., 2022). By evaluating 16S, ITS, 18S, 28S and COI, these findings showed that no single marker alone can provide complete eukaryotic coverage, reaffirming the necessity of a complementary multi-primer approach. The 515F primer pair successfully amplified bacterial 16S rRNA, yielding high ASV richness and detecting phyla such as Pseudomonadota, Bacteroidota and Planctomycetota (Needham et al., 2019, Polinski et al., 2019, Li et al., 2022). Similarly, the ARF primer enriched archaeal 16S rRNA genes, capturing Thermoproteota and Nanoarchaeota (Hadziavdic et al., 2014, Hugerth et al., 2014) and Subsequent studies (Di Cesare et al., 2020, Schenk et al., 2022, Grant et al., 2023, Zhang et al., 2024a) confirm that archaeal diversity was often underestimated without specialised primers. Refining clustering thresholds from 97 to 99-100% similarity notably improves resolution within Nanoarchaeota and Woesearchaeales, which dominated the archaeal OTUs in this study (Edgar, 2017, Grant et al., 2023). Meanwhile, the ITS primer pair yielded an average of 41K reads per sample, yet only 4.6% were assigned to fungi, while 52% remained unclassified at the kingdom level, possibly due to incomplete marine-fungal databases (Derycke et al., 2005, Tytgat et al., 2019). Many ITS-based ASVs mapped to groups such as Stramenopila and Viridiplantae (Schmidt et al., 2013, Harnelly et al., 2022) highlighting factors such as sample type, DNA yield, preservation methods and database limitations can all influence diversity assessment. VRAIN primers generated fewer reads than earlier nematode studies (Vrain et al., 1992, Derycke et al., 2010), pointing to primer efficiency or database coverage issues.

Focusing on 18S primers (G18S, TAR, NEM, E1391, D512) revealed differences in read depth and breadth. The G18S primer amplifies the V1-V2 loop (~400 bp), The V1–V2 primers are widely used in meiofaunal surveys because they capture a high fraction of nematodes and other benthic animals but they recover fewer protist lineages than V4 or V9 markers, partly owing to primer mismatches in excavates and haptophytes (Sikder et al., 2020, Liu and Zhang, 2021, Gattoni et al., 2023). Accordingly, G18S produced high read counts across metazoans, though it underrepresented some protists (Amaral-Zettler et al., 2009, Stoeck et al., 2010, Tytgat et al., 2019). Both TAR and D512 target the V4 loop (350-420 bp). V4 is generally considered the

best “all-round” region for whole-community work because it balances broad eukaryote coverage with high entropy and good taxonomic resolution (Owens et al., 2024). The TAR primers amplified diverse suite eukaryotes here but with slightly lower efficiency (Fonseca et al., 2022, Maosa et al., 2024). Previous research often used TAR in fractionated water samples (Stoeck et al., 2010, Kim et al., 2012, Shi et al., 2022), excluding certain groups with a 63 µm sieve, here, TAR was applied to unfractionated benthic sediment, revealing a broader suite of taxa including metazoans. Similarly, the D512, also V4-based, was dominated by stramenopiles, consistent with earlier reports (Zimmermann et al., 2011, Luddington et al., 2012, Kim et al., 2017). Primers targeting the long V6-V8 region (~600 bp), such as NEM, have consistently outperformed other 18S markers for nematode detection, with numerous mock-community and field trials reporting over 70% nematode reads in soil or benthic samples (Waeyenberge et al., 2019, Sikder et al., 2020, Sikder et al., 2021). Here, NEM was robust for nematodes (Porazinska et al., 2009, Sapkota and Nicolaisen, 2015) but returned fewer ASVs when sample sizes were standardised to 100 reads, suggesting the possibility of underestimating rare protists. Targeting the short V9 loop (310 bp), the E1391 primer pair is well-suited for detecting pico-eukaryotes and degraded DNA, capturing alveolate and rhizarian clades often missed by V4 or V1-V2 (Amaral-Zettler et al., 2009, Choi and Park, 2020). Its taxonomic resolution for animals, however, is generally limited to the class level. Its greater sensitivity helps recover overlooked taxa (Choi and Park, 2020). In this study, the E1391 primer pair primarily amplified diatoms, indicating its potential use for broad phylogenetic patterns rather than fine-scale profiling (Amaral-Zettler et al., 2009, Stoeck et al., 2010, Liu et al., 2019),

DNA extracted from sediment is often challenging to amplify, thus primer mismatches and copy-number differences can distort read counts (Elbrecht and Leese, 2015). Each primer was applied successfully in different contexts, but careful attention to sample type and extraction protocols was necessary to minimise underrepresentation (Lamb et al., 2019). Employing multiple loci is a straightforward way to mitigate bias, with multi-marker surveys typically recovering 25-40% more phyla than single primer sets and reducing locus-specific drop-outs (Pompanon et al., 2012, Leary et al., 2014). Ideally, PCR-free approaches such as shotgun metagenomics could more accurately capture taxon abundances by bypassing PCR biases (Losada et al., 2014, Eloe-Fadrosh et al., 2016), though higher costs and complex workflows limit routine adoption (Carew et al., 2018)**.** PCR-based approaches remain useful but can distort

abundances through mismatches and incomplete coverage, requiring researchers to refine primer designs and combine multiple markers where possible.

Sample preparation and extraction size can critically affect both the depth of coverage and the range of taxa detected (Barnes and Turner, 2016, Klunder et al., 2019). For instance, mall prokaryotes and meiofauna are often better detected in unsieved sediment, because a 63 µm sieve retains only the larger particles, the finer filtrate that contains nematodes and other microscopic taxa is commonly discarded (Klunder et al., 2019). Conversely, lower-density macrofauna may be underrepresented in unsieved protocols, since the probability of encountering larger organisms in any given volume of sediment was reduced (Delmont et al., 2011, Barnes and Turner, 2016). In this study, diversity metrics from the 2.7 g whole-sediment samples are presented descriptively because those extractions used an RNA-kit protocol that differed from the 0.25 g and sieved treatments. Metabarcoding also recovered far more nematode ASVs than the 21 species recorded morphologically from the same Breydon Water sample, including several rare or cryptic taxa. Smaller volumes (0.25 and 2.7 g) tended to be dominated by a few abundant ASVs, whereas the 10 g extractions displayed flatter k-dominance curves, indicating improved evenness and better recovery of rare taxa. Among primers, G18S yielded the highest richness across all volumes and out-performed TAR, NEM and JB3, especially at 10 g. Studies targeting these rarer, larger taxa may thus require sieving or enrichment steps to ensure more comprehensive representation. Extraction volume (e.g., 0.25 g, 2.7 g, or 10 g) further influences ASV or OTU recovery, with larger samples generally yielding higher diversity but also incurring greater costs in reagents and processing time (Gielings et al., 2021). Balancing the trade-offs between sample size, laboratory expenses and the target organisms of interest is essential for designing robust metabarcoding studies (Lafferty, 2024). In some cases, collecting multiple smaller replicates can also improve coverage by accounting for spatial heterogeneity in sediments (Brannock and Halanych, 2015). Ultimately, the chosen combination of sieving methods, extraction volumes and replication should align with the specific research objectives whether prioritising microfauna, macrofauna, or an inclusive survey of all size classes.

Beyond standard approaches, the widely used 18S, 28S and COI genes offer alternative routes for improving species-level resolution. For instance, the DM568 primer exploits the relatively faster evolutionary rate of 28S, thereby enhancing taxonomic discrimination among closely

related eukaryotic lineages (Derycke et al., 2010, Machida and Knowlton, 2012, Chaudhary and Singh, 2013, Shylla et al., 2013, Kounosu et al., 2019). In this study, the primer detected Platyhelminthes and cercozoan ASVs not recovered by other 18S datasets, contributing to finer taxonomic resolution. Nevertheless, databases contain less 28S sequences than 18S, which can limit the accuracy of taxonomic assignment (Watanabe et al., 2023). The JB3 primer set amplifies a segment of the COI gene that exhibits a higher mutation rate than 18S. The resulting rate of sequence divergence was sufficient to serve as a taxonomic barcode for closely related species (Derycke et al., 2005, Tytgat et al., 2019) Nonetheless, many reads remained unclassified, likely due to the primer's amplification bias and the limited availability of COI references at higher taxonomic ranks (Tytgat et al., 2019). Both 28S and COI can therefore complement 18S by providing finer resolution in certain taxa but depend on more complete sequence databases.

Several primers emerged as particularly effective: G18S and TAR offered broad eukaryotic coverage, NEM provided moderate resolution for nematodes, yet a standardised sample size of 100 reads yielded fewer detections than morphological methods, indicating potential underestimation. Additionally, the DM568 primer set contributed extra metazoan and protozoan lineages not captured by 18S. ITS amplified relatively few fungal sequences and ARF substantially enriched archaeal diversity. In contrast, E1391, D512, JB3 and VRAIN demonstrated limited utility in these datasets and were therefore excluded from further application. Collectively, these observations highlight persistent gaps in marine reference libraries (Porazinska et al., 2009, Sapkota and Nicolaisen, 2015) and underscore the value of a multi-primer strategy (Tytgat et al., 2019, Fonseca et al., 2022). Continuing to refine primer designs, integrate 18S with alternative markers such as 28S or COI and tailor sample preparation methods to specific research goals will further improve the accuracy and taxonomic breadth of metabarcoding in complex benthic ecosystems.

Chapter 4:

Impacts of Metal Pollution on Bacterial Community Composition, Diversity and Tolerance in Estuarine Sediments

4.1 Introduction

Metal pollution remains a pervasive threat to aquatic ecosystems worldwide, driven by sources such as mining, agriculture and industrial discharge (Järup, 2003, Alloway, 2012). Elevated concentrations of heavy metals, particularly copper, often reduce bacterial diversity and favour metal-tolerant phyla, including Proteobacteria (Pseudomonadota) and Acidobacteriota (Sazykin et al., 2023, Yin et al., 2024). By contrast, contamination levels that exceed background but remain below recognized toxic thresholds can, in some cases, amplify microbial abundance or even increase diversity, hinting at a possible dose-dependent response. Detecting these patterns in estuaries is challenging, however, because metal impacts may be masked by overlapping factors such as salinity, nutrient gradients and sediment composition (Di Cesare et al., 2020).

In southwestern England, Ogilvie & Grant (2008) examined 11 estuarine sites across the Hayle, Fal and Kingsbridge systems, each sampled with double replicates, finding that copper tolerance correlated strongly with porewater Cu concentrations. They identified a threshold around 15 µg/L porewater Cu above which Cu tolerance of the microbial community clearly increased markedly. Although terminal restriction fragment length polymorphism (T-RFLP) revealed limited changes in taxonomic composition, the physiological evidence underscored copper's ecological importance at those sites.

A number of studies reported relationships between bacterial community structure and metal concentrations. In "heavily polluted" marine sediments in Croatia, Di Cesare et al. (2020)

studied 14 sites in Pula Bay-Croatia with elevated Hg (up to 11.5 µg/L), Pb (395 µg/L) and Zn (931 µg/L) far above average shale levels of 0.18 µg/L, 20 µg/L and 95 µg/L, respectively, yet they concluded that depth and nutrients overshadowed direct metal effects, based on a small dataset and correlation-based analysis. Chen et al. (2022) found that low metal loads (e.g. copper only about twice local background) in nearshore East China Sea sediments coincided with certain community shifts, but those conclusions relied on ten sites spanning a strong nearshore–offshore gradient where metals correlated closely with water depth and salinity. Zhao et al. (2022) found in the Beibu Gulf that only arsenic exceeded average shale benchmarks, while other metals remained at or below these reference values. Their data showed that metal concentrations increased from nearshore to offshore, likely because of finer-grained sediments accumulating at greater depths. This pattern differs from what one would expect under a true pollution gradient, where nearshore sites typically have higher contamination. Instead, it indicated that the apparent offshore increase in metals mainly reflects natural depositional processes rather than genuine pollution. Du et al. (2022) identified arsenic and cadmium as key drivers of microbial differences in Bohai Sea sediments but did not address whether these concentrations exceeded recognised toxicity thresholds. Their data indicated that As varies from 27.2 to 35.4 ppm (about 2-2.5 times average shale) and Cd from 0.221 to 0.346 ppm (less than twice average shale), a relatively modest range considering the sediments were muddy (as suggested by high Fe). This narrow variation implies that average shale offers a reasonable benchmark for assessing metal loads, yet the study's data appears to lack replicate microbial community sampling at each site, which further limits the strength of its conclusions. These single-event, low-replication surveys often rely on correlation alone, making it difficult to disentangle mild metal elevations from broader estuarine gradients.

However, the metal concentrations reported by these studies were in most cases only slightly elevated above background concentrations; numbers of sites sampled was relatively small and metal concentrations were often correlated with other environmental variables. A more detailed assessment of these publications was given in the discussion. Besaury et al. (2014) investigated two coastal sites in Chile, each exposed to long-term, untreated discharges of copper-rich mining waste. One site showed extremely high total Cu (1,410-1,600 µg/g), whereas the other remained near background (65 µg/g). Although qPCR and RT-qPCR revealed archaeal and bacterial resilience at the contaminated site, only a single sampling event and one core per site were used, leaving possible confounders (organic matter, salinity) unaddressed. Like other

limited-replication studies, this approach could not isolate whether metal-driven changes would hold across broader gradients or under multi-factor scenarios. Similarly, Yin et al. (2022) noted (19-34 µg/g), Chen et al. (2022) observed (15-46 µg/g), Peng (2024) documented (16-37 µg/g) and Yi et al. (2021) recorded (28-81 µg/g). Such levels may not suffice to reveal thresholds for deeper community shifts. Thus, while these works hint at metal-driven effects, the combination of low metal elevations, correlation-based methods and confounding environmental variables makes it difficult to determine how far metals truly dictate microbial composition in estuarine ecosystems.

This study therefore investigates bacterial community responses along an exceptionally wide copper gradient (porewater Cu up to 90-fold, sediment extractable Cu up to 400-fold differences) in southwest-UK estuaries, using 16S rRNA amplicon sequencing to test whether severe copper loading produces clearer compositional thresholds than previously reported under milder contamination.

4.2 Methods

4.2.1 Sampling and DNA extraction

Three replicate samples were obtained from intertidal areas within estuaries in Southwest England, across 12 locations as detailed in (Udochi, 2020) and 34 locations as outlined in **Chapter 2**, along with two replicates from Breydon Water, Norfolk. The DNA extraction methods were described in **Chapter 2**.

4.2.2 PCR and sequencing

Hypervariable regions V3 and V4 of the 16S rRNA gene were amplified using modified versions of the primers 515F and 926R (Turner et al., 1999). The forward primer and the reverse primer contain an Illumina adaptor, 8 base barcode, 0-7 bases of length heterogeneity spacer and primer sequence (Fadrosh et al., 2014). Primer Specifications were given in more detail in **Chapter 3**. Primers were supplied by Sigma Aldrich Company limited.

Polymerase chain reaction (PCR) was conducted in a volume of 20 µl, including 10 µl of PhusionTM Flash master mix (Thermo Scientific, UK), 1 µl of forward primer and 1 µl of reverse primer (final concentration of primers: 100 µM), 7 µl of ultrapure sterile water (MilliQ water) and 1 µl of DNA, with concentrations varying from nearly zero to 9.3 ng/µl, was detailed further in **Chapter 2**. any reactions yielding no visible band on the gel or a concentration of 0 ng/µL were excluded, as outlined in **Chapter 3**. The amplification of the DNA templates was carried out in a Veriti™ HID 96-Well Thermal Cycler, 0.2mL system (Applied Biosystems, UK) in either 0.2mL x 96 well plates or 0.2 ml PCR tubes, depending upon the number of samples be amplified. The amplification consisted of an initial denaturation at 98°C for 10 minutes, followed by 35 cycles of denaturation at 98°C for 30 seconds; annealing at 67 °C for 30 seconds and extension at 72°C for 30 seconds. This was followed by a final extension at 72°C for 5 minutes, before being held at 4°C. Primers were supplied by SigmaAldrich Company limited.

The PCR products were visualised using gel electrophoresis as detailed in **Chapter 2**. Primers 515F and 926R amplified a product that is 411 bp. The sequencing adapters, barcodes and length heterogeneity spaces increased this length to approximately 550 bp. The PCR products were purified using Aline Biosciences PCRClean DX kit (Aline Biosciences, Woburn, USA) following the manufacturer protocol except that the ratio of bead suspension to PCR product which was 1.8:1 1st run, 1:1 2nd and 3rd runs and 0.7:1 at last run. This ratio was altered in the light of experience to improve removal of free primer adaptors and adaptor dimers, which were not fully removed when the ratio in the manufacturer's protocol was used (Quail et al., 2009). The PCR products quantification followed methods described in **Chapter 2**.

Sequencing of the purified 16S rRNA was performed at the Earlham Institute, Norwich, UK, using a pre-made library pool on a single lane of Illumina MiSeq with 300 bp paired-end sequencing in June 2021. This was followed by sequencing on a NovaSeq 6 000 flow cell with 250 bp paired-end in February 2022 and February 2023 and on a NextSeq 1 000 with 300 bp paired-end in April 2023. Sequencing of 16S rRNA amplicons yielded up to 399 598 reads per sample. A small number of samples yielded only very small numbers of reads ranged between 0 and 858 reads. When these were excluded the mean number of reads per sample was 21 601 on MiSeq; 226 438 on NovaSeq 6000 and 71 478 on NextSeq 1 000. This removal of low count samples maintained the integrity of the dataset and enhanced the accuracy of subsequent

analyses. The excluded samples included (HR2, HR3 and SJ2) from the 12-site dataset; (JB2, LF2 and RC1) from the 34-site dataset; and (TRD_LB2 and RD_Bag) from the experimental dataset. outlier samples (HA_Bag and BW_1_1) were also removed due to their clear separation from all other samples in the NMDS ordination.

Bioinformatic analysis was carried out initially using Mothur software, version v.1.45.3 (www.mothur.org) (Schloss et al., 2016) and sequences were aligned against the SILVA database (Pruesse et al., 2007). Subsequent analysis was carried out using the LotuS2 pipeline (Özkurt et al., 2022), with taxonomic assignment of sequences using KSGP version 1.0 (Grant et al., 2023). Bacterial ASVs were identified using furthest neighbour clustering, based on sequence similarity thresholds set during the clustering process.

4.2.3 Positive control

Microbial community standards (ZymoBIOMICS, Zymo Research, Irvine, CA, USA) were employed as positive controls to ensure the accuracy and reliability of the PCR and sequencing processes. Each of the expected ASVs was correctly identified and remained as a single ASV, all assigned to the species level. However, there was some variation in their relative abundances when compared to the standard, indicating minor discrepancies in abundance (**Table 4.1**).

Table 4.1 Comparison between the expected theoretical composition of 16S reads for the ZymoBIOMICS Microbial Community Standard and the proportion of reads observed in this study. The theoretical composition was sourced from the ZymoBIOMICS manual (Zymo Research. ZymoBIOMICS™ Microbial Community DNA Standard Manual).

Bacterial Sp.	Theoretical Composition %	This study Composition %	standard deviation
Bacillus subtilis	17.4	20.9	2.5
Enterococcus faecalis	9.9	12.0	1.5
Escherichia coli	10.1	1.8	5.8
Lactobacillus fermentum (Limosilactobacillus)	18.4	31.7	9.4
Listeria monocytogenes	14.1	16.3	1.6
Pseudomonas aeruginosa	4.2	0.6	2.5
Salmonella enterica	10.4	2.2	5.8
Staphylococcus aureus	15.5	14.4	0.8

4.3 Results

4.3.1 Phylogenetic Analysis of Bacterial Communities

A detailed assessment of phylogenetic relationships was conducted to provide essential background for examining how bacterial taxa responded to varying levels of metal pollution. The phylogenetic tree constructed using the 515F 16S rRNA dataset showed a highly diverse bacterial community (**Fig. 4.1**). Out of a total of 120,668 ASVs, all were successfully annotated at the phylum level, with no unclassified branches observed in the phylogenetic tree. Only 20 ASVs (0.03%) lacked domain-level classification and were excluded from domain-specific analyses, highlighting the comprehensive coverage of the reference database used in this analysis. Bacterial ASVs dominated of the dataset, accounting for the vast majority of ASVs, with Archaea and Eukaryota made up 5.6% and 0.5% of the total respectively. This contrasted with the archaeal phylogenetic analysis presented in Chapter 5, where unclassified branches were more prevalent. The bacterial ASVs demonstrated higher levels of taxonomic resolution, further emphasising the robustness of the bacterial classifications in this study.

The refined bacterial phylogenetic tree (**Fig. 4.2**), after removing non-bacterial ASVs, revealed extensive branching within the dominant phyla Pseudomonadota and Bacteroidota, which together accounted for roughly half of the bacterial ASVs. Pseudomonadota contained the highest number of branches, while Bacteroidota showed a slightly less complex topology. The phyla Planctomycetota and Acidobacteriota formed comparatively compact clustered with markedly fewer branches, consistent with their lower ASV counts. A distinct clade composed of ASVs lacking domain-level classification (medium blue in **Fig. 4.2**) was also present.

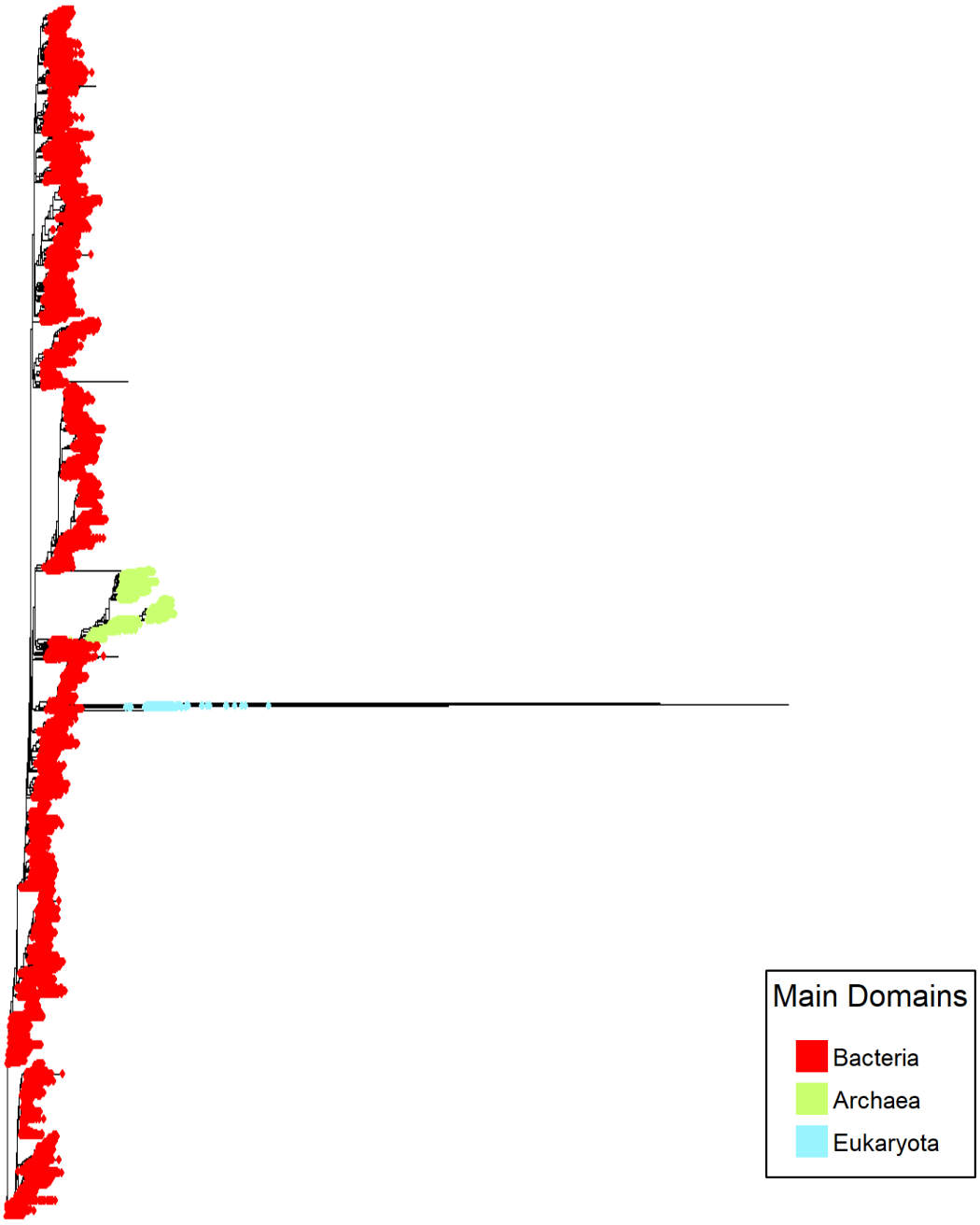


Fig. 4.1 Overview of a phylogenetic tree constructed using 515F 16S rRNA data across all datasets. The tree includes all ASVs and highlights the presence of Bacteria (red), Archaea (green) and Eukaryota (blue). Taxonomic assignments were performed using the KSPG database.

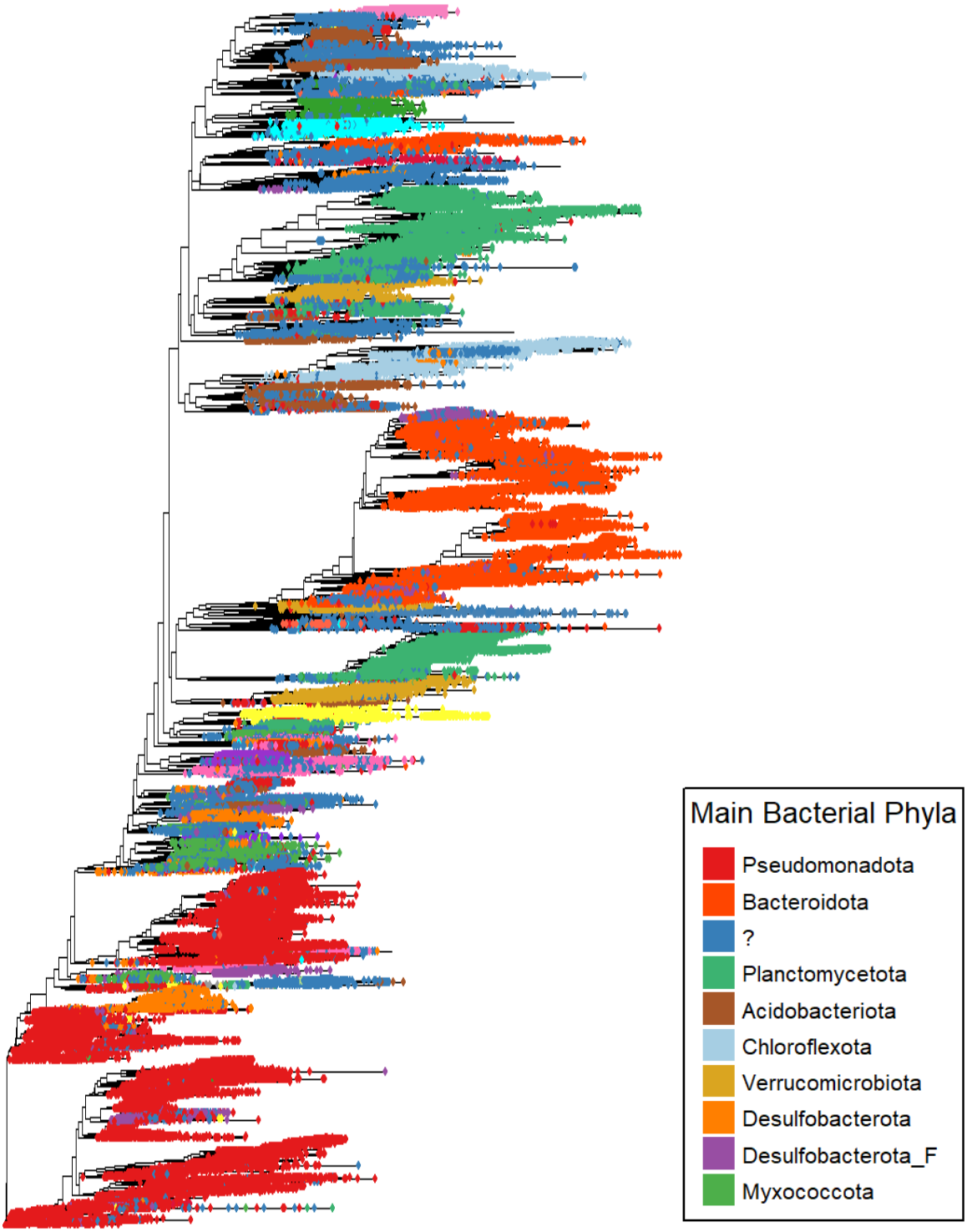


Fig. 4.2 Phylogenetic representation of bacterial phyla derived from 515F 16S rRNA data, following the removal of non-bacterial branches. The main phyla identified are Pseudomonadota (red), Bacteroidota (dark orange), Planctomycetota (green) and Acidobacteriota (brown), with unknown phyla shown in medium blue. Taxonomic assignments were performed using the KSPG database.

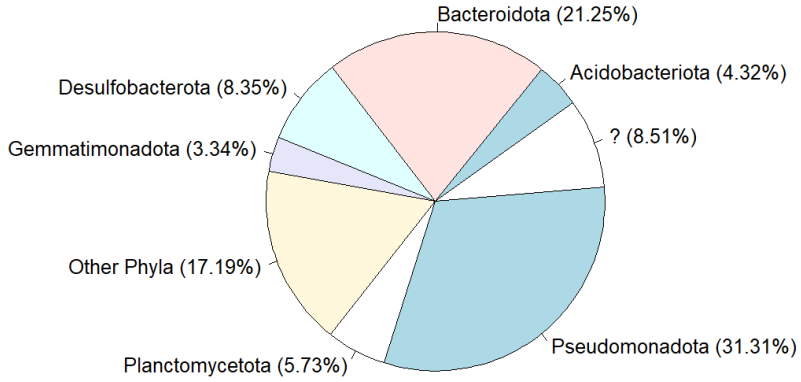
4.3.2 Taxonomic Composition

To address narrower versus broader metal contamination gradients, subsequent analyses drawn on both a 12-site dataset focused on the Fal and Hayle estuaries and a 34-site dataset including samples from several other clean and moderately contaminated estuaries. This design enabled a multi-scale exploration of how pollution influenced bacterial communities, providing a comparative framework for later sections. The dominant bacterial phyla, Pseudomonadota and Bacteroidota, together accounted for nearly half of the 36 million reads, reflecting their consistent abundance across all datasets (**Fig.4.3 A** and **B**). Pseudomonadota alone contributed over a third of the total reads, while other phyla such as Planctomycetota, Desulfobacterota and Acidobacteriota were less prevalent, each representing only about 4 - 8% of the community.

Assessment of bacterial ASVs revealed a nuanced picture of diversity, as differences between read abundance and ASV richness highlighted the complexity of community structure (**Fig. 4.3 A** and **4.4 A**) and (**Fig. 4.3 B** and **4.4 B**). Pseudomonadota and Bacteroidota exhibited high read abundance but lower ASV richness, while Actinomycetota and Verrucomicrobiota had more evenly distributed reads across ASVs.

Comparisons between the KSPG and SILVA databases showed a similar predominance of bacterial reads at 96.2% and 97.5%, respectively, with archaeal reads consistently around 2.4%. Pseudomonadota and Bacteroidota remained the dominant phyla in both datasets, although subtle differences in the relative ordering of Desulfobacterota and Acidobacteriota (**Fig. 4.4 A** and **B**). Unclassified phyla constituted 8-12% of reads and ASVs, all of which were classified as bacteria but remained unresolved at the phylum level.

A (KSPG)



B (SILVA)

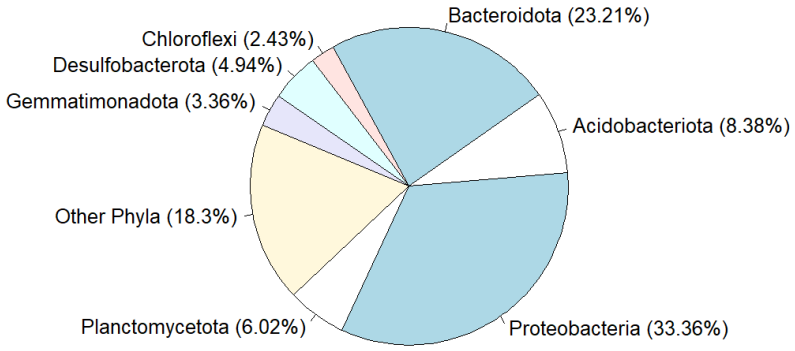
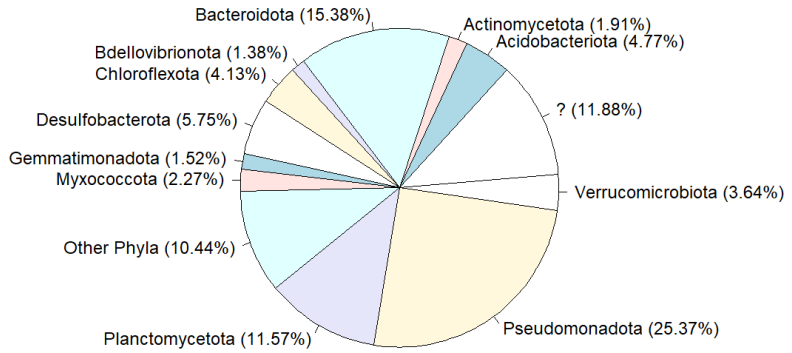


Fig. 4.3 Distribution of reads across phyla for all datasets after excluding non-bacterial reads. Based on the total number of reads, taxonomic assignments were made using the KSPG database for **(A)** and the SILVA database for **(B)**. Phyla making up less than 2.2% of the total are combined into "other phyla." The '?' symbol indicates that taxonomy was not resolved at the phylum level by LCA classification based on matches in the assigned databases.

A (KSPG)



B (SILVA)

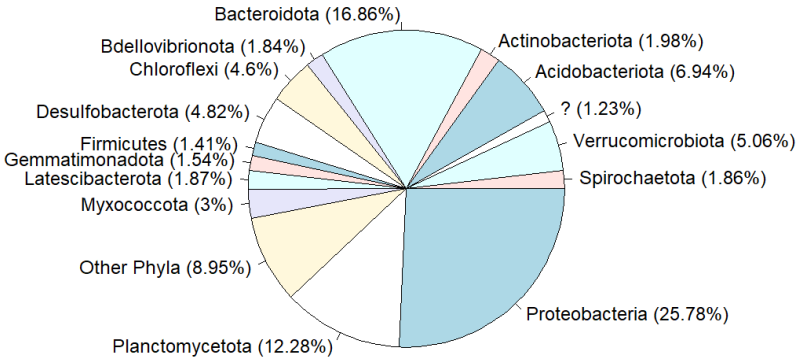


Fig. 4.4 Distribution of ASVs across phyla for all datasets after excluding non-bacterial reads. Taxonomic assignments were made using the KSPG database for **(A)** and the SILVA database for **(B)**. Phyla making up less than 1.2% of the total are combined into "other phyla." The '?' symbol indicates that taxonomy was not resolved at the phylum level by LCA classification based on matches in the assigned databases.

4.3.3 NMDS Analysis of Replicate Consistency and Similarities Between Datasets

Analyses of replicate consistency and cluster patterns illustrated how bacterial communities responded to contamination gradients within both the 12-site and 34-site datasets. NMDS and ANOSIM showed that replicate samples from each site clustered closely. The **Fig. 4.5** also showed partial overlap among the 12-site, 34-site and experimental datasets. The 12-site dataset captured a narrower contamination gradient, whereas the 34-site dataset covers broader spatial and metal variability. By contrast, pilot and experimental samples plotted apart from field-collected data. Polluted sites occupied broadly similar positions within both datasets,

although subtle differences emerged among samples from the same location (see **Section 4.3.7**).

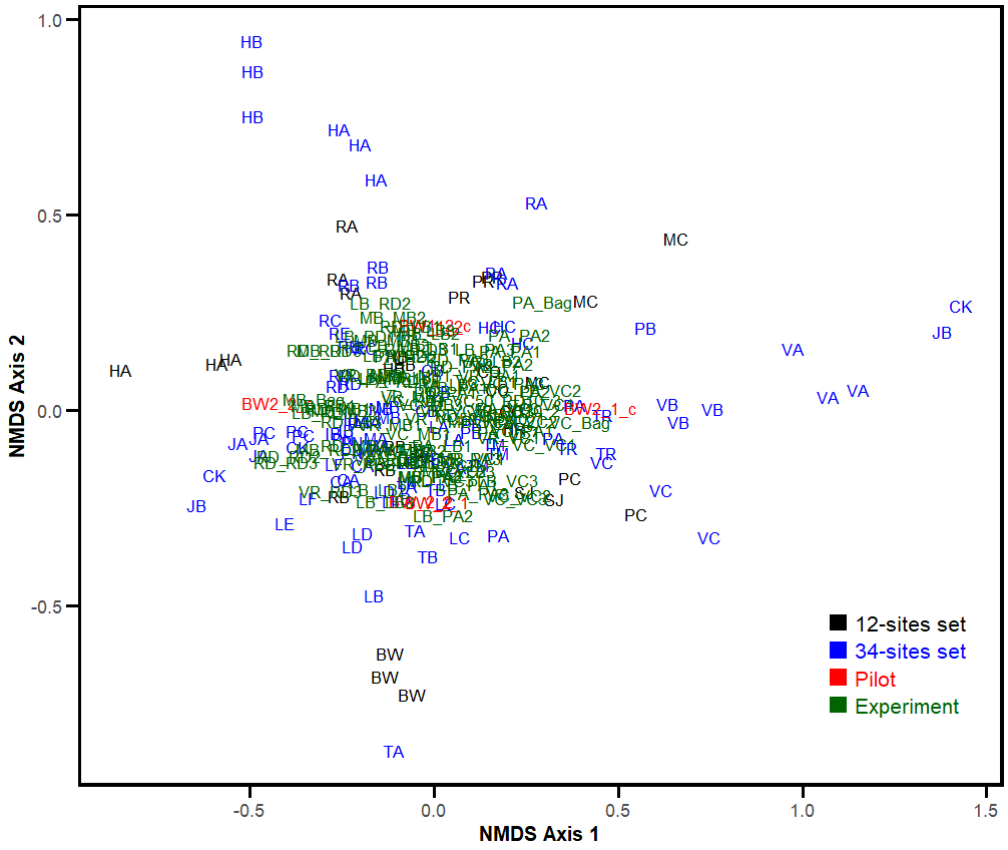


Fig. 4.5 Sample ordination using Non-metric Multi-dimensional Scaling (NMDS) of Bray-Curtis similarity matrix of bacterial abundances, based on square root transformed data, derived from all datasets; Pilot, 12-sites, 34-sites and experiment. Each two-letter sample label represents site, colours represent the different datasets. Outlier samples were excluded as described in the methods section. All datasets performed with 0.1% ASV pruning. Data processed using LotuS2 in conjunction with the KSPG database. Site codes as in **Table 2.1**.

When considering the 12-site and 34-site datasets (**Fig. 4.6**), replicates from each site clustered closely together. Sites with high contamination levels, such as Hayle (HA, HB) and Restronguet Creek (RA), plotted towards the upper region of the ordination, while cleaner sites, such as Breydon Water (BW) and the Avon River sites (VA, VB, VC) laid towards the lower or right areas. Sites in intermediate positions reflected a gradient of contamination levels.

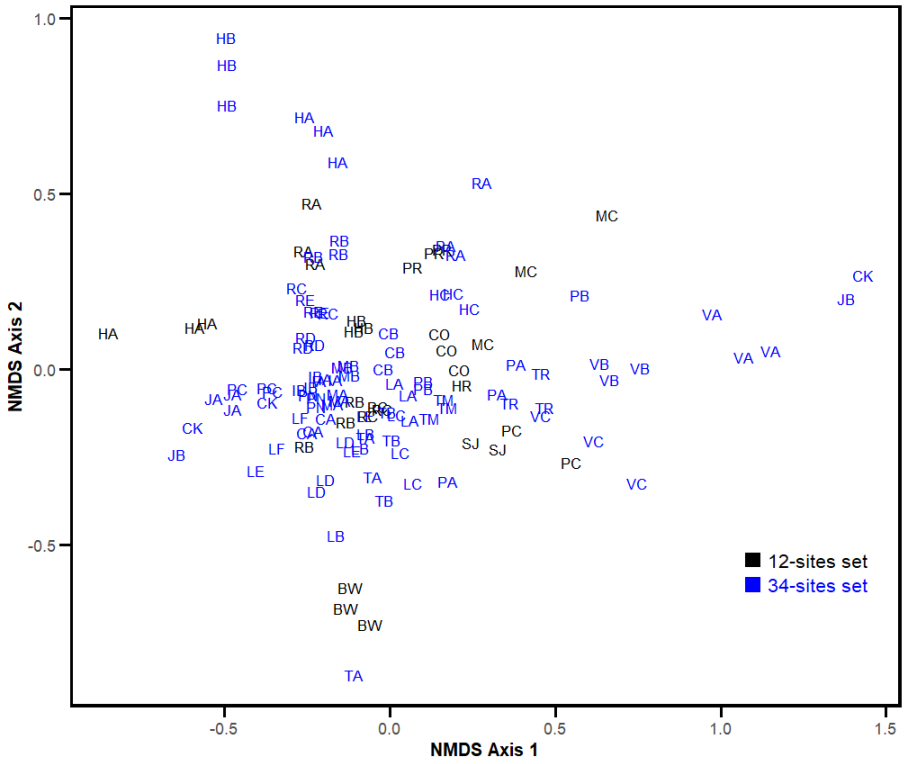


Fig. 4.6 Sample ordination using NMDS of Bray-Curtis similarity matrix of square root transformed bacterial abundance data, showing the 12 and 34 site datasets. Each two-letter sample label represents site, colours represent the different datasets. All datasets performed with 0.1% ASV pruning. Data processed using LotuS2 in conjunction with the KSPG database. Site codes as in **Table 2.1**.

In the 12-site dataset (**Fig. 4.7 A**), samples from Breydon Water (BW) clustered at the bottom of the plot and were separated from all Southwest sites. Excluding BW samples (**Fig. 4.7 B**) resolved the ordination, with polluted sites (HA, HB, RA) clustering to the left, distinct from cleaner sites. Replicates remained closely grouped. The BW samples clustered closely together in the plots, showing similar community compositions among themselves.

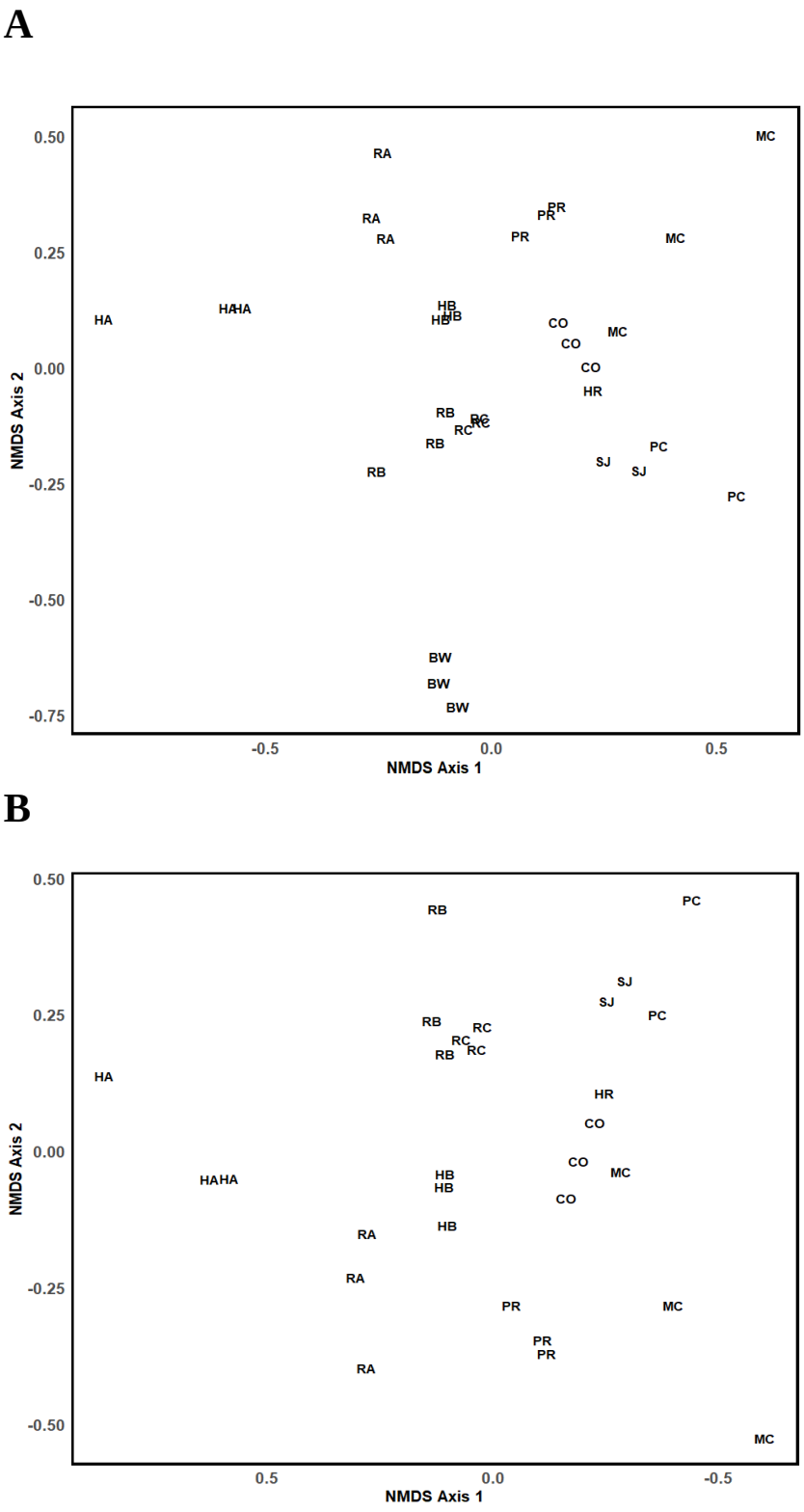


Fig. 4.7 Sample ordination using NMDS of Bray-Curtis similarity matrix of square root transformed bacterial abundance data from the 12-site dataset. **A)** includes Breydon Water sites (stress value = 0.18), while **B)** excludes them (stress value = 0.16). Each two-letter sample label represents site, colours represent the different datasets. All datasets performed with 0.1% ASV pruning. Data processed using LotuS2 in conjunction with the KSPG database. Site codes as in **Table 2.1**.

The NMDS plot for the 34-site dataset (**Fig. 4.8**) showed consistent clustering of replicates, with minimal within-site variability in most cases. Sites with the highest pore-water copper concentrations, including Hayle (HA, HB) and Restronguet Creek (RA, RB, RC, RD), were positioned towards the upper and left regions of the ordination. Conversely, cleaner sites, including River Avon (VA, VB, VC), were positioned towards the far right of the plot, clearly separated from the polluted sites.

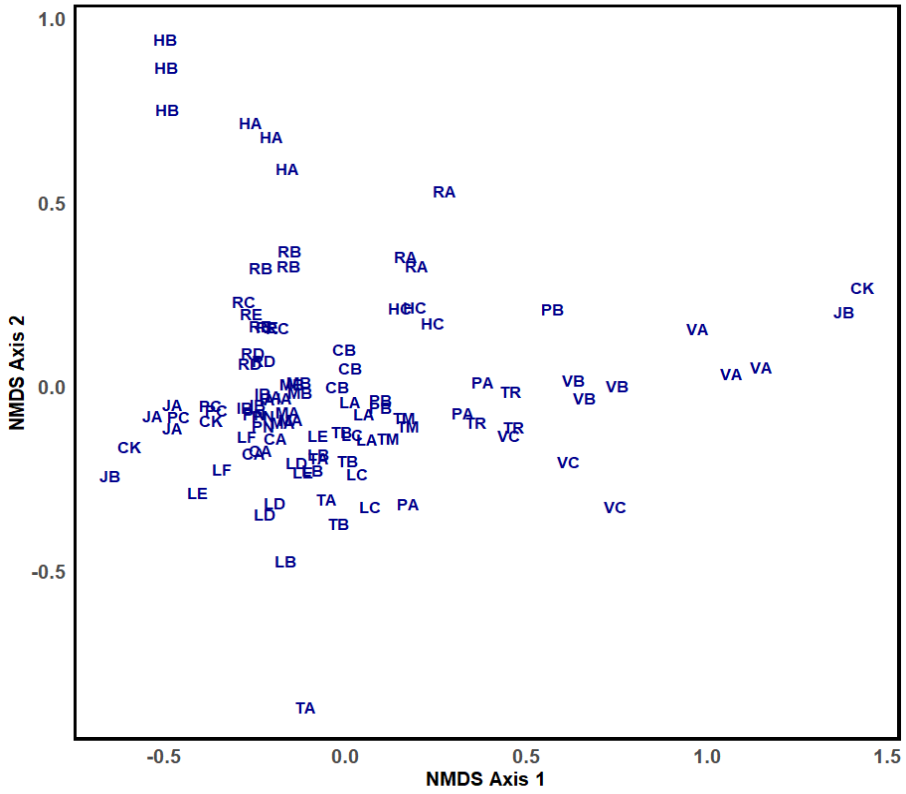


Fig. 4.8 Sample ordination using NMDS of Bray-Curtis similarity matrix of square root transformed bacterial abundance data from the 34-site dataset (stress value = 0.14). Each two-letter sample label represents site. All datasets performed with 0.1% ASV pruning. Data processed using LotuS2 in conjunction with the KSPG database. Site codes as in **Table 2.1**.

For the experimental samples, (**Fig. 4.9**) illustrated the differences in community composition between treatments. The first part of the treatment name indicated the inoculum source, with "bag" samples reflecting the initial community composition. The second part identified the defaunated sediment source in each replicate, as distinguished by colour. Treatments clustered more strongly by the sediment's site of origin rather than by inoculum source. Consistently,

NMDS ordination indicated a stronger effect of sediment source on community structure than inoculum source, which was supported by ANOSIM results in **Section 4.3.5**. The Sediment Source factor yielded a considerably higher R-value (0.85) than Inoculum Source (0.16), although both were statistically significant ($p < 0.001$). Clustering patterns followed a gradient of decreasing metal concentration from left to right, with heavily polluted sediments (red) forming one cluster and the cleanest sediments (green) on the opposite side. Helford (LB) and Percuil (PA) samples grouped near Avon sites.

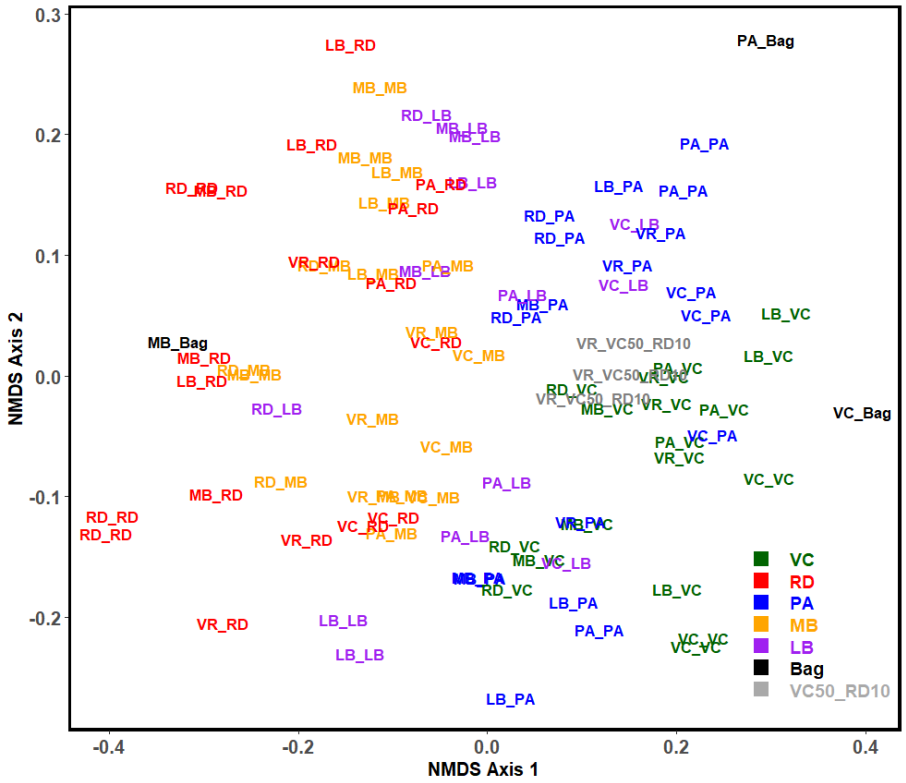


Fig. 4.9 Sample ordination using NMDS of Bray-Curtis similarity matrix of square root transformed bacterial abundance data from the experiment set samples (stress value = 0.17). Colours represent the different test sediments. The first two letters in the sample labels indicate the source of the inoculum, while the second two letters represent the sediment that makes up 75% of the total composition. Outlier samples were excluded as described in the methods section. All datasets performed with 0.1% ASV pruning. Data processed using LotuS2 in conjunction with the KSPG database. Site codes as in **Table 2.1**.

4.3.4 Microbial Diversity Across Pollution Gradients

Rarefaction curves and rarefied ASV counts were generated to standardise sequencing depth across samples in both the 12-site (**Fig. 4.10 A** and **B**) and 34-site datasets (**Fig. 4.11 A** and **B**). To evaluate environmental influence, Spearman's rank correlation was then applied to test the association between PWCu concentrations and bacterial ASV richness. The analysis revealed a significant yet weak negative association between PWCu and bacterial richness; for the 12 sites, rho = -0.23 ($p \leq 0.05$); for the 34 sites, rho = -0.16 ($p \leq 0.01$).

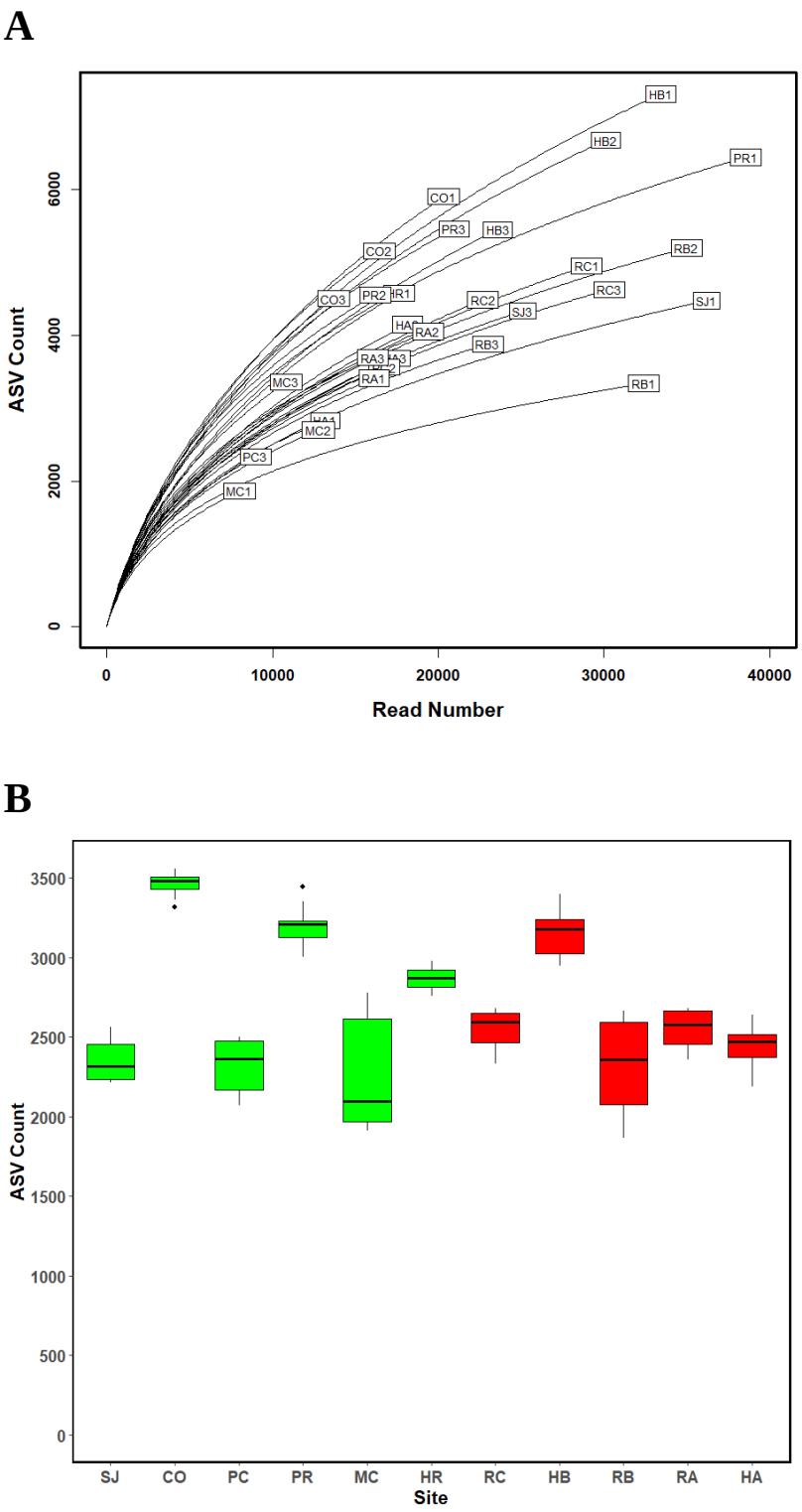


Fig. 4.10 A) Rarefaction curves for bacterial samples for 12 sites, illustrating ASVs richness in relation to sample size. **B)** Box and whisker plot of ASV counts rarefied to a sample size equivalent to the lowest sample size (7 989) across different sites, sites were arranged in increasing order of porewater copper concentrations with sites above and below 20 µg/L indicated in red and green respectively. ANOVA results ($F = 39.8$, $p < 0.001$). Spearman's rank correlation between ASV richness and porewater copper concentration ($rho = -0.23$, $p \leq 0.05$). Site codes as in **Table 2.1**.

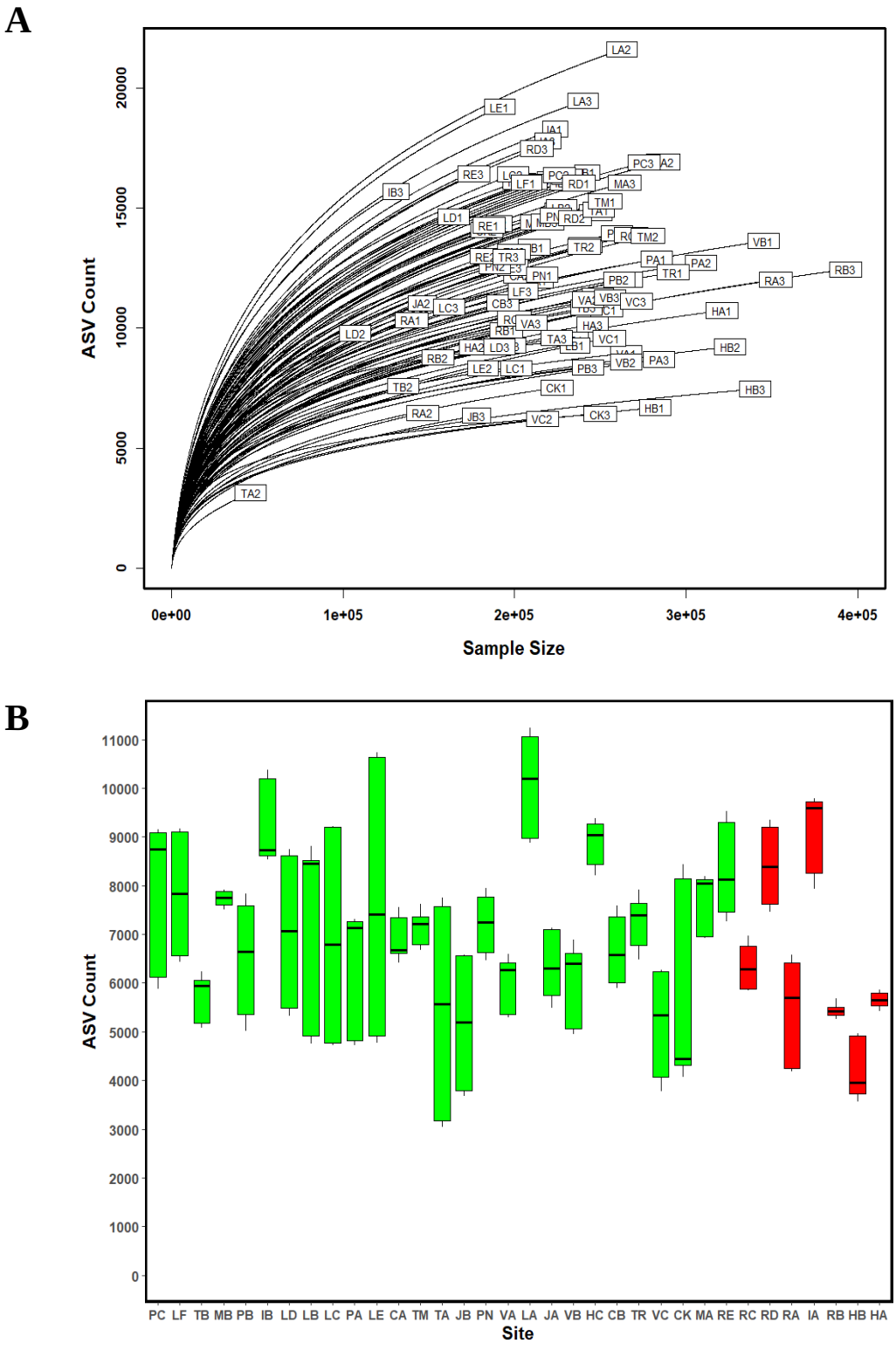


Fig. 4.11 A) Rarefaction curves for bacterial samples for 34 sites, illustrating ASVs richness in relation to sample size. **B)** Box and whisker plot of ASV counts rarefied to a sample size equivalent to the lowest sample size (46 193) across different sites, sites are arranged in increasing order of porewater copper concentrations with sites above and below 20 µg/L indicated in red and green respectively. ANOVA results ($F = 12.8$, $p < 0.001$). Spearman's rank correlation between ASV richness and porewater copper concentration ($rho = -0.16$, $p \leq 0.01$). Site codes as in **Table 2.1**.

In the 12-site dataset, the ANOVA yielded the highest F value (39.8). However, (**Fig. 4.10 B**) does not clearly separate heavily contaminated locations from less polluted ones. In the 34-site analysis, the ANOVA also pointed to strong inter-site differences (F = 12.8), yet some highly contaminated sites displayed richness levels comparable to cleaner areas.

Sample-based accumulation curves (**Fig. 4.12**) were constructed to assess whether ASV richness approached an asymptote as additional sites were included in the survey. The bacterial richness climbed quickly in the 12-site set and added few new ASVs after 25 samples (**Fig. 4.12 A**), while the 34-site curve flattened by 80 samples with a narrower confidence band; both curves leveled off before a full plateau, indicating most common taxa were captured, with rare taxa still under-represented, especially in the smaller survey (**Fig. 4.12 B**).

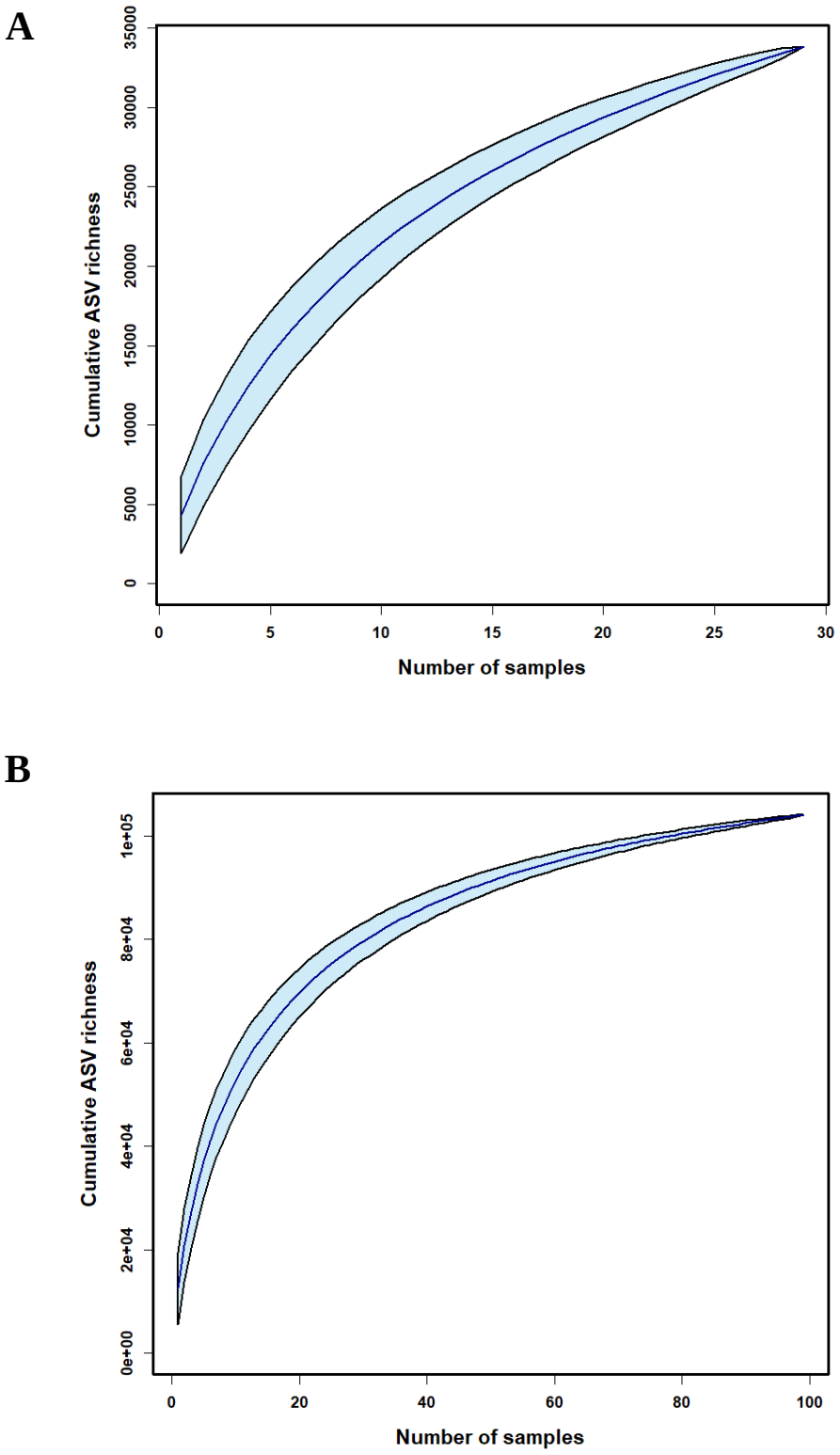
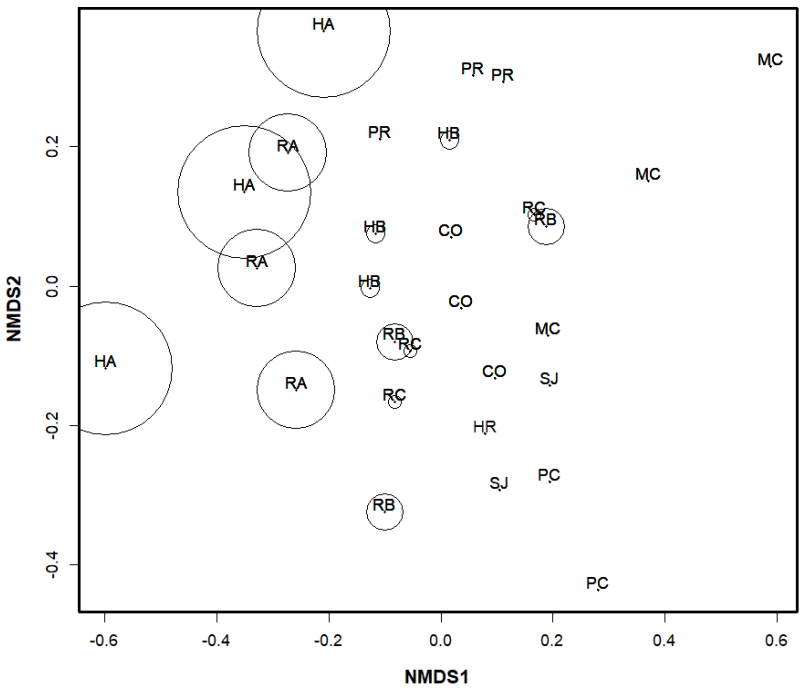


Fig. 4.12 Sample-based species-accumulation curve for 16S rRNA ASVs recovered from **A)** the 12-site and **B)** 34-site bacterial surveys. The solid line represents the mean cumulative richness and the light-blue polygon the 95 % confidence envelope. Site codes as in **Table 2.1**.

4.3.5 Effect of ASV Rare Removal on Analysis

Further filtering was applied by retaining only ASVs with a relative abundance of 0.1% or 1% or higher, ensuring that rare or insignificant taxa were excluded from the analysis. In the 12-site dataset (**Fig. 4.13**), the NMDS ordination that included all ASVs (**Fig. 4.13 A**) showed a wider spread of sample points. Applying a 0.1 % threshold (**Fig. 4.13 B**) reduced the spread and brought replicates from the same site closer together. Increasing the threshold to 1 % (**Fig. 4.14**) further reduced the number of ASVs and compressed the ordination.

A All ASVs Retained



B Pruned (<0.1%)

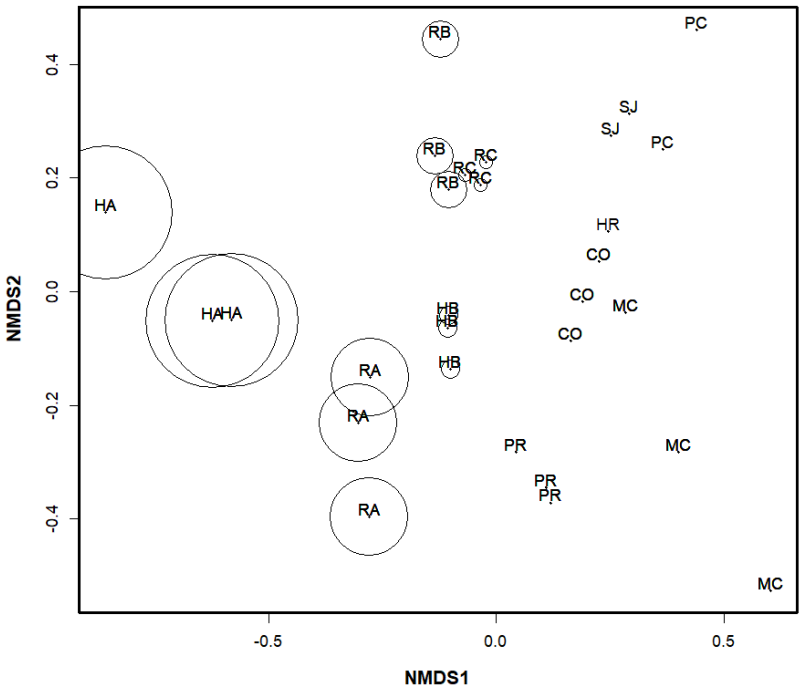


Fig. 4.13 NMDS bubble plots of the 12-site dataset using square root transformed data based on Bray-Curtis similarity of bacterial abundances. It compares ASV counts across all sites under two conditions: **A)** without removing rare ASVs and **B)** with ASVs occurring at less than 0.1% abundance in all samples removed. Stress values are 0.22 and 0.16, respectively. Bubble sizes represent the pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

Pruned (<1%)

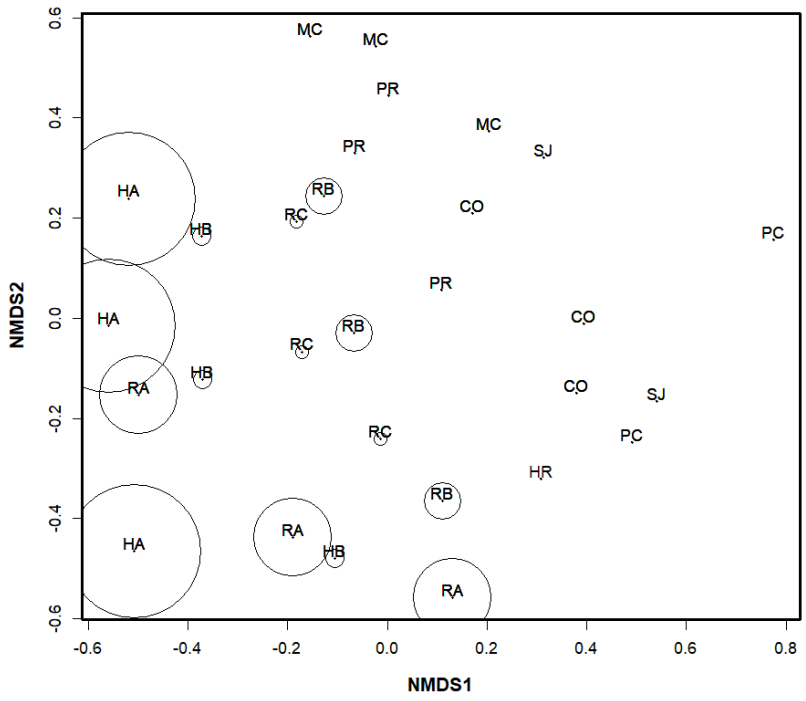
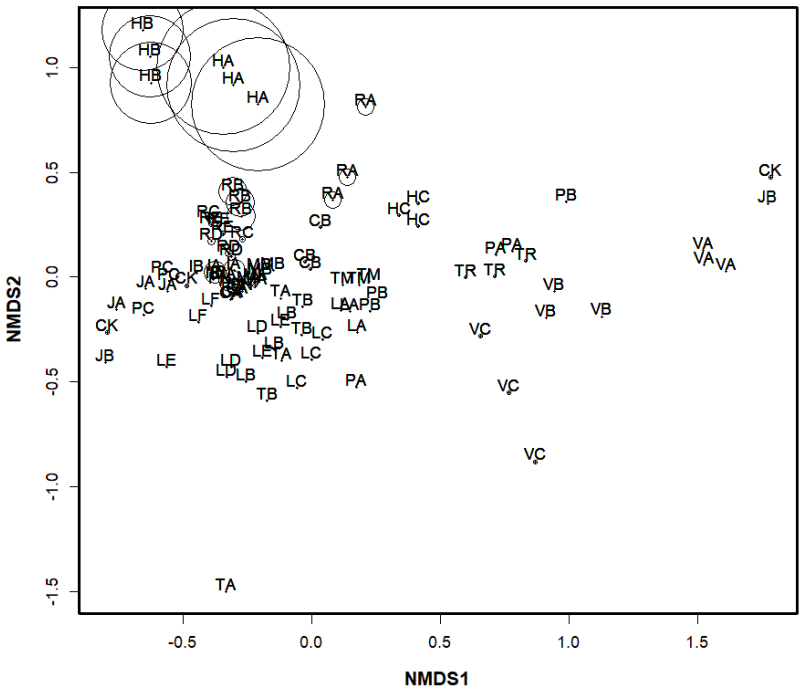


Fig. 4.14 NMDS bubble plot of the 12-site dataset using square root transformed data based on Bray-Curtis similarity of bacterial abundances. It shows ASV counts across all sites with ASVs occurring at less than 1% abundance in all samples removed. Stress value is 0.20. Bubble sizes represent the pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

In the 34-site dataset (**Fig. 4.15**), ordinations produced with no pruning (**Fig. 4.15 A**) and with a 0.1 % threshold (**Fig. 4.15 B**) were similar, although replicate clusters were slightly tighter after pruning. With a 1 % threshold (**Fig. 4.16**), the ordination displayed less dispersion among sites.

A All ASVs Retained



B Pruned (<0.1%)

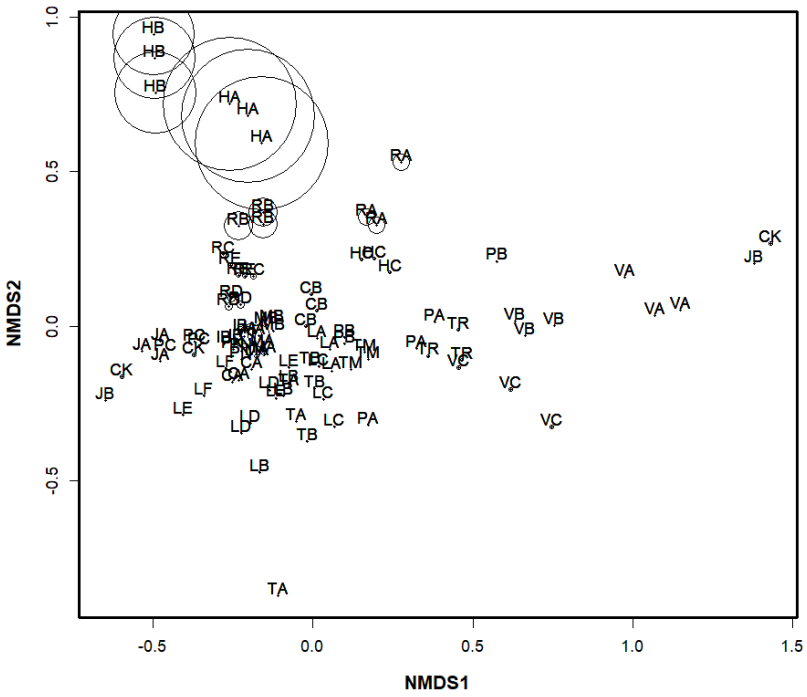


Fig. 4.15 NMDS bubble plots of the 34-site dataset using square root transformed data based on Bray-Curtis similarity of bacterial abundances. It compares ASV counts across all sites under two conditions: **A)** without removing rare ASVs and **B)** with ASVs occurring at less than 0.1% abundance in all samples removed. Stress values are 0.13 and 0.14, respectively. Bubble sizes represent the pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

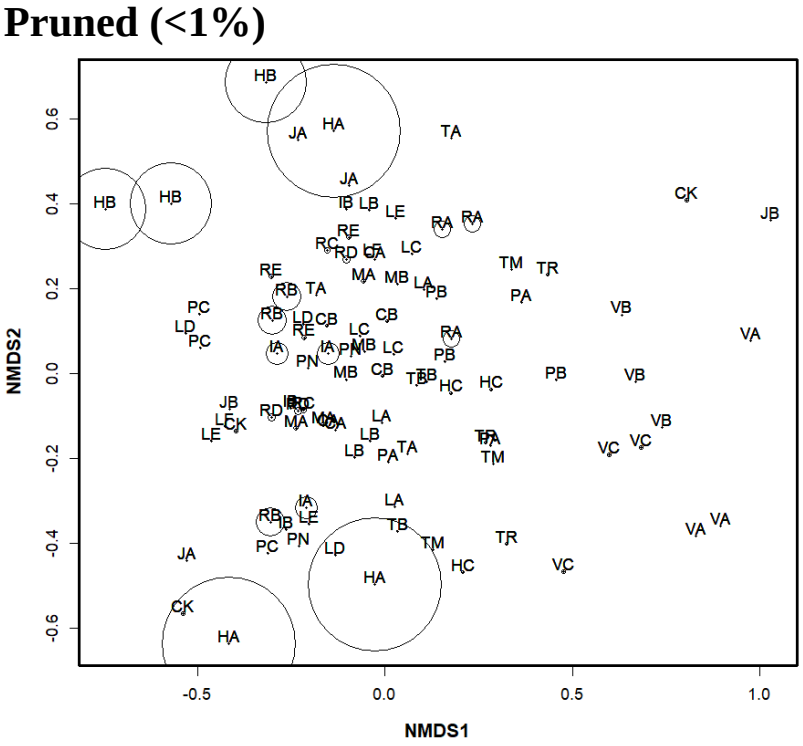


Fig. 4.16 NMDS bubble plot of the 34-site dataset using square root transformed data based on Bray-Curtis similarity of bacterial abundances. It shows ASV counts across all sites with ASVs occurring at less than 1% abundance in all samples removed. Stress value is 0.20. Bubble sizes represent the pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

Results from ANOSIM (**Fig. 4.17**) indicated that R-values in the 12-site dataset peaked at 0.1 % and remained comparable across other thresholds. In the 34-site dataset, R-values for no pruning and 0.1 % were similar, whereas the 1 % threshold gave a lower R-value.

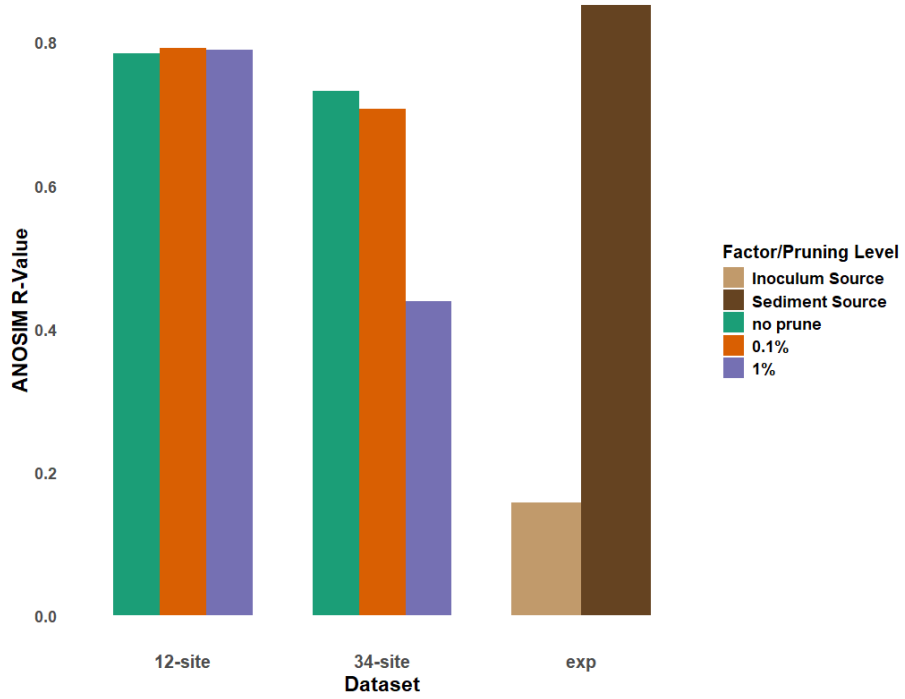
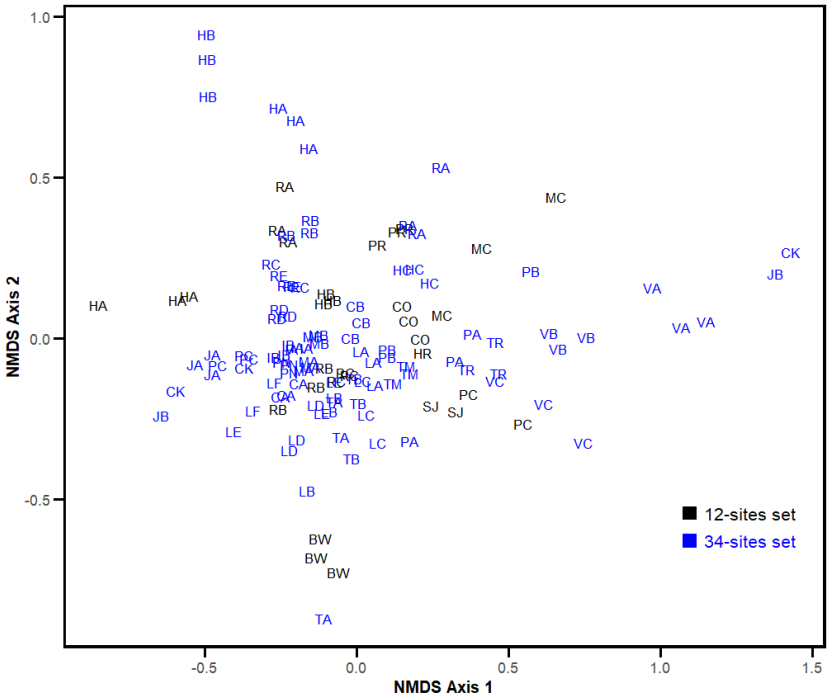


Fig. 4.17 ANOSIM R-values for 12-site, 34-site and experimental datasets generated using the 515F primer under three pruning thresholds (no pruning, 0.1% and 1%). The experimental dataset includes inoculum source and sediment source factors. The results illustrate the degree of dissimilarity between sites, with R statistics quantifying the strength of clustering. R-values > 0.75 indicate very strong differences between groups and all results are statistically significant ($p \leq 0.0001$).

4.3.6 Evaluating Site Grouping Consistency

To compare community-level and phylogenetic ordinations, Bray-Curtis and UniFrac metrics were applied to the bacterial datasets. In the Bray-Curtis ordination (**Fig. 4.18 A**), samples from sites with higher pore-water copper concentrations were positioned nearer each other, whereas samples from lower-copper sites were placed farther away. The UniFrac ordination (**Fig. 4.18 B**) showed less alignment with copper levels, and sample positions did not separate strongly along the contamination gradient. .

A Bray-Curtis



B UniFrac

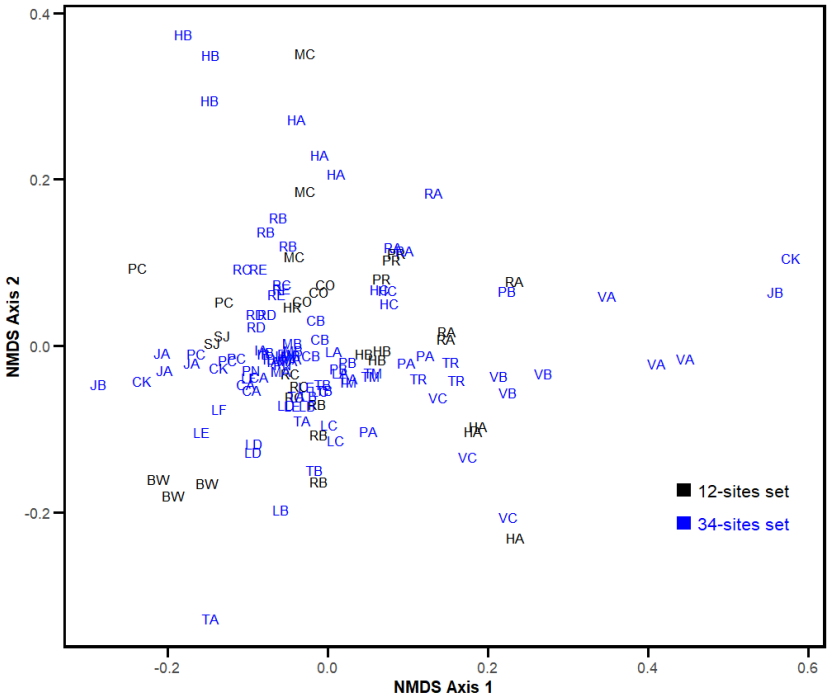


Fig. 4.18 Comparison of **A)** Bray-Curtis and **B)** UniFrac NMDS similarity matrices of square root transformed plots. derived from 12 and 34 site datasets, each two-letter sample label represents site, colours represent the different datasets. No ASVs were pruned for both datasets. Data processed using LotuS2 in conjunction with the KSPG database. Site codes as in **Table 2.1**.

4.3.7 Comparison of Clustering Patterns Between NMDS and PCoA Analyses

To compare the effectiveness of ordination methods for bacterial community composition across pollution gradients, NMDS and Principal Coordinates Analysis (PCoA) were applied. In the 12-site dataset, the NMDS ordination (**Fig. 4.19 A**) separated sites with higher pore-water copper (PWCu) concentrations (HA, HB, RA) from cleaner sites (PC, SJ, CO) along the NMDS1 axis. The corresponding PCoA ordination (**Fig. 4.19 B**) showed the same general pattern, but polluted and clean sites were less distinct. For the 34-site dataset, NMDS (**Fig. 4.20 A**) separated heavily polluted sites (HA, HB, RA, RB) from less polluted ones. The PCoA ordination (**Fig. 4.20 B**) displayed broader trends but less separation among sites with lower PWCu. Hierarchical clustering supported the ordinations. In the 12-site dendrogram (**Fig. 4.21 A**), sites HA, HB and RA formed one cluster, whereas PC, SJ and CO clustered separately. Replicates from each site grouped together. In the 34-site dendrogram (**Fig. 4.21 B**), HA, HB and RA clustered together, while cleaner sites showed more dispersion. Overall, NMDS separated sites along the PWCu gradient more than PCoA. Hierarchical clustering showed a similar grouping pattern.

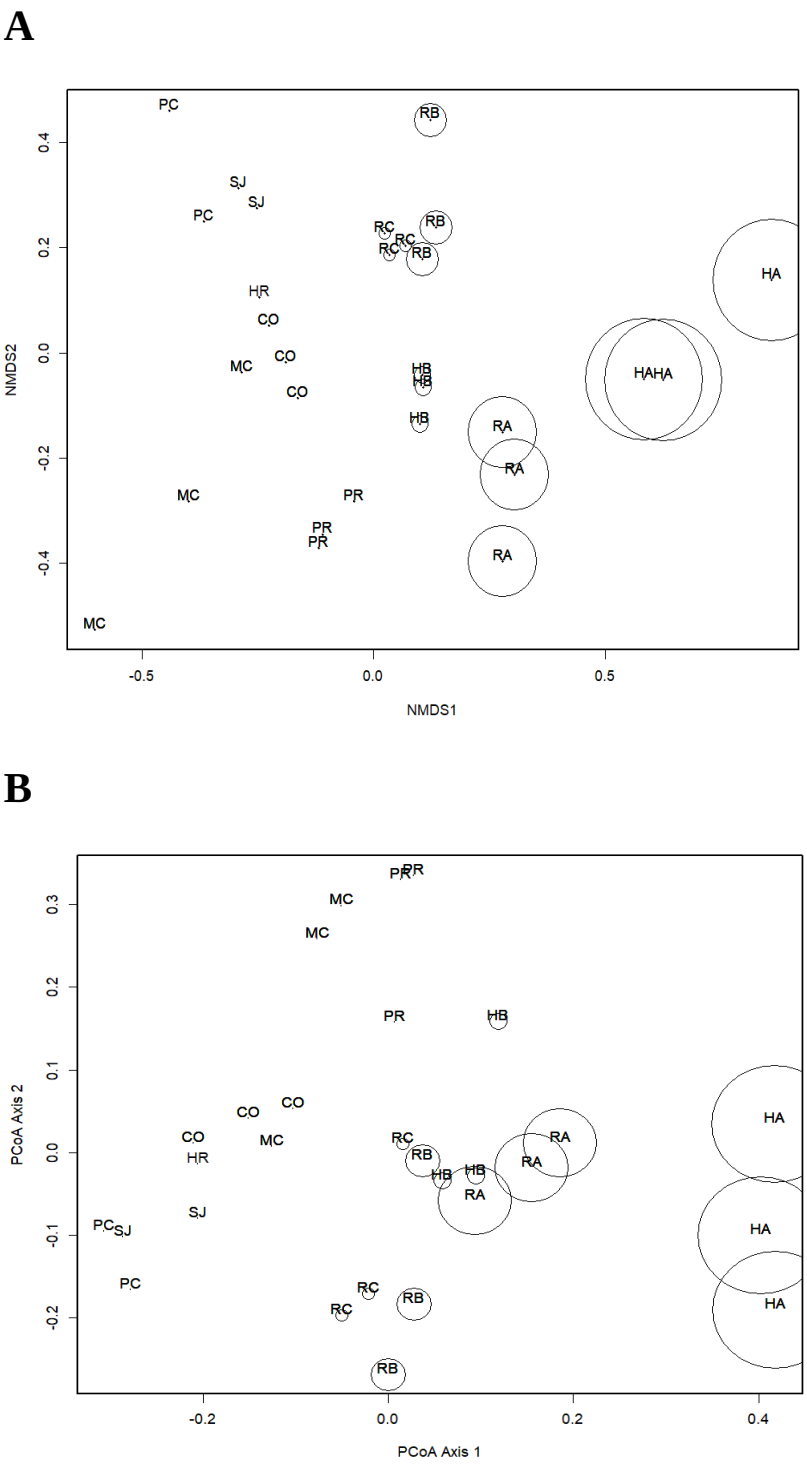


Fig. 4.19 Comparison of **A)** NMDS (Non-metric Multidimensional Scaling) and **B)** PCoA (Principal Coordinates Analysis) on Bray-Curtis distances for square root transformed 12 site dataset. Bubble sizes represent PWCu levels. A 0.1% ASV pruning threshold was applied in these analyses. Site codes as in **Table 2.1**.

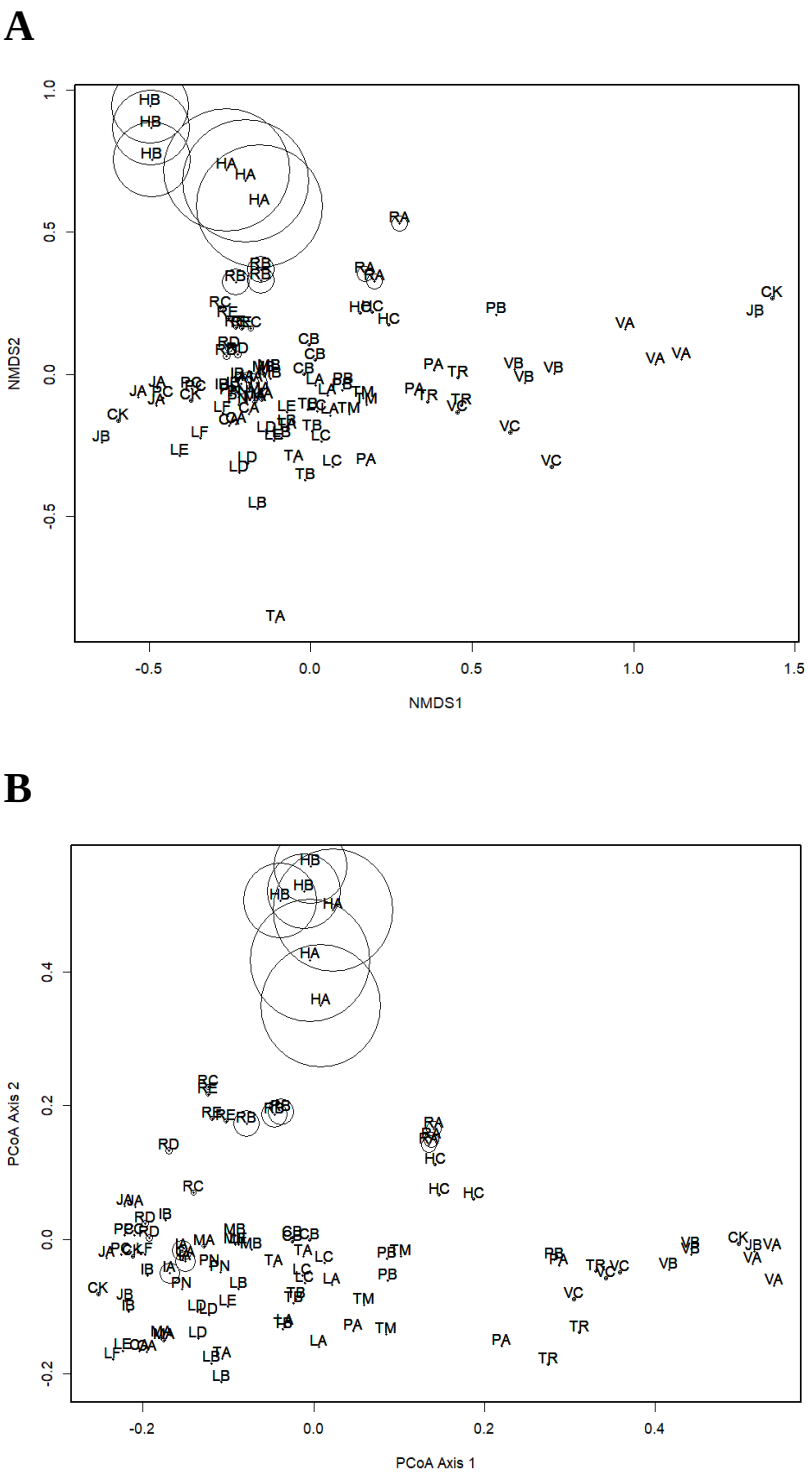


Fig. 4.20 Comparison of **A)** NMDS and **B)** PCoA on Bray-Curtis distances for square root transformed 34 site dataset. Bubble sizes represent PWCu levels. A 0.1% ASV pruning threshold was applied in these analyses. Site codes as in **Table 2.1**

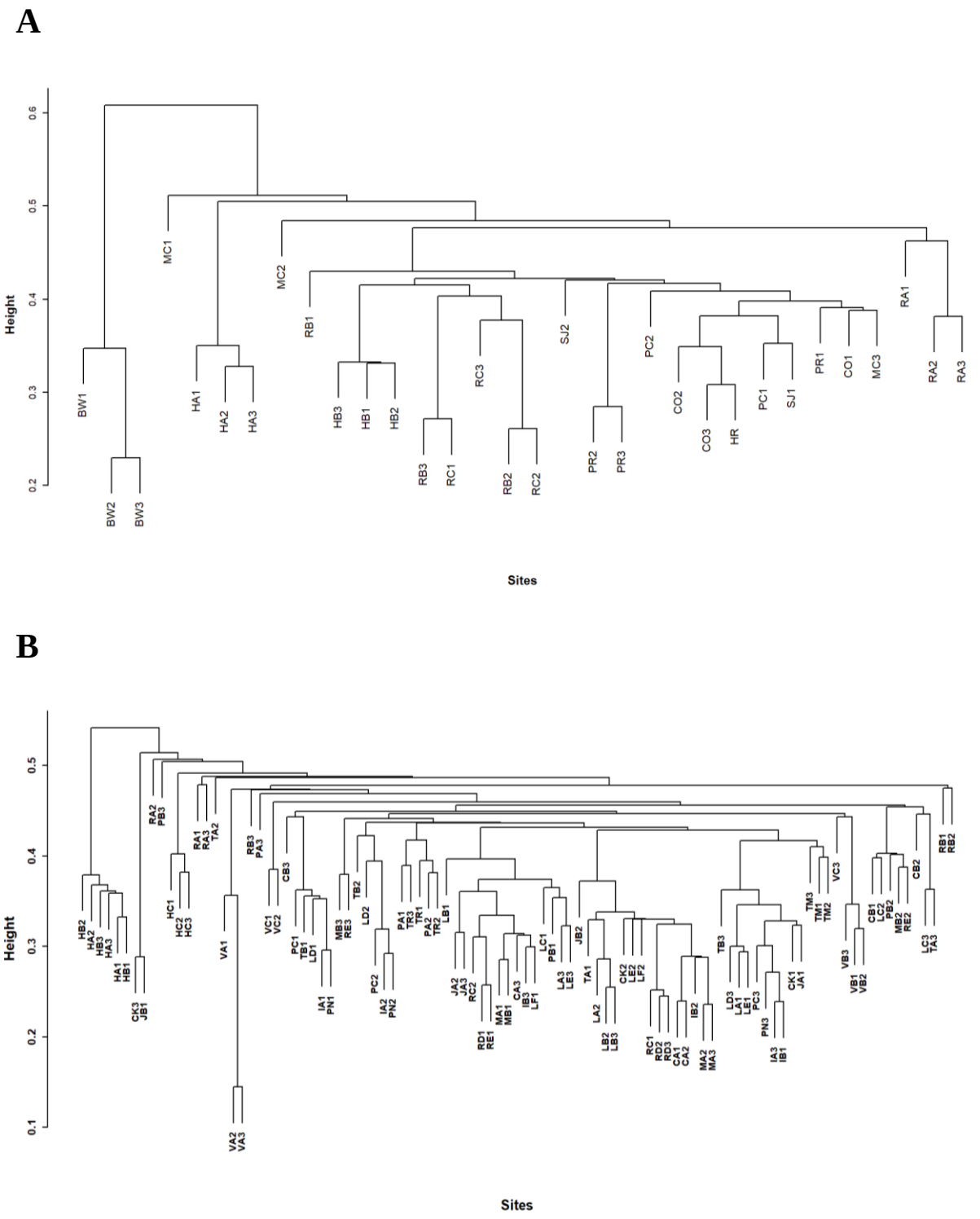


Fig. 4.21 Single linkage hierarchical clustering dendrogram of **A)** 12-Site and **B)** 34-site datasets. Branch heights represent the degree of dissimilarity between site clusters. Bray-Curtis similarity matrices were used, with 0.1% ASV pruning. Site codes as in **Table 2.1**

4.3.8 The Relationships Between Ecological Patterns and Environmental Variables

The relationships between environmental variables and microbial community composition were assessed using the BIOENV method (Clarke and Ainsworth, 1993), based on Spearman's rank correlation. For the bacterial communities at the 12 sites, the highest correlation with community composition was achieved by a combination of variables; AVS, PWCu, DOC, PW_OC_Cu and Salinity (**Fig. 4.22**). Among individual variables, PWCu gave the highest correlation value.

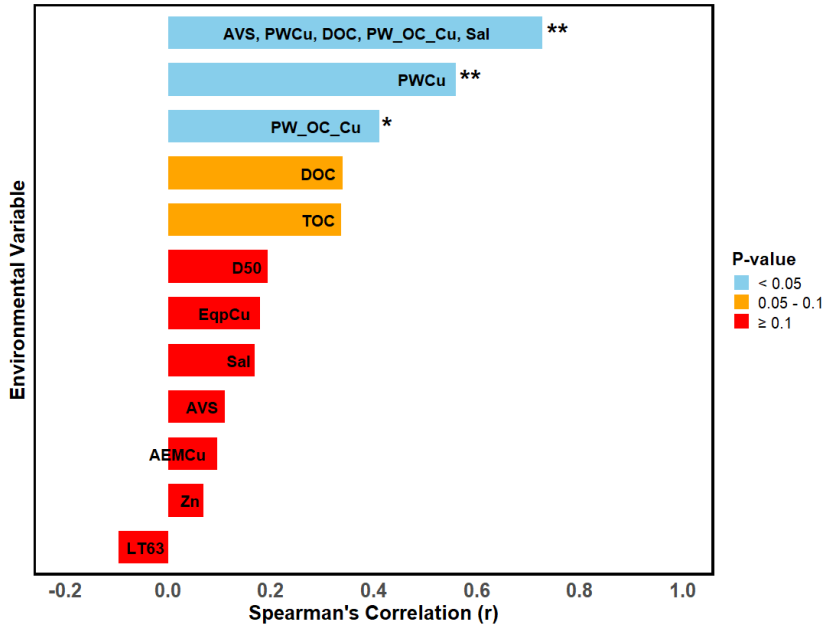
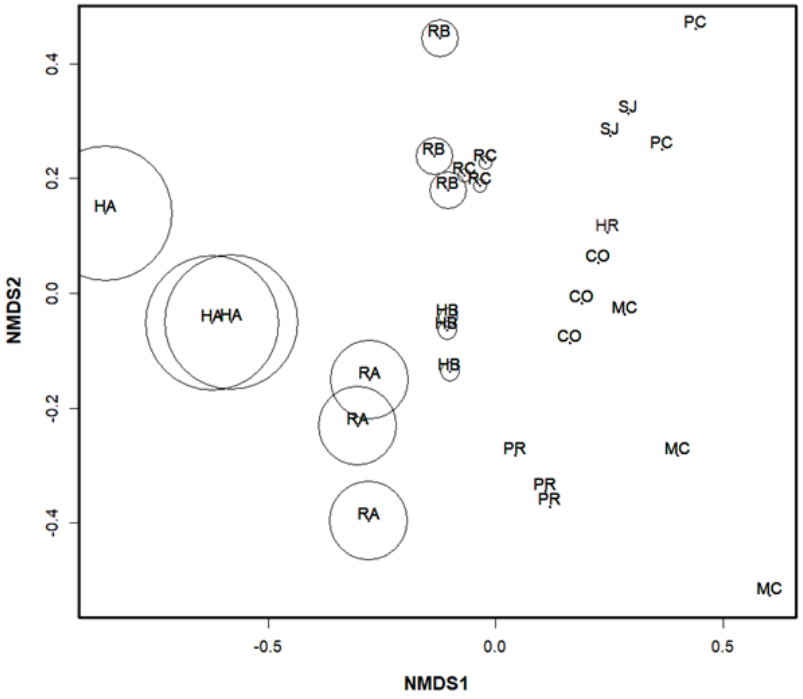


Fig. 4.22 Spearman's rank correlation (Rho) for BIOENV analysis of the relationships between environmental variables and microbial community composition for the 12-sites dataset. Bar colours indicate statistical significance: blue; most significant correlations ($p < 0.05$); orange; moderate significance ($0.05 < p < 0.1$); and red; non-significant ($p \geq 0.1$). Site codes as in **Table 2.1**. AEMCu (Acid-Extractable Copper), LT63 (<63 µm fines), AVS (Acid Volatile Sulphide), PWCu (Porewater Cu), TOC (Total Organic Carbon), DOC (Dissolved Organic Carbon), PW_OC_Cu (DOC normalised PWCu), EqPCu (Equilibrium Partitioning Cu), D50 (Median Grain Size), Sal (Salinity), and pH.

However, the multivariable combination provided additional context by incorporating factors such as organic-bound copper and sediment characteristics, which collectively enhanced explanatory power. The bubble plots for PWCu and PW_OC_Cu (**Fig. 4.23 A** and **B**) showed a clear alignment with the BIOENV findings. Sites with higher porewater copper and organically bound copper form distinct clusters.

A PWCu



B PW_OC_Cu

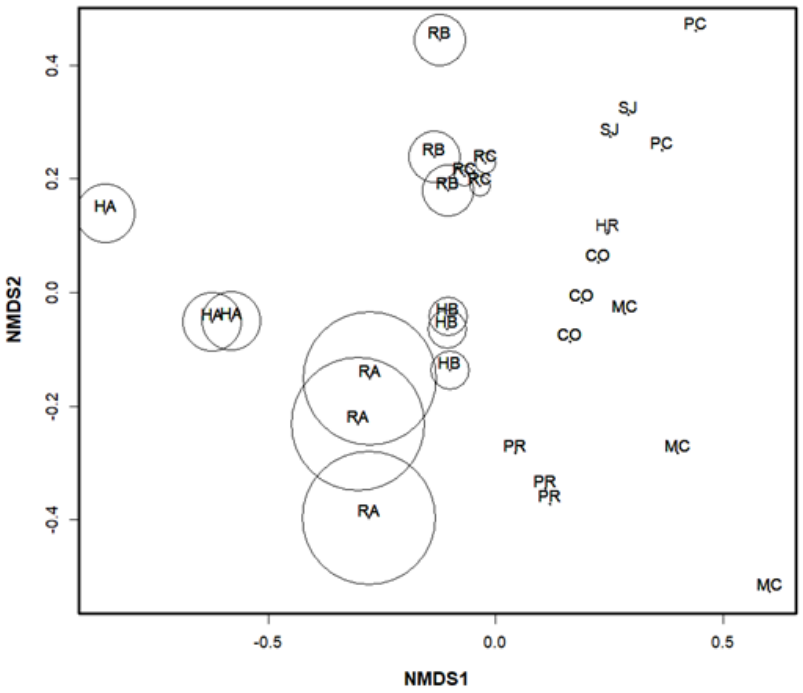


Fig. 4.23 NMDS plots of square root transformed data of bacterial abundances illustrating the spatial patterns of the 515F bacterial community structure derived from the 12-site dataset in relation to key environmental variables: **A)** PWCu, **B)** PW_OC_Cu. Bray-Curtis similarity matrices were used, with a 0.1% ASV pruning threshold.

In contrast, the 34-site dataset displayed a broader range of significant environmental parameters. Not only did porewater copper (PWCu), fine sediment fraction (LT63) and Salinity collectively yielded the highest correlation (**Fig. 4.24 A**), but each of these variables, along with median grain size (D50), also exhibited individually significant relationships with bacterial community composition. D50, in particular, showed the strongest single-variable correlation.

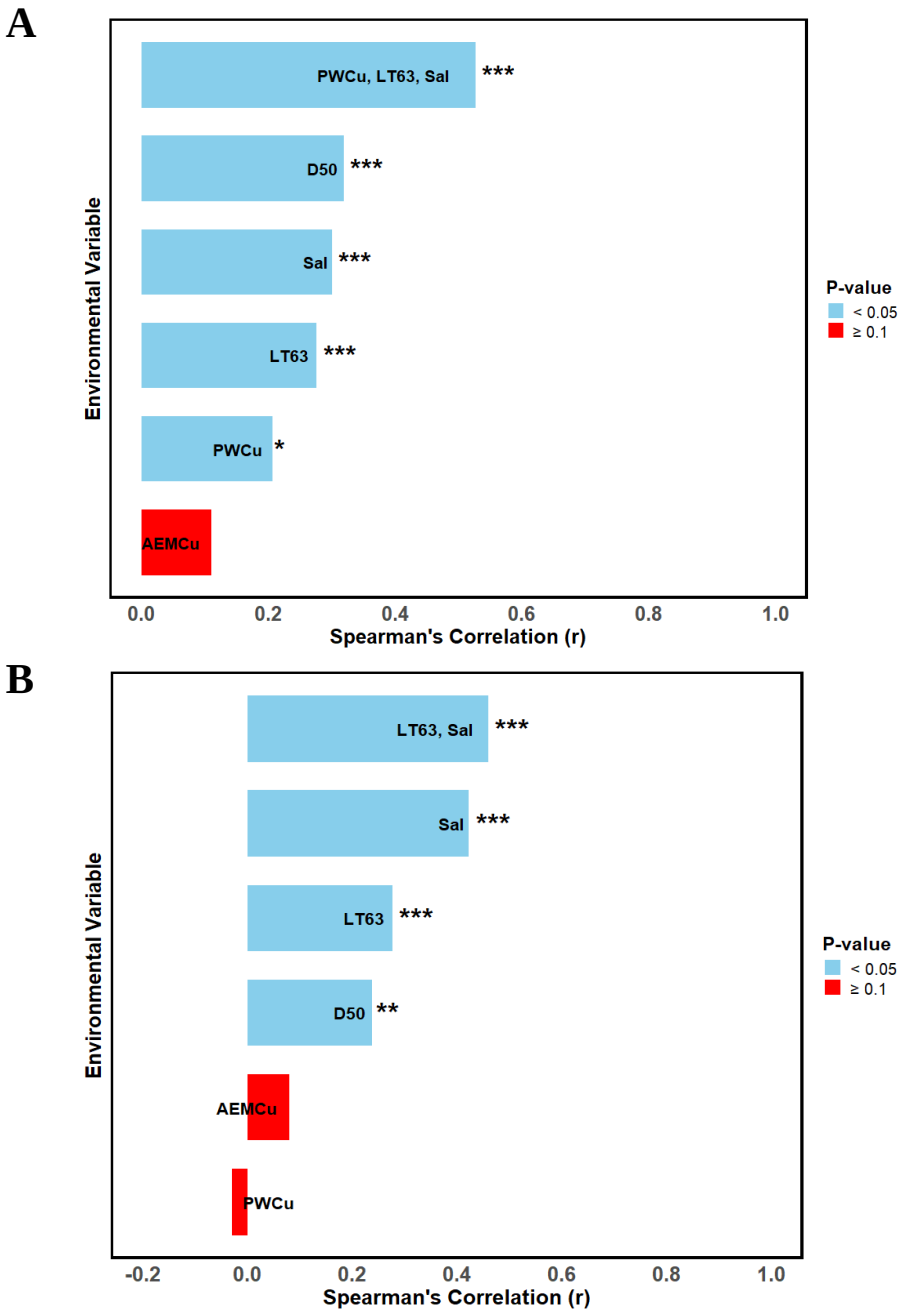


Fig. 4.24 Spearman's rank correlation (Rho) for BIOENV analysis of the relationships between environmental variables and microbial community composition. **A)** shows the results for the 34-sites dataset, and **B)** for the 34-sites dataset excluding highly polluted sites (HA, HB, RA and RB). Bar colours indicate statistical significance: blue; most significant correlations ($p < 0.05$); orange; moderate significance ($0.05 < p < 0.1$); and red; non-significant ($p \geq 0.1$). PWCu (Porewater Cu), AEMCu (Acid-Extractable Cu), D50 (Median Grain Size), LT63 (<63 µm Fines), Sal (Salinity).

The NMDS plots confirmed the overarching effect of PWCu, LT63 and Salinity on community patterns but also revealed that the most heavily polluted sites (HA, HB, RA and RB) dominated the overall ordination (**Figs. 4.25 A, B** and **4.26 A)**.

When HA, HB, RA and RB were excluded **(Figs. 4.26 B, 4.27 A, B** and **Fig. 4.28)**. BIOENV pointed to LT63 and Salinity as the most influential variables, with PWCu becoming non-significant (**Fig. 4.24 B**). Meanwhile, D50 remained important but displayed a more diffuse pattern. In support of these findings, Principal Components Analysis (PCA) further highlighted the influence of D50 on the 34-site dataset: the pronounced vector in the PCA biplot signified a strong correlation with one of the principal components and a distinct effect on how sites clustered. This pattern persisted even in the face of limited variability in PWCu across certain sites. Although D50 was statistically significant in BIOENV.

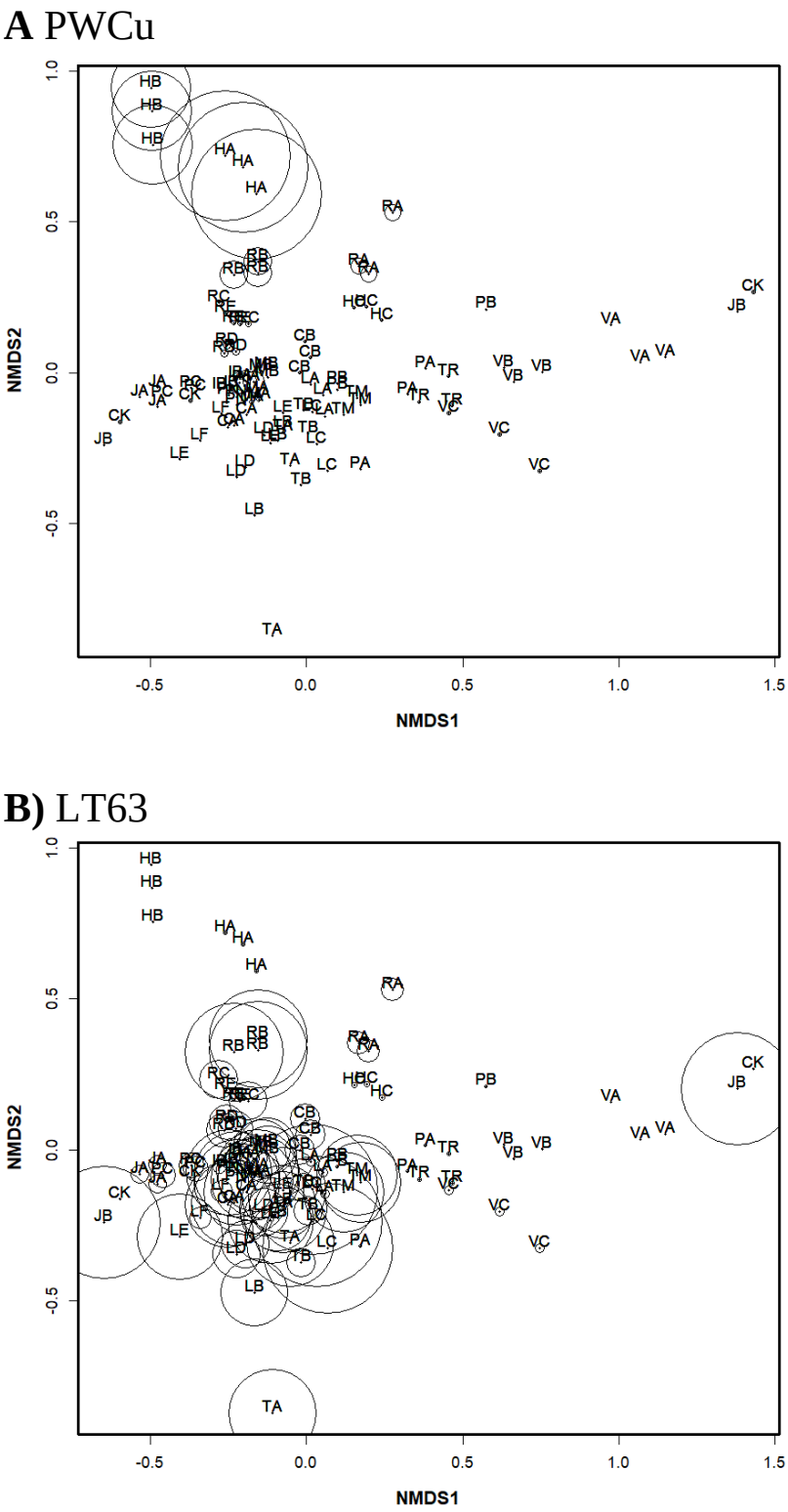


Fig. 4.25 NMDS plots of square root transformed data of bacterial abundances illustrating the spatial patterns of the 515F bacterial community structure derived from the 34-site dataset in relation to significant environmental variables. **A)** shows the relationship with PWCu, and **B)** with LT63. Both plots include all sites. Bray-Curtis similarity matrices were used, with a 0.1% ASV pruning threshold. Site codes as in **Table 2.1**.

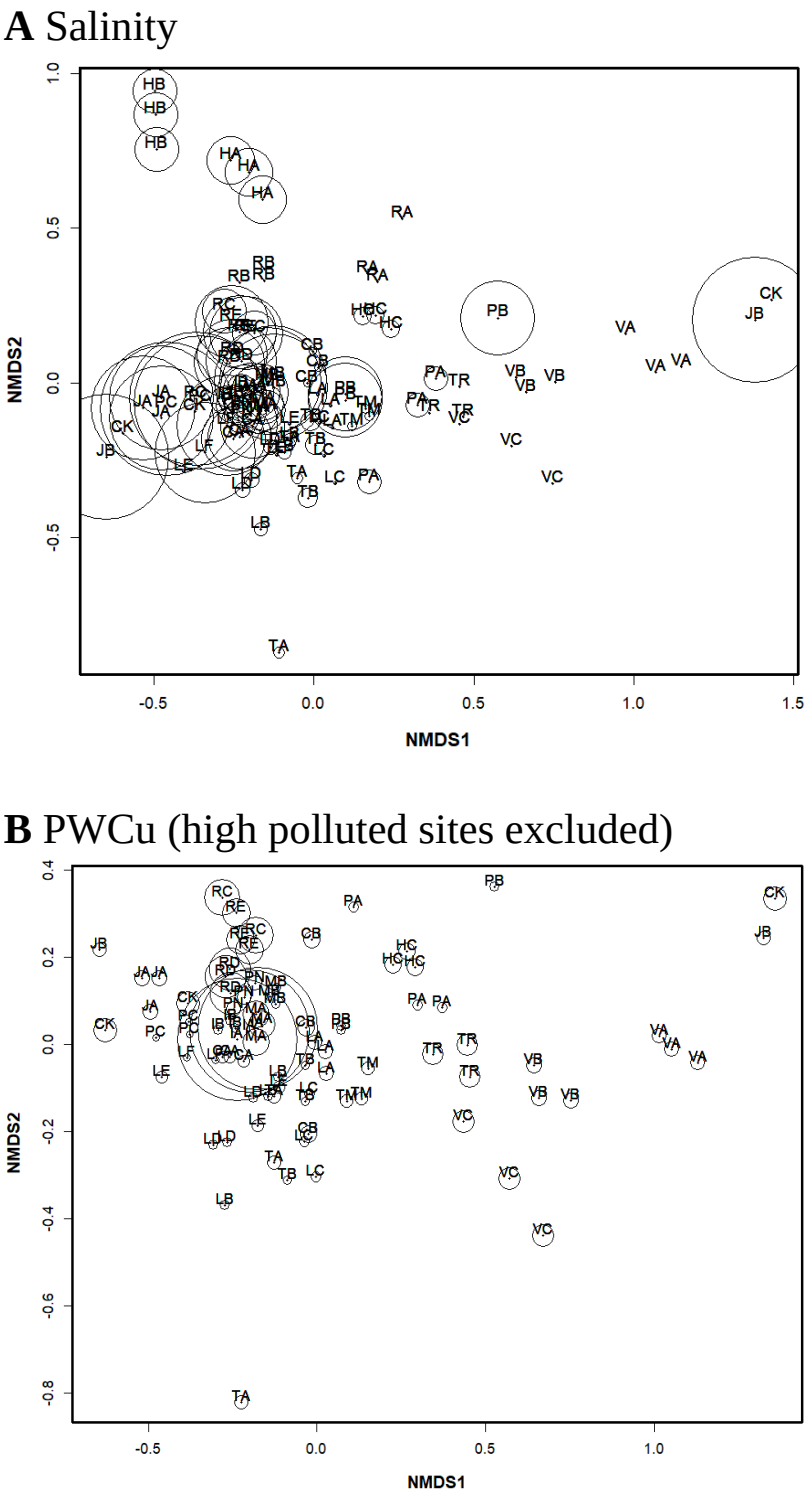
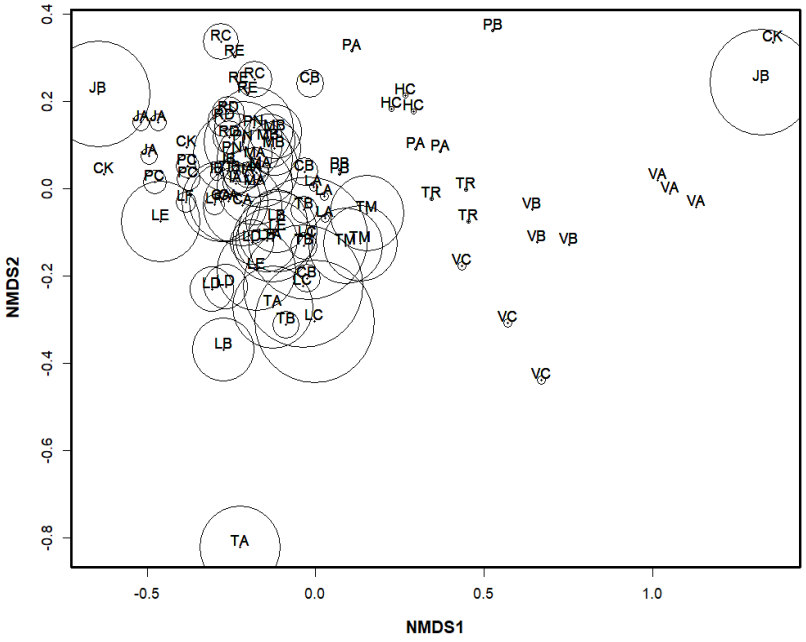


Fig. 4.26 NMDS plots of square root transformed data of bacterial abundances illustrating the spatial patterns of the 515F bacterial community structure. **A)** shows the relationship with salinity across all sites, while **B)** shows the relationship with PWCu after excluding highly polluted sites (HA, HB, RA, RB). Bray-Curtis similarity matrices were used, with a 0.1% ASV pruning threshold. Site codes as in **Table 2.1**.

A LT63 (high polluted sites excluded)



B Salinity (high polluted sites excluded)

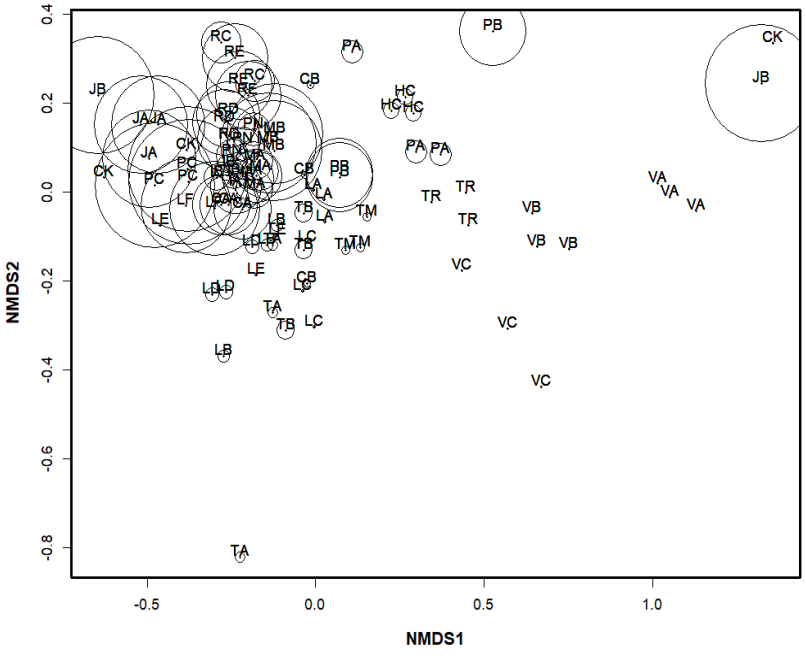


Fig. 4.27 NMDS plots of square root transformed bacterial abundances from the 34-site dataset after excluding highly polluted sites. **A)** shows the relationship with LT63, and **B)** shows the relationship with salinity. Bray-Curtis similarity matrices were used, with a 0.1% ASV pruning threshold. Site codes as in **Table 2.1**.

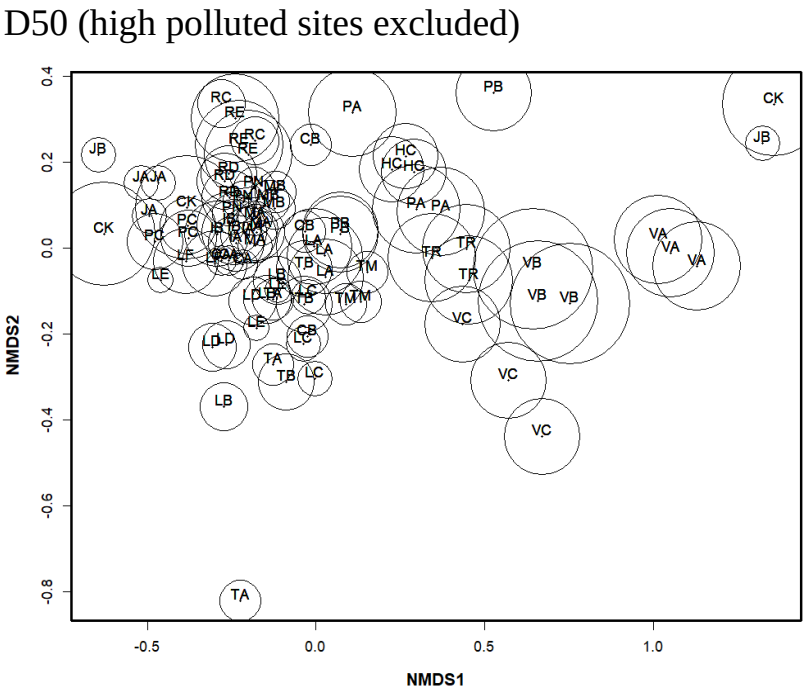


Fig. 4.28 NMDS plot of square root transformed bacterial abundances from the 34-site dataset showing the relationship with D50 after excluding highly polluted sites (HA, HB, RA, RB). Bray-Curtis similarity matrices were used, with a 0.1% ASV pruning threshold. Site codes as in **Table 2.1**.

4.3.9 Identifying Copper-Tolerant and Sensitive Phyla

To identify phyla that were tolerant or sensitive to copper pollution, ASV counts were aggregated to the phylum level. Average abundances in clean sites (ava) and polluted sites (avb) were calculated for each phylum and the fold change between polluted and clean sites was determined along with p-values. In the 12-site dataset, two bacterial phyla displayed significant differences in abundance between clean and polluted environments (**Fig. 4.29 A**). Cyanobacteriota showed a strong positive fold change and was significantly more abundant in clean sites. In contrast, Pseudomonadota exhibited a significant negative fold change. A phylum with unresolved taxonomy ('?') also showed a significant increase in clean sites. Other major phyla, such as Desulfobacterota and Spirochaetota, displayed abundance differences but did not reach statistical significance.

In the 34-site dataset, fewer phyla demonstrated statistically significant relationships with copper levels (**Fig. 4.29 B**). Pseudomonadota, Planctomycetota, Gemmatimonadota and

Acidobacteriota were all significantly less abundant in clean sites. Remarkably, Pseudomonadota was the only phylum consistently identified as significantly less abundant in clean sites across both the 12-site and 34-site analyses. Unlike the 12-site dataset, no phyla in the 34-site analysis showed a significant increase in unpolluted environments, with most significant findings indicating higher abundance under polluted conditions.

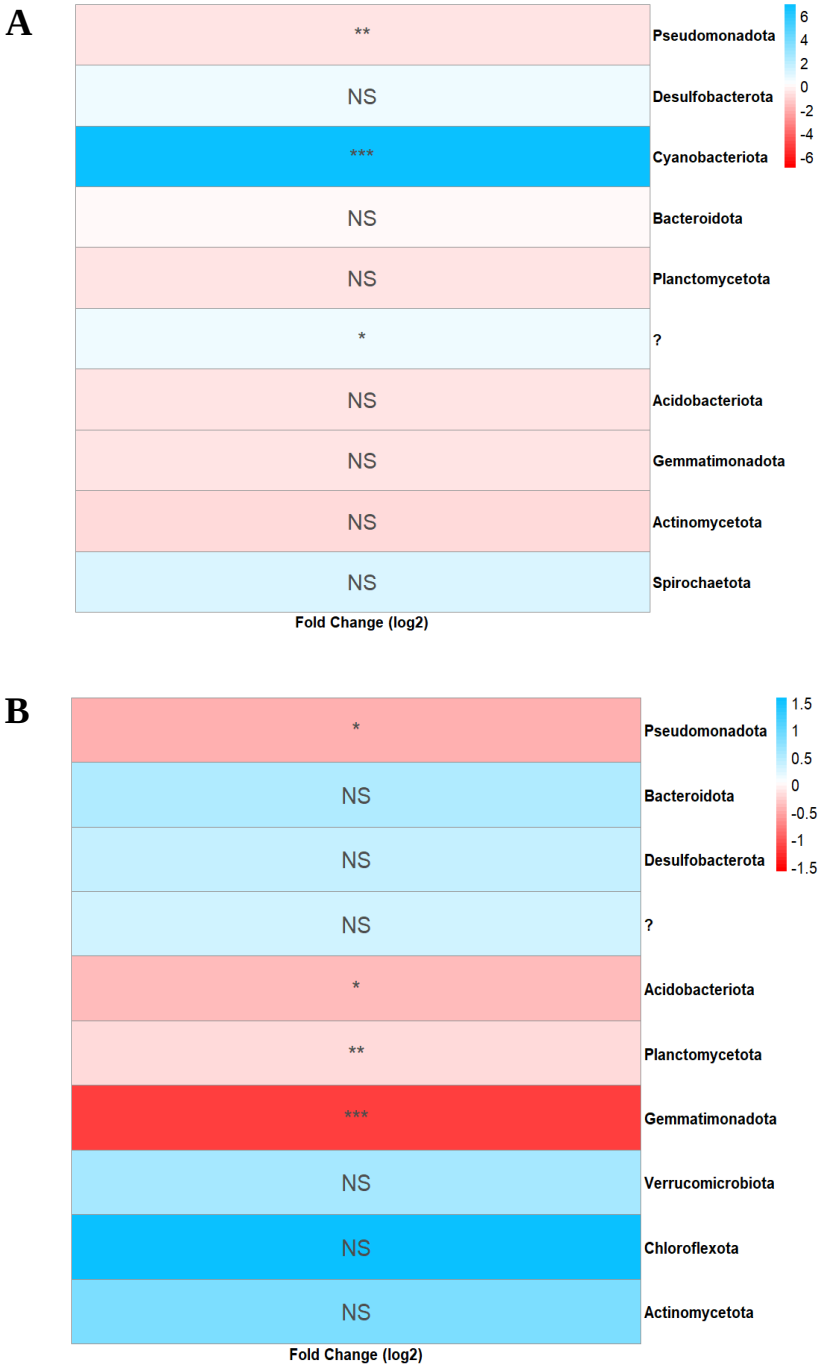


Fig. 4.29 Heatmap illustrating the top 10 phyla exhibiting significant differences in abundance between polluted and clean sites across **A)** 12-site and **B)** 34-site datasets, along with their p-values and raw fold changes. The phyla were sorted based on their average SIMPER contribution to dissimilarity between groups. Statistical significance is indicated by asterisks (*$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$), with "NS" denoting non-significant results. Darker colour intensities represent higher raw fold changes, blue shades indicate negative fold changes (more abundant in clean sites) and red shades indicate positive fold changes (more abundant in polluted sites). Raw fold changes [$\log_2(ava^2/avb^2)$] revert the square-root-transformed averages to the original abundance scale. Taxonomic assignments were made using the KSPG database. '?' means unresolved taxonomy by LCA in KSGP.

4.3.10 Identifying Copper-Tolerant and Sensitive Taxa at the ASV Level

In the 12-site dataset (**Fig. 4.30 A**), all highlighted ASVs displayed statistically significant differences in abundance between clean and polluted environments. Cyanobacteriota-affiliated ASVs (e.g., ASV13 and ASV310) were reasonably more abundant in clean sediments. In contrast, multiple Pseudomonadota-affiliated ASVs (e.g., ASV294 and ASV53) were consistently more abundant in polluted sites. Although several Desulfobacterota and Bacteroidota ASVs also differed significantly, their patterns were less uniform. In the broader 34-site dataset (**Fig. 4.30 B**), all identified ASVs were likewise significant but showed a tendency to be more abundant in polluted sediments. Here, Pseudomonadota and Desulfobacterota ASVs consistently exhibited negative fold changes in clean environments.

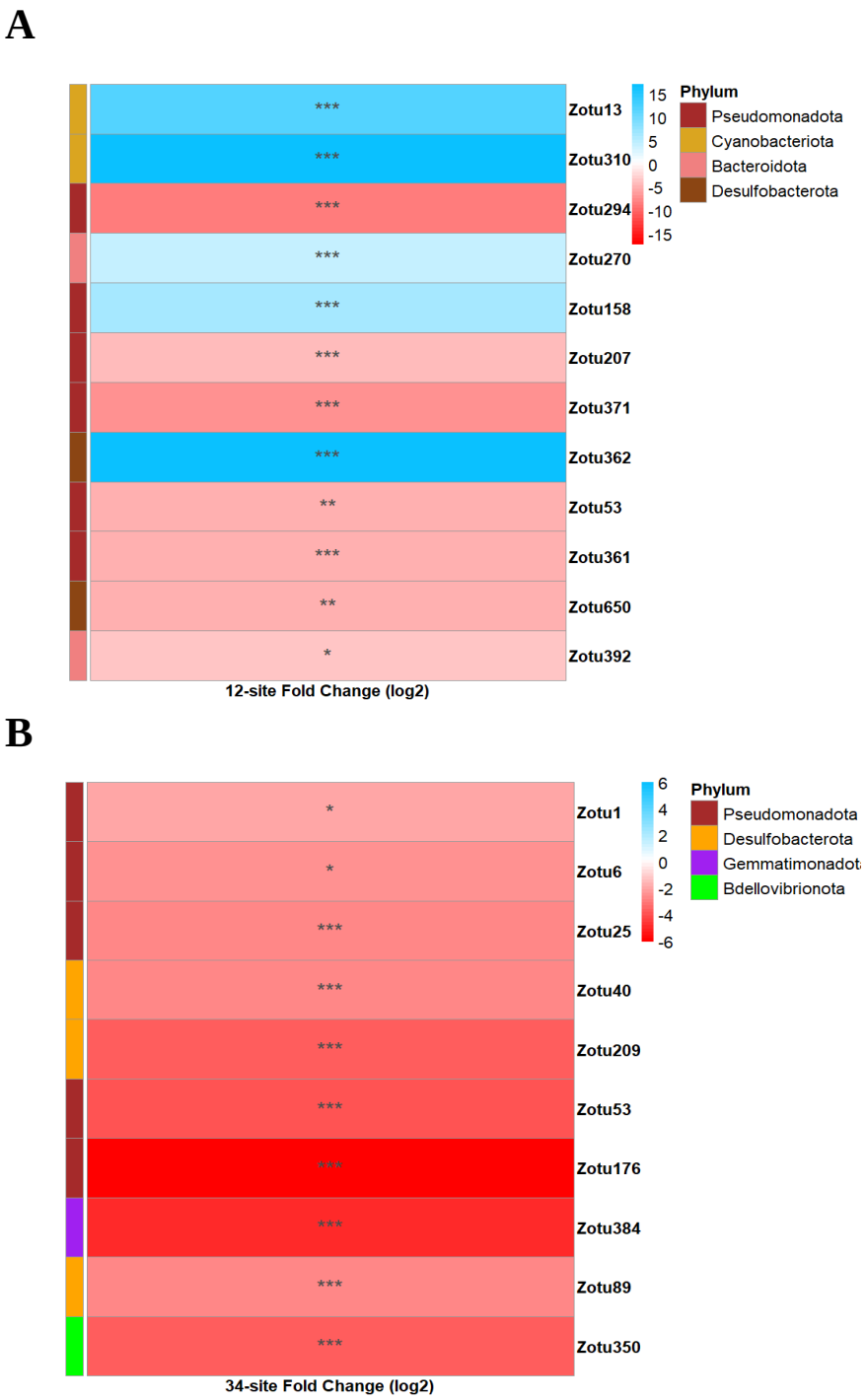


Fig. 4.30 Heatmap showing the top 10 significant ASVs for **A)** the 12-site and **B)** the 34-site datasets, each selected based on their average SIMPER contribution to differences between polluted and clean sites. Asterisks indicate statistical significance ($p \leq 0.05$; *$p \leq 0.01$; **$p \leq 0.001$). Colour intensity corresponds to raw fold changes, with blue indicating ASVs more abundant in clean sites and red indicating ASVs more abundant in polluted sites. Raw fold changes [$\log_2(ava^2/avb^2)$] revert the square-root-transformed averages to the original abundance scale. For ASVs with zero counts in clean sites (ASV310 and ASV362 in the 12-site dataset), an upper raw fold change limit of 15 was assigned. Taxonomic assignments were determined using the KSPG database.

4.4 Discussion

The study encompassed 34 estuarine sites in Southwest England, each sampled in triplicate, surpassing many earlier T-RFLP-based or single-site investigations in both scope and replication and thereby providing robust statistical power. Earlier surveys that used low-resolution techniques such as PLFA profiling (Agnihotri et al., 2023) or ARISA (Abed et al., 2015), reported broad variability but could not identify which taxa or functions tracked metal loads, illustrating the need for the high-throughput sequencing approach adopted here. Variation among replicates was consistently lower than variation among sites, showed that differences in environmental variables, including copper (Cu), primarily drive community changes. A significant but relatively weak negative correlation emerged between porewater Cu and alpha-diversity. Community composition at the most heavily contaminated sites diverged sharply from that at cleaner locations, with HA, HB, RA, RB, RC and RD showing particularly pronounced responses to elevated Cu levels. However, there were no clear relationships between community composition and Cu concentration at sites where contamination was more moderate so where effects may be limited or obscured by salinity, sediment texture and organic content. Pseudomonadota, especially Pseudomonas spp., are well documented to proliferate under heavy-metal stress because they combine broad substrate versatility with copper-efflux systems and pollutant-degrading pathways, allowing them to out-compete less tolerant taxa at highly contaminated sites (Zhang et al., 2007, Vojtková and Janulková, 2012, Panov et al., 2013).

These findings highlight that bacterial communities in estuarine sediments were influenced by copper contamination alongside salinity, sediment texture and other environmental gradients. Across up to 34 sites in Southwest England, porewater copper ranged from about 5 to 431 µg/L, while extractable copper reached up to 1797 µg/g, enabling a significantly wider contamination gradient than many earlier marine or estuarine surveys. Pseudomonadota and Bacteroidota together accounted for roughly half of the total sequence reads, mirroring other marine settings (Polinski et al., 2019). Yet at the ASV level these two phyla contributed a much smaller fraction of the total richness, indicating that their dominance in reads was driven by a relatively small number of very abundant ASVs. Smaller phyla such as Actinomycetota and Verrucomicrobiota, by contrast, showed more even read-ASV ratios, implying a broader spread of taxa and fewer single-lineage dominants. Considering both read abundance and ASV

diversity is therefore essential for capturing the full complexity of community composition. Compared with the 12-site survey, the 34-site accumulation curve approached its asymptote earlier and with a narrower confidence band, indicating that broader spatial coverage rather than deeper sequencing was pivotal for capturing the bulk of bacterial diversity. Pseudomonadota typically became more abundant at severely polluted sites, whereas Cyanobacteriota declined, in line with findings that chronic metal exposure selects for resistant taxa (Pennanen et al., 1996, Ranjard et al., 2013, Goswami et al., 2023, Sazykin et al., 2023, Yin et al., 2024). Some Desulfobacterota also appeared at heavily polluted locations, though not consistently across all replicates. Such taxon-specific patterns underscore how only a subset of bacterial groups shifts strongly in response to metal stress, while others remain comparatively resilient. A significant but relatively weak negative correlation emerged between porewater copper and alpha-diversity, indicating that high copper levels reduce richness but that moderate contamination can be masked by factors like salinity, grain size and organic content. Notably, only a small subset of abundant amplicon sequence variants (ASVs) showed a pronounced shift between clean and polluted sites, suggesting many dominant bacterial taxa remain resilient to intermediate copper concentrations. These patterns may reflect functional redundancy, where multiple taxa occupy similar ecological niches, or they may arise from other environmental drivers overshadowing copper effects in moderately contaminated sediments (Millward and Grant, 1995, Olsgard and Gray, 1995). Furthermore, comparisons of archaeal and bacterial responses (detailed in **Chapter 5**) indicate that archaeal diversity was often more sensitive to copper, consistent with findings that archaea sometimes exhibit less functional redundancy under metal stress (Gupta et al., 2021). Bacterial assemblages at the most contaminated sites (HA, HB, RA, RB, RC and RD) clustered separately, reinforcing that extreme pollution can outweigh other confounding variables. Filtering rare taxa at a moderate threshold (0.1%) helped reduce noise while preserving subtle shifts; excessively stringent filters risk losing important but low-abundance ASVs that may respond to copper.

In comparing these outcomes with other studies, the threshold around 20 µg/L porewater copper aligns well with Ogilvie and Grant (2008), who showed via Pollution-Induced Community Tolerance (PICT) that bacterial tolerance increases above approximately 15 µg/L yet found minimal structural changes using T-RFLP. Because T-RFLP missed many compositional details that PICT detected, it appears that functional tolerance measures can be more sensitive at lower Cu concentrations, whereas substantial community reorganisation

becomes evident only once Cu surpasses 15 µg/L. Similarly, observations in Restronguet Creek and Hayle parallel Ogilvie and Grant's detection of copper-driven effects in Pill Creek, suggesting that these southwestern England estuaries share a consistent threshold phenomenon, with tolerance shifts arising before large-scale compositional turnover. Here, including sites exceeding 100 µg/L PWCu and employing deeper sequencing enabled the detection of pronounced compositional shifts that extend Ogilvie and Grant's earlier findings, confirming a threshold-like response while revealing which taxa drive these community reconfigurations.

A number of authors have reported correlations between metals and microbial community composition at concentrations substantially lower than the levels at which clear relationships were observed. However, there were some important limitations to a number of these studies, including small sample sizes, lack of replication and failure to take into account potential correlations between metal concentrations and other environmental factors. For instance, Di Cesare et al. (2020) labelled their 14-site area "heavily polluted" because Hg, Pb and Zn reached up to 11.5 µg/L, 395 µg/L and 931 µg/L, respectively-well above Turekian and Wedepohl (1961) average shale values (0.18, 20 and 95 µg/L). Even the lower end of their gradient remained somewhat elevated in Hg and Pb, yet the study relied on a single sampling event and correlated metals with depth and nutrients, with no replicate community analyses per site. Consequently, their conclusion that depth and nutrient factors overshadowed metal impacts rests on correlative evidence, mirroring the present study's finding that moderate-to-high pollution effects can easily be concealed by other environmental variables when replication was limited. Similarly, Chen et al. (2022) sampled ten nearshore–offshore stations in the East China Sea where copper was only about twice local background (15-46 µg/g) and heavily confounded by water depth and salinity, an onshore-offshore gradient that further complicates attributing changes exclusively to metals. Zhao et al. (2022) showed metal levels in the Beibu Gulf were mostly below average shale references, rising from near to far shore likely due to finer sediment particle sizes, casting doubt on whether metals were truly driving community changes. Du et al. (2022) noted arsenic and cadmium correlated with microbial shifts in the Bohai Sea, yet their data suggest only moderate elevations (2-2.5 times average shale for As, under twice average shale for Cd and lacked replicate microbial sampling to confirm a metal-driven threshold. Meanwhile, Besaury et al. (2014) investigated two coastal sites in Chile exposed to long-term, untreated discharges of copper-rich mining waste, with one site showed extremely high total Cu (1410-1600 µg/g) and the other near background (65 µg/g).

Although their data indicated archaeal and bacterial resilience under heavy contamination, the single sampling event and one core per site left confounders like salinity or organic matter unaddressed, illustrating again how limited replication and narrow coverage can hinder definitive conclusions about metals' direct effects. Olsgard and Gray (1995) demonstrated how local sediment variability around North Sea drilling sites could produce "shotgun" ordination patterns that obscure distinct pollution effects, a phenomenon likewise seen in these estuaries where salinity and sediment properties complicate copper-specific signals. Taken together, these comparisons emphasise the distinct advantage conferred by this study's extensive copper range and multi-replicate design, which together provide the statistical power to confirm threshold effects and examine how partial shifts under moderate contamination can be concealed by environmental heterogeneity.

Overall, this work extends previous knowledge by showing that while bacterial communities often remain resilient at moderate copper concentrations, they undergo notable restructuring once porewater Cu surpasses about 20 µg/L, aligning with the tolerance thresholds reported by Ogilvie and Grant (2008). Although both tolerance and compositional changes can emerge around this same 15-20 µg/L range, the shift in overall community structure becomes especially pronounced once copper concentrations move well beyond that level, with severe contamination above 100 µg/L driving a more dramatic shift. This two-tiered response suggests that early functional tolerance can appear at lower levels, while pronounced taxonomic changes arise under significantly higher copper loads. Nonetheless, even at the highest measured Cu concentrations in these estuaries, overall bacterial alpha-diversity showed only a modest decline-markedly less severe than the diversity collapses documented for macrofaunal assemblages near offshore oil platforms (Olsgard and Gray, 1995). This discrepancy indicated that bacterial communities, despite responding compositionally, retain tolerant or functionally redundant taxa that buffer extreme pollution stress, contrasting with the sharp biodiversity losses often observed among marine macrofauna under heavy contamination. These findings contrast with some earlier reports that attributed community shifts to metals at comparatively low concentrations without robust replication or without accounting for confounders such as salinity and grain size. A broader contamination spectrum and higher replication demonstrate that copper-driven changes become unambiguous only when loads were substantially above background, helping to resolve the more ambiguous results from single-time or small-scale surveys. Confounding gradients can hide intermediate pollution effects, underscoring the need

for integrated approaches that combine high-throughput sequencing with detailed environmental data to capture both subtle and overt ecological shifts. By encompassing a broader contamination spectrum and employing rigorous replication, the study refines T-RFLP-based results from earlier work in this region and demonstrates how advanced metabarcoding can pinpoint threshold-level changes more reliably.

Chapter 5:

Impacts of Heavy Metal Pollution on Archaeal Diversity and Community Structure in Estuarine Sediments

5.1 Introduction

Archaea often comprise only a small proportion of total prokaryotic cells in moderate estuarine settings, typically in the 1-14% range (Webster et al., 2015). Although the salinity ranges reported for Tibetan Plateau lakes (0.6-324.8 g/l) (He et al., 2022) and solar saltern sediments (13-248 g/l) (Mani et al., 2020) are actually similar at the upper end, archaea account for only 2.41% of the microbial community in these Tibetan lakes, whereas they can reach 85% in the solar saltern sediments and even surpass bacterial abundance. This illustrates how environmental pressures such as elevated salinity can confer a significant advantage on archaea, aligning with other findings that these organisms, though often rare under benign conditions, can dominate when stresses intensify (Huang et al., 2016, Korzhenkov et al., 2019). Much of this variability appears linked to their functional breadth, which includes methanogenesis and anaerobic ammonia oxidation (Liu et al., 2010, Wang et al., 2020). Several studies suggest that archaeal lineages, including Thaumarchaeota and Bathyarchaeota, are prevalent in estuarine sediments (Zou et al., 2020), yet they may exhibit lower abundance compared to bacteria under moderate salinity or low-pollution conditions giving the impression that they are overshadowed by bacterial populations. There is also debate regarding archaea's relative sensitivity: some reports indicate that certain archaeal groups tolerate to heavy metals better than co-occurring bacteria (Salgaonkar et al., 2016, Korzhenkov et al., 2019) whereas others hint that archaea can exhibit equal or greater vulnerability depending on the specific contaminant and environmental conditions (Deng et al., 2018, Wang et al., 2023a). These contrasting observations highlight gaps in the current understanding of how archaeal abundance, distribution and taxonomic composition respond to estuarine pollution gradients, underscoring the need for robust, estuary-focused studies that identify which archaeal clades

proliferate, or recede under metal-enriched scenarios and whether their response thresholds differ markedly from those of other microbial groups (Wang et al., 2020, Zou et al., 2020).

Metal pollution is a widespread environmental issue threatening aquatic ecosystems globally (Hama Aziz et al., 2023, Singh et al., 2023). Heavy metals such as copper, zinc, cadmium and mercury accumulate in estuarine sediments, disrupting microbial community balance (Gillan et al., 2005). These metals are toxic to microorganisms, leading to reduced biodiversity and altered community structures (Jackson et al., 2015, Pan et al., 2022). Detecting the ecological impacts of metal pollution is challenging due to the difficulty in linking specific pollutants to community changes in environments that are variable in both time and space (Grant, 2010).

Heavy metal pollution, particularly with metals like copper, significantly decreases archaeal diversity and promotes the proliferation of metal-resistant groups such as Euryarchaeota (Gupta et al., 2021). In marine sediments, copper pollution enhances the activity of copper-adapted archaea, including unclassified Euryarchaeota and methanogenic archaea, demonstrating their resilience in metal-rich environments (Besaury et al., 2014). In contrast, in terrestrial ecosystems, heavy metal contamination reduces archaeal populations and alters community structures. In particular, Sandaa et al. (1999) reported that Archaea declined from about 1.3% of the total cells in uncontaminated soils to below detection limits in metal-contaminated soils, even though metal concentrations were below regulatory thresholds and community-level DNA profiling revealed distinct structures between treatments. The occurrence of organisms with elevated metal tolerance can indicate that ecological stress was favouring these over more sensitive species or strains (Grant et al., 1989, Blanck, 2002). However, pollution-tolerant populations were typically confined to heavily contaminated sites, limiting their utility as widespread monitoring tools (Grant, 2002).

Despite documented shifts in microbial communities within polluted estuarine sediments, the specific responses of archaeal communities to varying levels of metal pollution remain poorly understood (Zhang et al., 2024b). Heavy metal contamination markedly alters both archaeal and bacterial community structures, with some evidence showing temporal variations in benthic microbial communities in polluted areas (Yin et al., 2015, Coppo et al., 2023, Yu et al., 2024). While heavy metal contamination can have pronounced effects, many recent studies report only moderate metal concentrations in marine sediments e.g., (Chen et al., 2022, Zhao

et al., 2022), leaving gaps in understanding how archaea respond under more severe pollution scenarios. Nutrient inputs further complicate the picture by enhancing microbial activity and altering community structures, whereas long-term metal exposure selects for metal-resistant species (Voica et al., 2016, Zou et al., 2020). At the same time, some investigations suggest that bacteria might be as sensitive or even more affected by these stressors than archaea (Birrer et al., 2021), indicating that the relative sensitivity of archaea in heavily impacted estuaries remains an open question (Euler et al., 2020, Zou et al., 2020).

Additionally, archaea play important roles in bioremediation and wastewater treatment, providing insights into microbial ecology and serving as potential markers for assessing environmental quality and guiding remediation efforts (Krzmarzick et al., 2018). However, the lack of comprehensive archaeal reference sequences and outdated classifications in databases means much of their diversity remains uncharted (Grant et al., 2023). Identifying archaeal taxa tolerant to pollution will help establish reliable indicators, enhancing the ability to accurately assess environmental quality and implement effective remediation strategies.

Archaeal communities are increasingly recognised as key contributors to biogeochemical processes in estuarine sediments, yet the extent of their sensitivity to metal pollution remains partially unresolved. Existing microbial-focused studies e.g., (Chen et al., 2019) have highlighted the impact of copper (Cu) and other metals on overall community shifts, but often emphasise bacterial or resistome patterns rather than specific archaeal responses. By contrast, Korzhenkov et al. (2019) provide direct evidence of archaeal dominance under metal-rich, acidic conditions, underscoring their capacity to thrive where copper and zinc reach toxic levels. However, such investigations typically involve either narrow sampling scopes or single-extreme sites, as demonstrated by Coppo et al. (2023) who assessed Cu-influenced benthic shifts but primarily emphasized meiofaunal eukaryotes. Broader-scale reviews from Zou et al. (2020) and Voica et al. (2016) revealed that salinity and halophilic adaptations can mask or alter archaeal responses to metals, though they did not quantitatively specify which contamination thresholds specifically trigger archaeal community restructuring.

To address these gaps, the present work examines two distinct datasets (12 versus 34 estuarine sites) with porewater copper levels ranging from 2 to >400 µg/L and AEMCu (4 to ~1800 µg/g). This design surpasses previous research by covering both relatively pristine and severely

contaminated sediments, employing recently developed archaeal-specific primers that target a greater range of Archaea phylogenetic diversity than older primers and deeper sequencing approaches to capture potentially subtle or threshold-level community shifts. Such an expanded site coverage and refined molecular analysis enable a clearer identification of copper-tolerant versus copper-sensitive archaeal taxa, providing new insights into how multiple gradients may interact with metal stress in shaping archaeal assemblages.

This chapter therefore examines how archaeal diversity and community composition respond to metal pollution across a wide gradient of contamination in Southwest England estuaries, including sites with copper concentrations that exceed those investigated in most recent marine studies (17-34) µg/L in (Chen et al., 2019) and (16-37) µg/L in (Peng et al., 2024). Specifically, porewater copper levels differ by a factor of 90, while extractable metal concentrations vary by a factor of 400, offering a comprehensive assessment of pollution effects on archaeal communities. By analysing overall community patterns and the abundance of individual OTUs, this work evaluates the extent to which archaeal diversity reflects severe metal contamination and determines which contamination measures most strongly correlate with observed ecological differences. Building on these considerations, the central question was whether archaeal communities show a clear and measurable response to the highest copper concentrations and if so, at what contamination level that response emerges and how it compares with less polluted sites.

5.2 Methods

5.2.1 Sampling and DNA extraction

Three replicate samples were obtained from intertidal areas within estuaries in Southwest England, across 12 locations as detailed in (Udochi, 2020) and 34 locations as outlined in **Chapter 2**, along with two replicates from Breydon Water, Norfolk. The DNA extraction methods are described in **Chapter 2**.

5.2.2 PCR and sequencing

Hypervariable regions V1 and V2 of the 16S rRNA gene were amplified using primers SSU1ArF and SSU520R (Bahram et al., 2019). The forward primer and the reverse primer contained an Illumina adaptor, 8 base barcode, 0-7 bases of length heterogeneity spacer and primer sequence (Fadrosh et al., 2014). Primer Specifications were given in more detail in **Chapter 3**. Primers were supplied by Sigma Aldrich Company limited.

The polymerase chain reaction (PCR) was conducted in a volume of 20 µl, including 10 µl of PhusionTM Flash master mix (Thermo Scientific, UK), 1 µl of forward primer and 1 µl of reverse primer (final concentration of primers: 100 µM), 7 µl of ultrapure sterile water (MilliQ water) and 1 µl of DNA, with concentrations varying from nearly zero to 9.3 ng/µl, was detailed further in **Chapter 2** Samples with zero concentration and no amplified PCR products visible on the gel were excluded, as outlined in **Chapter 3**.

The amplification of the DNA templates was carried out in a Veriti™ HID 96-Well Thermal Cycler, 0.2mL system (Applied Biosystems, UK) in either 0.2mL x 96 well plates or 0.2 ml PCR tubes, depending upon the number of samples be amplified.

The amplification consisted of an initial denaturation at 98°C for 10 minutes, followed by 35 cycles of denaturation at 98°C for 30 seconds, annealing at 50 °C for 30 seconds and extension at 72°C for 30 seconds. This was followed by a final extension at 72°C for 5 minutes, before being held at 4°C. Primers were supplied by SigmaAldrich Company limited.

The PCR products were visualised using gel electrophoresis as detailed in **Chapter 2**. Primers SSU1ArF and SSU520R amplify a product that is 520 bp. The sequencing adapters, barcodes and length heterogeneity spaces increased this length to approximately 660 bp. The PCR products were purified using Aline Biosciences PCRClean DX kit (Aline Biosciences, Woburn, USA) following the manufacturer protocol except that the ratio of beads to PCR product which was 1:1 and then modified to 0.7:1 at latest run. This ratio was altered in the light of experience to improve removal of free primer adaptors and adaptor dimers, which were not fully removed when the ratio in the manufacturer's protocol was used (Quail et al., 2009). Quantification of the PCR products followed the methods described in **Chapter 2**.

Sequencing of the purified 16S rRNA was performed at the Earlham Institute, Norwich, UK, using a pre-made library pool on a single lane of the NovaSeq 6000 flow cell with 250 bp paired-end in February 2022 and February 2023 and on a NextSeq 1000 with 300 bp paired-end in April 2023. Sequencing of 16S rRNA amplicons yielded up to 4 870 265 reads per sample. To ensure data quality and reliability. Samples with low sequencing read counts (ranging from 56 to 3,400 reads) were excluded from the analysis, resulting in an average of 471,095 reads per remaining sample. When these were excluded the mean number of reads per sample was 691 759 on NovaSeq and 101 503 on NextSeq. This removal of low count samples maintained the integrity of the dataset and enhanced the accuracy of subsequent analyses. The excluded samples were BW1_1_a and BW2_2_s from the pilot samples; HB3 and CO2 from the 12-site set; JA2, MB2, PN3, RC2 and VC2 from the 34-site set; and LB_VC3, LB_PA2, MB_VC2 and RD_Bag from the experimental set.

Bioinformatic analysis was carried out using the LotuS2 pipeline (Özkurt et al., 2022), with taxonomic assignment of sequences using KSGP version 1.0 (Grant et al., 2023). Archaeal OTUs were clustered at a similarity threshold of 97%.

5.3 Results

5.3.1 Phylogenetic Analysis of Archaeal Communities

A preliminary phylogenetic overview of archaeal assemblages was conducted to establish baseline taxonomic diversity and identify potential novel lineages. This phylogenetic tree (**Fig. 5.1**), constructed using the 16S ARF rRNA data, revealed a highly diverse Archaeal community. Archaea made up approximately two-thirds of the tree, with a few bacterial OTUs less than (0.1%) and a slightly larger number of eukaryotes (2%). The great majority of unclassified OTUs lied on long branches in the upper section of the tree, with a small number lying on long branches within the Archaea section of the tree. Manual BLAST searches against the NCBI nt database showed that none of these unclassified OTUs were common. Manual BLAST searches against the full NCBI nt database showed that none of these unclassified OTUs were common.

The refined phylogenetic tree, excluding OTUs not classified as Archaea, was presented in **Fig. 5.2**. A total of 61% of ASVs were classified as Archaea, with all archaeal ASVs were resolved at the phylum level. Only 20 ASVs remained completely unclassified, lacking any domain-level assignment. The tree highlighted the remarkable diversity of Nanoarchaeota, with extensive branching within this group. However, other main phyla such as Thermoproteota, Aenigmatarchaeota and Thermoplasmatota exhibited higher proportions of unclassified OTUs (12%, 43% and 30%) respectively compared to Nanoarchaeota's 12%.

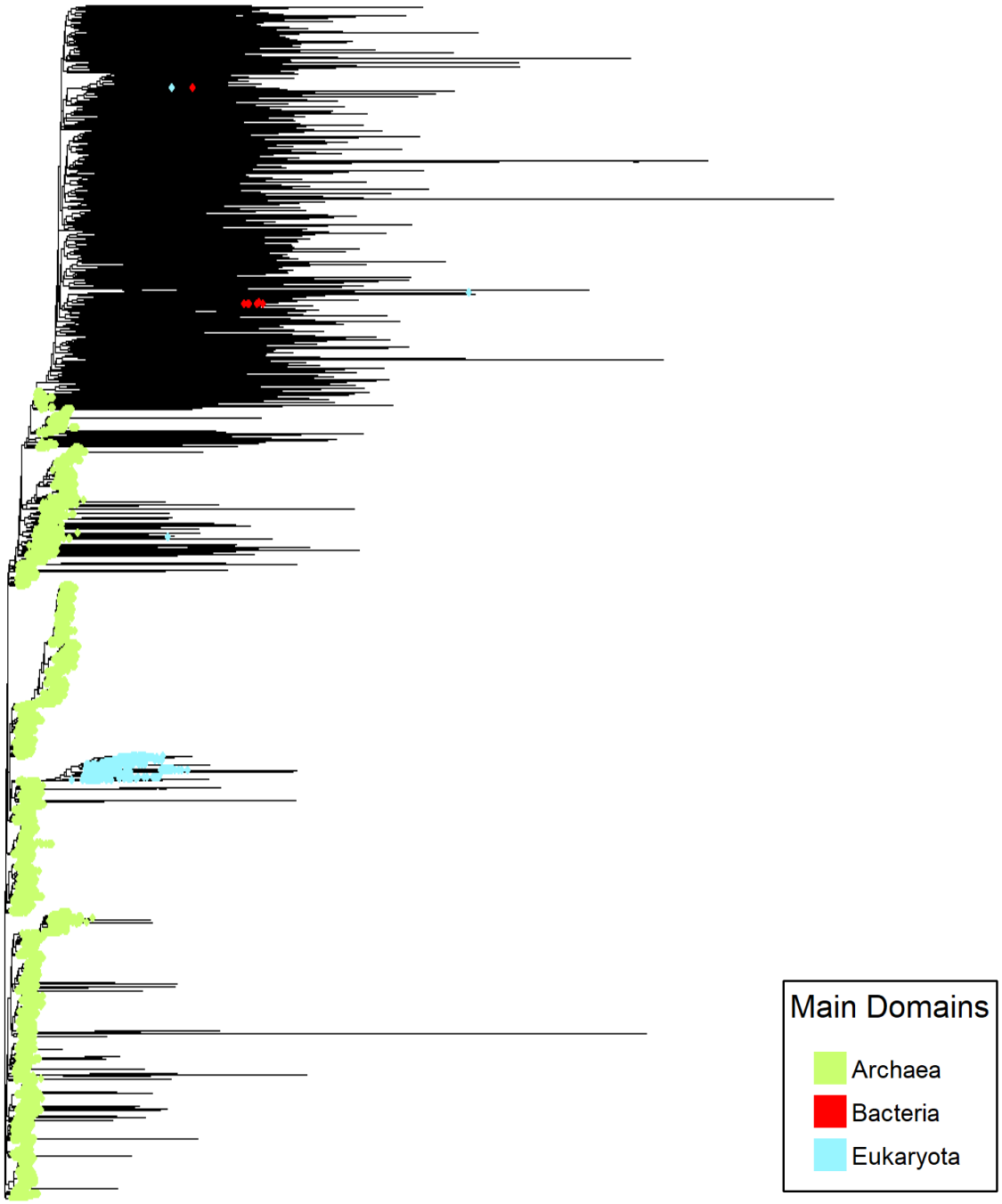


Fig. 5.1 Phylogenetic tree constructed using ARF 16S rRNA data across all datasets, showing all OTUs featuring Archaea (green), Bacteria (red) and Eukaryota (blue). Taxonomic assignments were performed using the KSPG database.

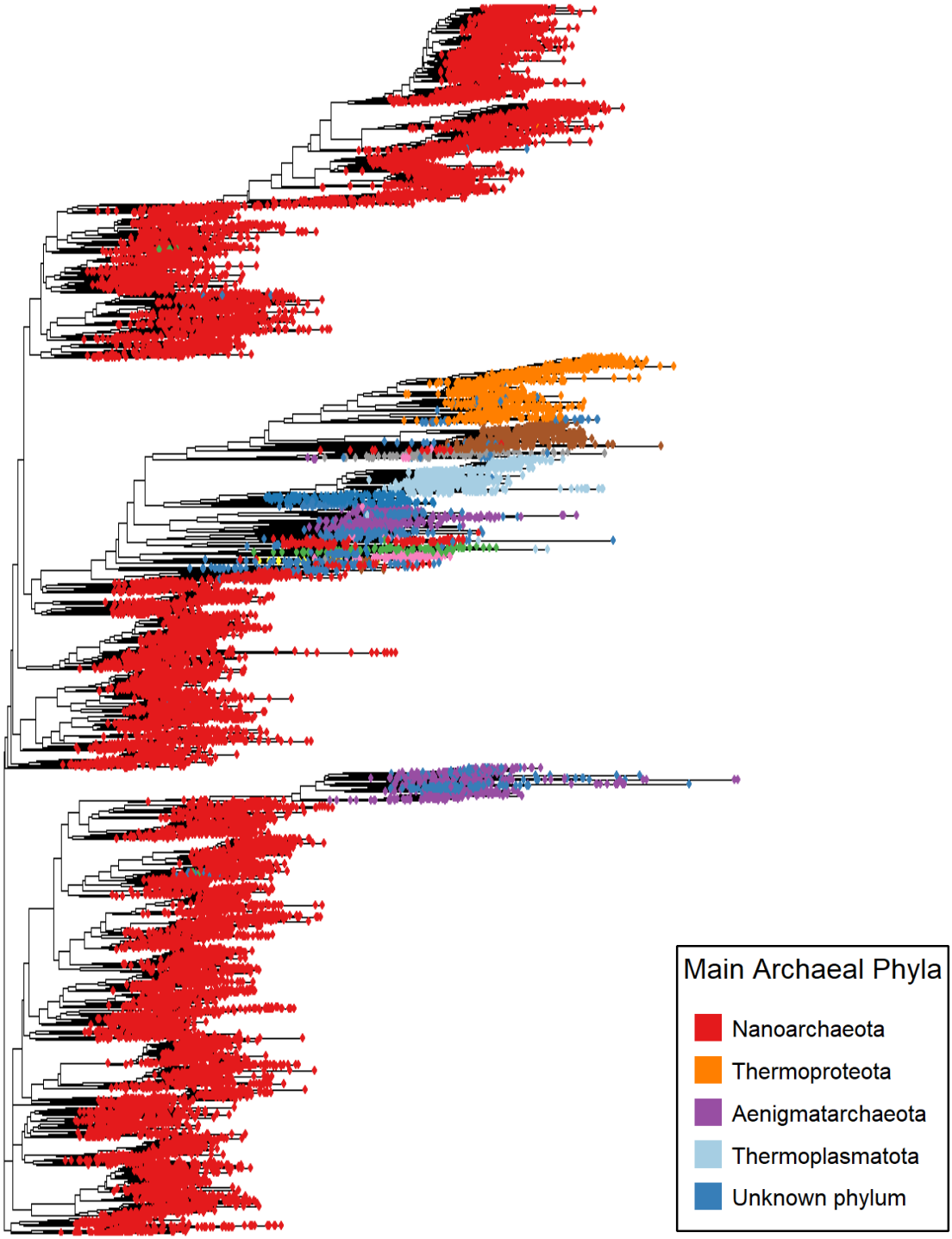


Fig. 5.2 Phylogenetic tree for Archaeal phyla based on ARF 16S rRNA data, following the removal of non-Archaeal branches. The main phyla represented are Nanoarchaeota (red), Thermoproteota (orange), Aenigmatarchaeota (purple), Thermoplasmatota (light blue), and an unknown phylum (blue). Taxonomic assignments were performed using the KSPG database.

5.3.2 Taxonomic composition

Two principal datasets (12-site and 34-site) were evaluated to capture archaeal community dynamics under narrower and broader pollution gradients, mirroring the strategy applied to bacterial analyses. The dominant archaeal phyla were Thermoproteota and Nanoarchaeota, which each made up around 40% of reads, with Thermoplasmatota contributing a further 9% (**Fig. 5.3 A**). Nanoarchaeota was the most diverse phylum, made up 80% of OTUs in contrast to 40% of reads (**Fig. 5.3 B**). By contrast, Thermoproteota and Thermoplasmatota made up only 5 and 3% of OTUs respectively, indicating that the average abundance of OTUs in these phyla was 16 and 6 times higher than it was for Nanoarchaeota. The taxonomy of nearly 5% of reads was not resolved at phylum level.

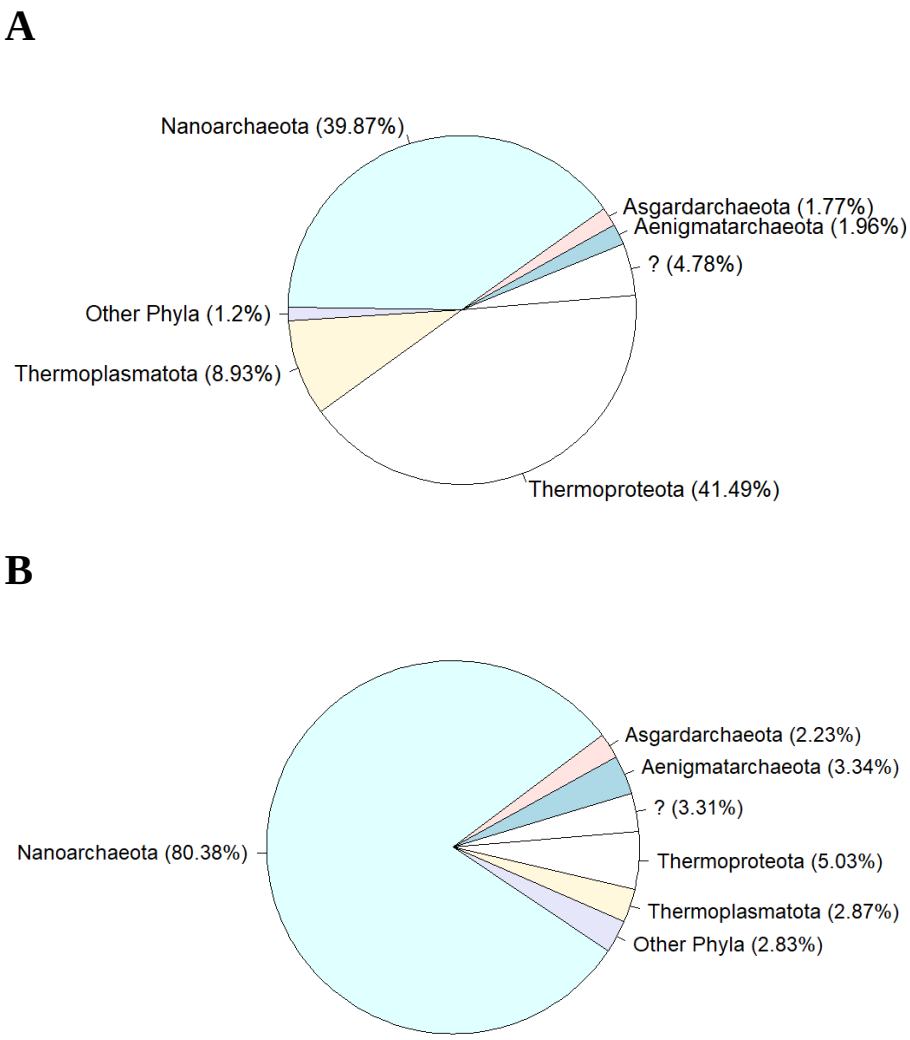


Fig. 5.3 Distribution of reads across phyla for all datasets after excluding non-archaeal reads, based on the total number of **A)** reads and **B)** OTUs, Taxonomic assignments were made using the KSPG database. Phyla making up less than 0.7% and 1.2% of the total respectively are combined into 'other phyla'. '?' indicates that taxonomy is not resolved at phylum level by LCA classification based on matches in KSGP.

5.3.3 NMDS analysis of replicate consistency and similarities between datasets

Analyses of replicate consistency and cluster patterns highlighted how archaeal phyla respond to contamination gradients in both the 12-site and 34-site datasets. By examining site-level differences, NMDS plots revealed spatial variability, pollution influences and dataset-specific distinctions. The **Fig. 5.4 A** presented the NMDS plot of all four datasets: pilot, 12-site, 34-site and experimental samples. The plot revealed partial overlap among these datasets. The 12-site

dataset was fully encompassed within the larger 34-site dataset, with several locations shared between them. Details and labels for both datasets were provided in **Chapter 2**.

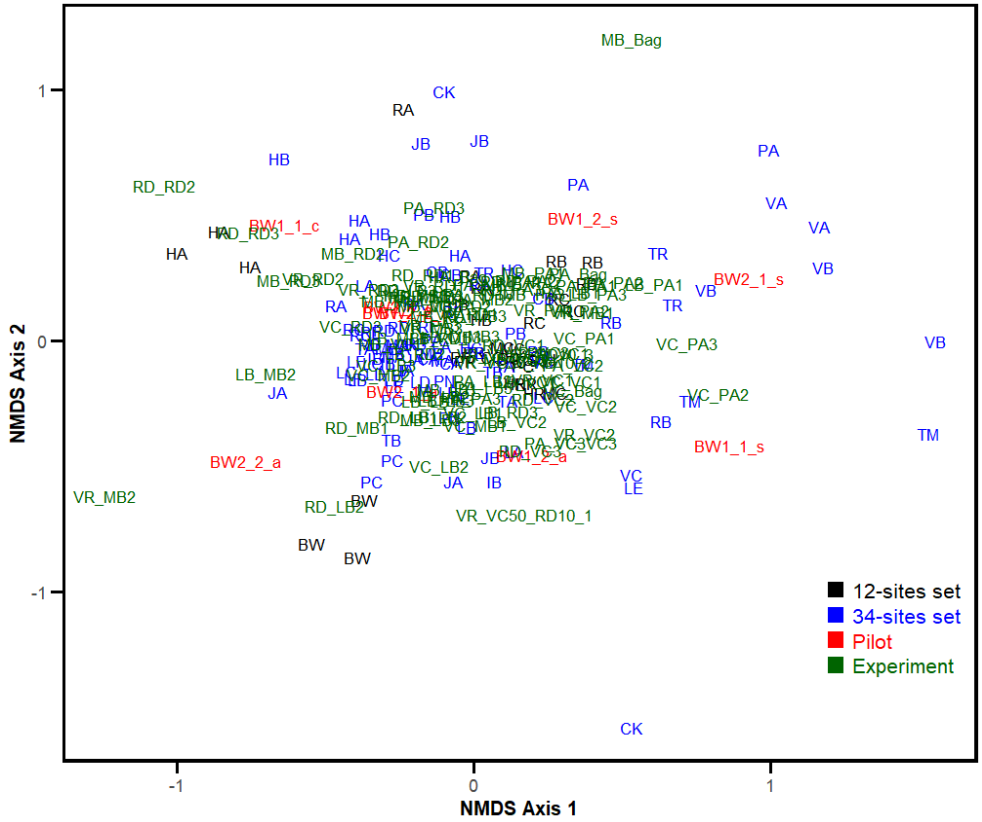


Fig. 5.4 Sample ordination using Non-metric Multi-dimensional Scaling (NMDS) of Bray-Curtis similarity matrix of Archaeal abundances, based on square root transformed data, derived from all datasets; Pilot, 12-sites, 34-sites and experiment. Colours represent the different datasets. Each two-letter sample label represents site. All datasets performed without the pruning of OTUs. Data processed using LotuS2 in conjunction with the KSPG database. Site codes as in **Table 2.1**.

When MDS is carried out including just these two datasets (**Fig. 5.5**), replicated clustered closely together within each dataset. However, in most cases samples from the same site collected at different times (2017 and 2022, respectively) did not plot in the same position on the NMDS plot. The exception to this was that samples in both datasets from sites in the Hayle were plotted towards the top left hand corner of the ordination. These were sites where pore water copper concentrations were at their highest.

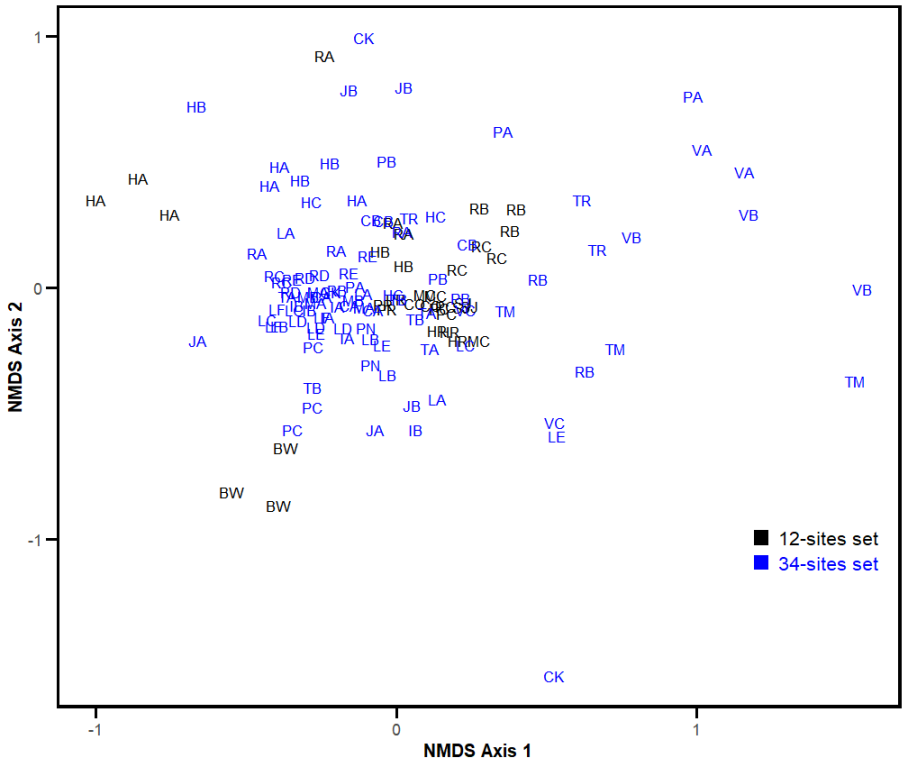


Fig. 5.5 Sample ordination using NMDS of Bray-Curtis similarity matrix of square root transformed Archaeal abundance data, showing the 12 and 34 site datasets. Colours represent the different datasets. Each two-letter sample label represents site. All datasets performed without the pruning of OTUs. Data processed using LotuS2 in conjunction with the KSPG database. Site codes as in **Table 2.1**.

In the 12-site dataset, samples from Breydon Water (BW) were clustered together, but were separated from all the Southwest (SW) samples (**Fig. 5.6 A**). The NMDS plot of the 12-site dataset (**Fig. 5.6 B**) showed that polluted sites were grouped on the left side of the plot. Additionally, the three replicated from each site clustered closely together, indicating that within-site variations were minimal compared to differences between sites. In the NMDS plot of the 34-site dataset (**Fig. 5.7**), replicated from the same site clustered closely together in most cases.

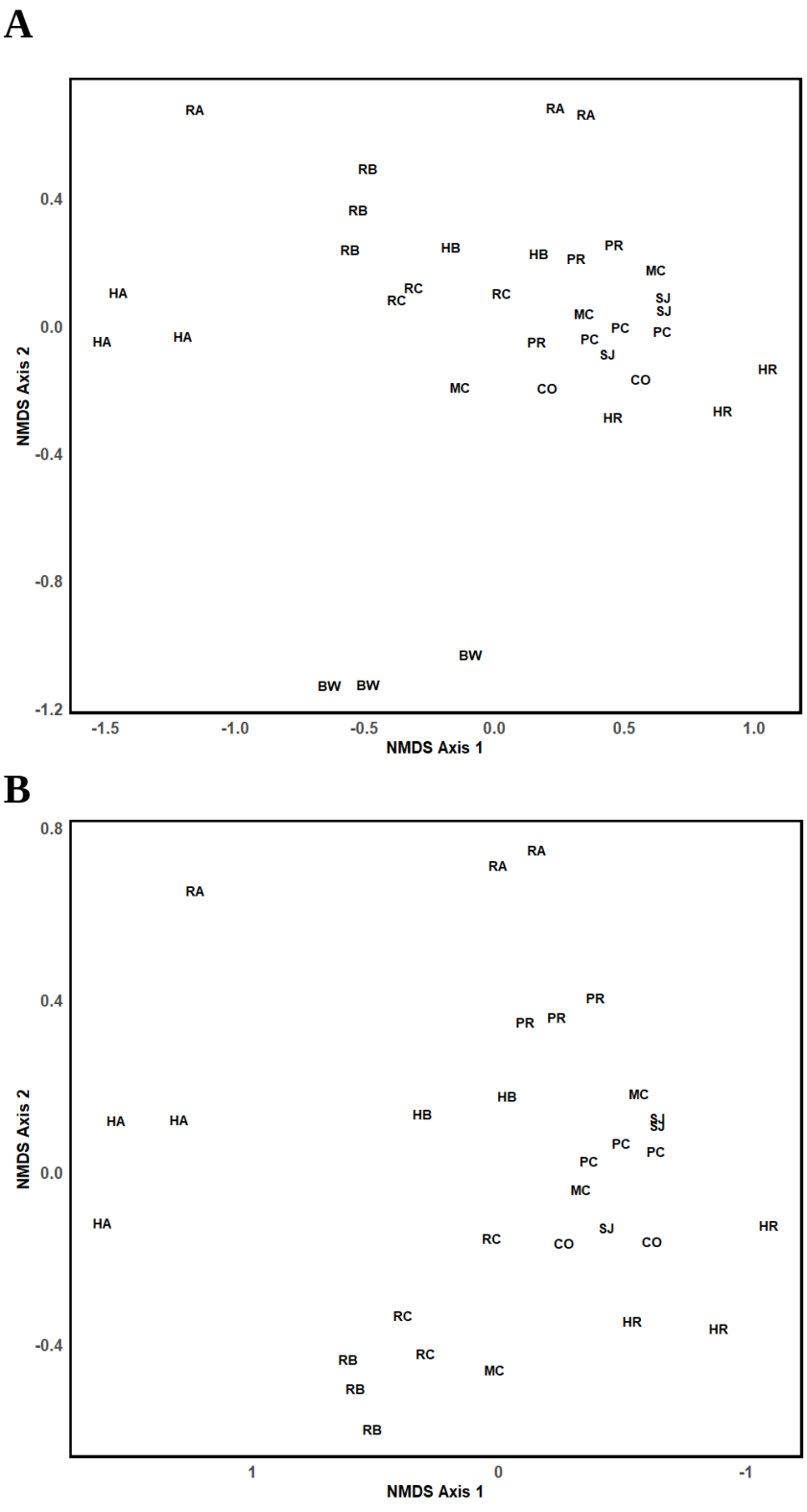


Fig. 5.6 Sample ordination using NMDS of Bray-Curtis similarity matrix of square root transformed Archaeal abundance data from the 12-site dataset. **A)** includes Breydon Water sites (stress value = 0.13), while **B)** excludes them (stress value = 0.12). All datasets performed without the pruning of OTUs. Data processed using LotuS2 in conjunction with the KSPG database. Site codes as in **Table 2.1**.

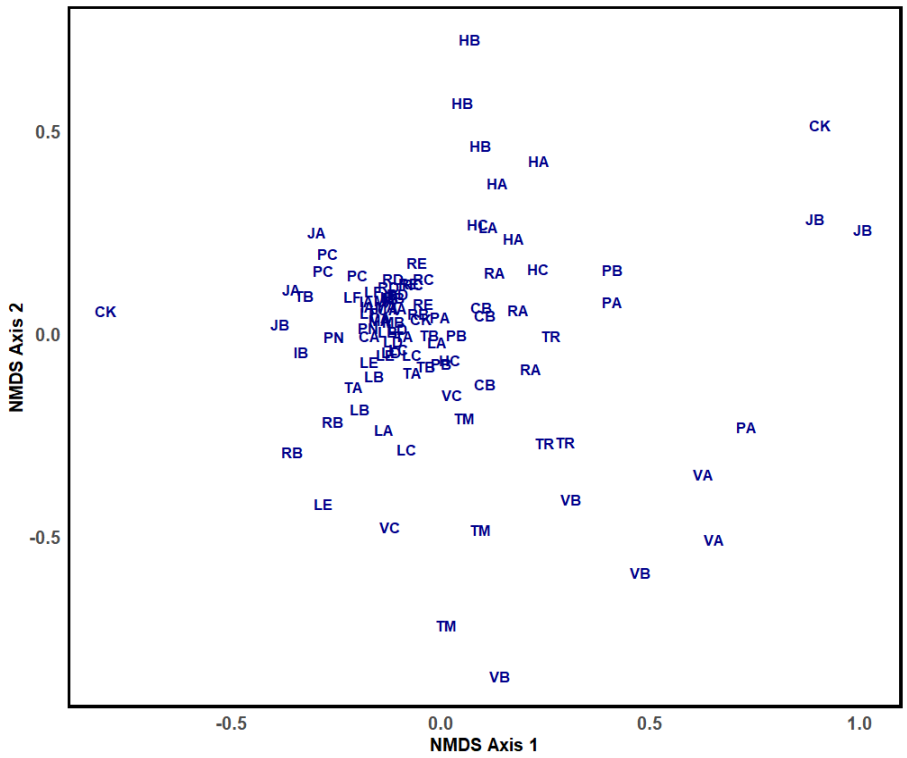


Fig. 5.7 Sample ordination using NMDS of Bray-Curtis similarity matrix of square root transformed Archaeal abundance data from the 34-site dataset (stress value = 0.16). Each two-letter sample label represents site. All datasets performed without the pruning of OTUs. Data processed using LotuS2 in conjunction with the KSPG database. Site codes as in **Table 2.1**.

For the experimental samples, the NMDS plot (**Fig. 5.8**) illustrated the differences in community composition between treatments. The first part of each treatment name indicated the source of the inoculum, with "bag" samples representing the initial community composition. The second part specifies the origin of the defaunated sediment in each replicate, distinguished by colour in **Fig. 5.8** Treatments primarily clustered by the site of the defaunated sediment rather than by the inoculum. Treatments with sediment from the Avon estuary (VC) lied towards the bottom right of the ordination, whereas those with Helford and Percuil sediments (LB, PA) appeared on either side of the Avon cluster. ANOSIM analysis (**Fig. 5.8**) confirmed this pattern, with Sediment Source yielding a higher R-value (0.542) than Inoculum Source (0.119), both statistically significant ($p < 0.001$).

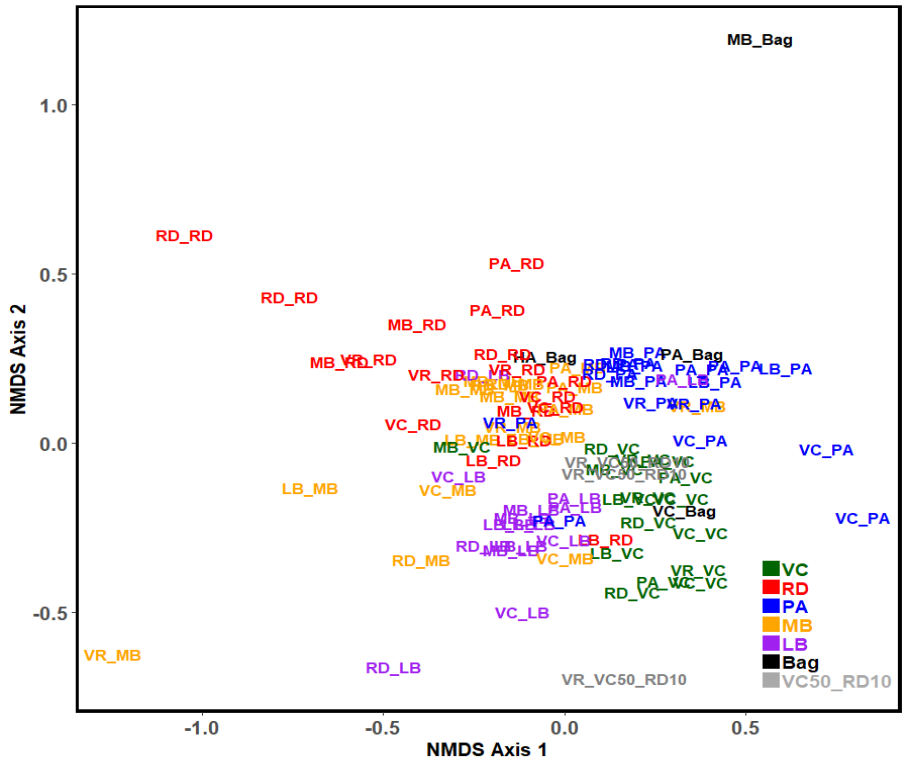


Fig. 5.8 Sample ordination using NMDS of Bray-Curtis similarity matrix of square root transformed Archaeal abundance data from the experiment set samples (stress value = 0.19). Colours represent the different test sediments. The first two letters in the sample labels indicate the source of the inoculum, while the second two letters represent the sediment that makes up 75% of the total composition. Each two-letter sample label represents site. All datasets performed without the pruning of OTUs. Data processed using LotuS2 in conjunction with the KSPG database. Site codes as in **Table 2.1**.

5.3.4 Archaeal Diversity Across Pollution Gradients

Rarefaction curves were used solely to standardise sequencing depth across samples, and the resulting rarefied ASV counts were then examined for patterns in archaeal community richness relative to porewater copper (PWCu) concentrations in both the 12-site (**Fig. 5.9 A** and **B**) and 34-site (**Fig. 5.10 A** and **B**) datasets. In the 12-site set, the standardised rarefaction curves (**Fig. 5.9 A**) indicated that sites with lower PWCu (i.e., cleaner environments) generally support higher archaeal diversity, whereas sites with elevated PWCu (notably in Hayle and Restronguet Creek) occupy the lower portion of the curves. A corresponding box-and-whisker plot (**Fig. 5.9 B**), based on rarefaction to 149 595 reads per sample, showed that mean OTU counts differ significantly among sites (F = 16). Low-copper sites (green) typically exhibited higher OTU

counts with minimal variability, while heavily contaminated sites (red) tended to have fewer OTUs and, in some cases (e.g. MC, HR and RA), showed greater variability among replicates. Additionally, the relationship between OTU richness and PWCu (excluding BW) was tested using Spearman's rank correlation, which showed a strong negative association (rho = -0.64, $p < 0.001$), confirming that archaeal diversity declined as copper levels increased.

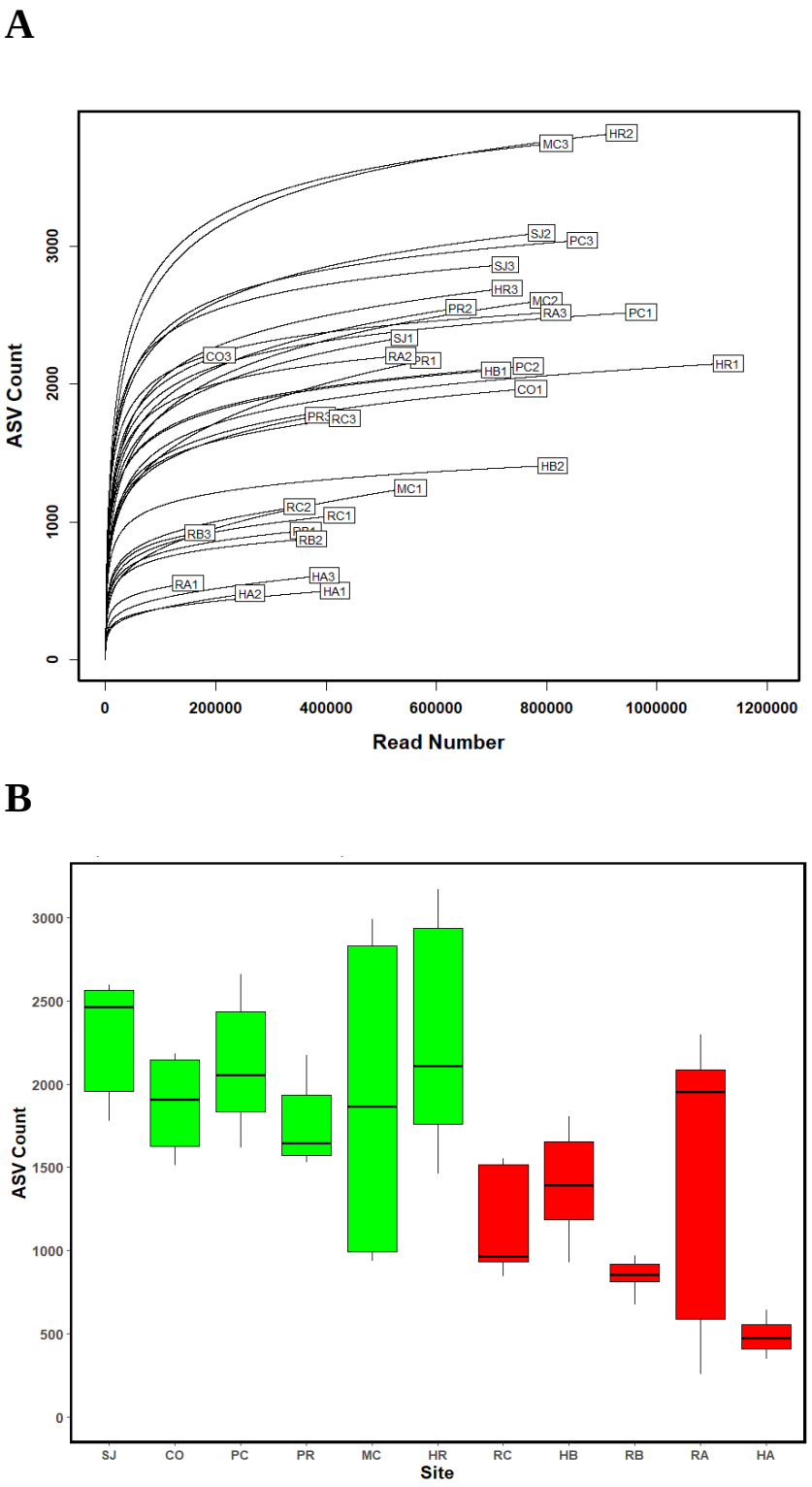


Fig. 5.9 A) Rarefaction curves for archaeal samples for 12 sites, illustrating OTUs richness in relation to sample size. **B)** Box and whisker plot of OTU counts rarefied to a sample size equivalent to the lowest sample size (149 595) across different sites, sites are arranged in increasing order of porewater copper concentrations with sites above and below 20 µg/L indicated in red and green respectively. ANOVA results ($F = 13.2$, $p < 0.001$). Spearman's rank correlation between OTU richness and porewater copper concentration ($rho = -0.64$, $p < 0.001$). Site codes as in **Table 2.1.**

In the 34-site dataset, rarefaction curves (**Fig. 5.10 A**) spanned a wide range of archaeal OTU richness without neatly separating clean and polluted sites. A box and whisker plot (**Fig. 5.10 B**) illustrated OTU counts for each site, based on rarefaction to 35,155 reads in the 34-site dataset. Although an ANOVA again confirms significant inter-site differences ($F = 8.4$), the Spearman's rank correlation between OTU richness and PWCu was remarkably weaker than in the 12-site dataset ($rho = -0.25$, $p < 0.001$).

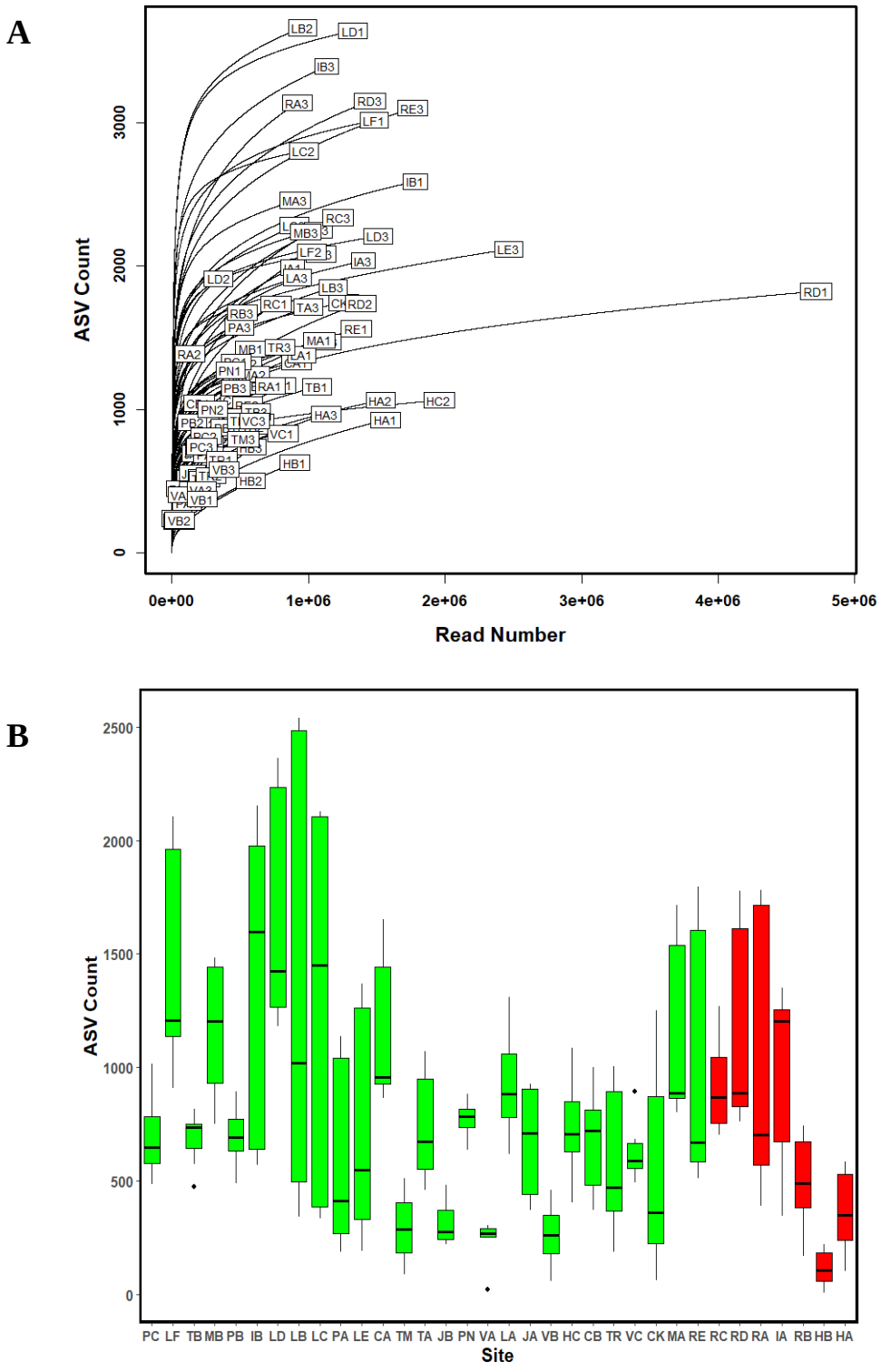


Fig. 5.10 A) Rarefaction curves for archaeal samples for 34 sites, illustrating OTUs richness in relation to sample size. **B)** Box and whisker plot of OTU counts rarefied to a sample size equivalent to the lowest sample size (35 155) across different sites, sites are arranged in increasing order of porewater copper concentrations with sites above and below 20 µg/L indicated in red and green respectively. ANOVA results ($F = 8.4$, $p < 0.001$). Spearman's rank correlation between OTU richness and porewater copper concentration ($rho = -0.25$, $p < 0.001$). Site codes as in **Table 2.1**.

Sample-based accumulation curves (**Fig. 5.11**) were generated to evaluate how cumulative archaeal ASV richness increased with additional sites. In the 12-site dataset (**Fig. 5.11 A**) the curve rose steeply through - roughly - the first 15 samples and then bent gradually toward a plateau, while the 95 % confidence band remained relatively wide. In the 34-site dataset (**Fig. 5.11 B**) the curve began at a higher richness, its slope became shallow by around 80 samples and the confidence band narrowed.

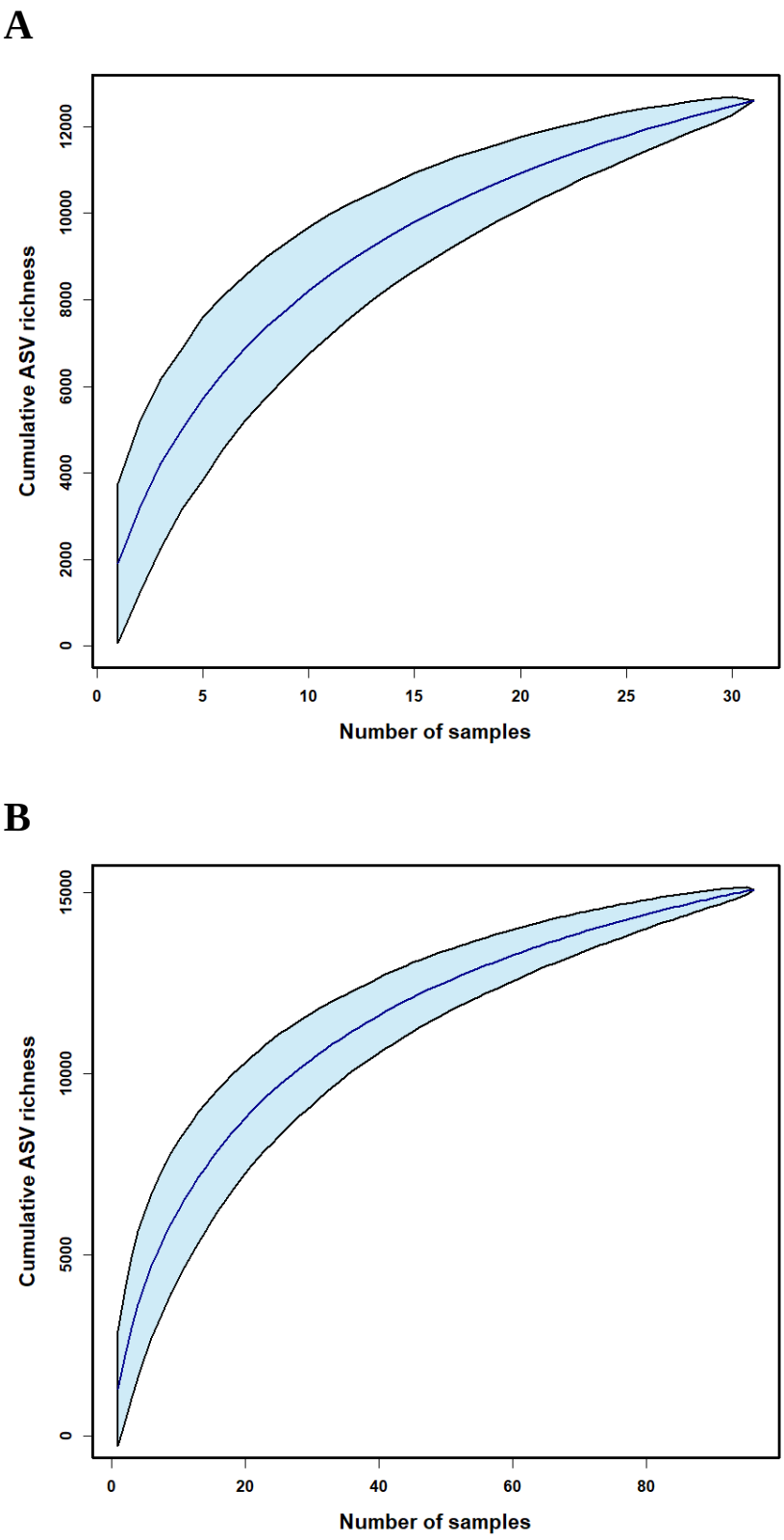
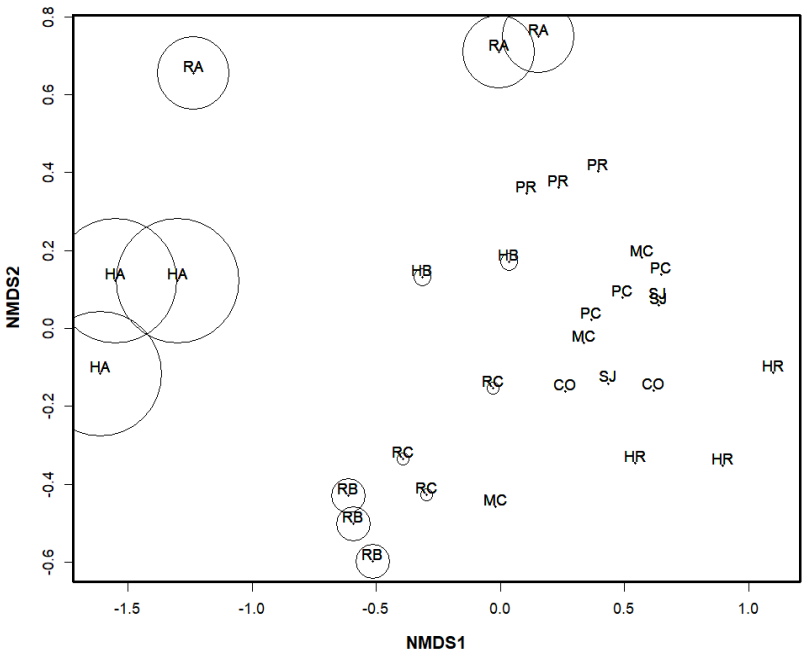


Fig. 5.11 Sample-based species-accumulation curve for archaeal 16S rRNA ASVs recovered from **A)** the 12-site and **B)** 34-site bacterial surveys. The solid line represents the mean cumulative richness and the light-blue polygon the 95 % confidence envelope.

5.3.5 Effect of Rare OTU removal on analysis

In the 12-site dataset (**Fig. 5.12**), removing rare OTUs affected the NMDS ordinations of archaeal communities. Including all OTUs (**Fig. 5.12 A**) produced clustering of samples in which heavily contaminated sites (e.g., HA) were not fully separated from less contaminated locations. Excluding OTUs below the 0.1 % threshold (**Fig. 5.12 B**) increased separation for HA, and a 1 % threshold (**Fig. 5.13**) separated HA further. Raising the threshold removed additional OTUs and reduced overall community richness. At 1 %, heavily contaminated and cleaner sites were separated more than with the lower thresholds.

A All ASVs Retained



B Pruned (<0.1%)

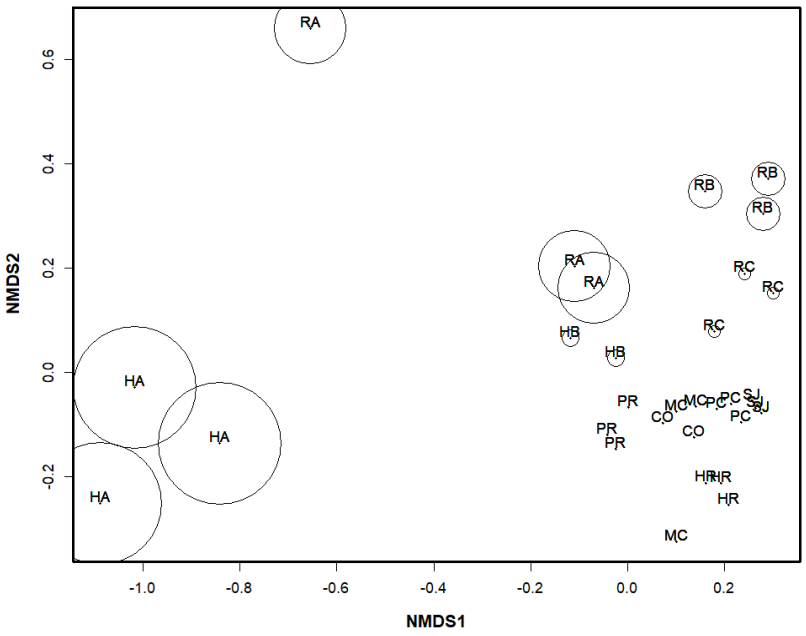


Fig. 5.12 NMDS bubble plots of the 12-site dataset using square root transformed data based on Bray-Curtis similarity of Archaeal abundances. It compares OTU counts across all sites under two conditions: **A)** without removing rare OTUs and **B)** with OTUs occurring at less than 0.1% abundance in all samples removed. Stress values are 0.12 and 0.13, respectively. Bubble sizes represent the pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

Pruned (<1%)

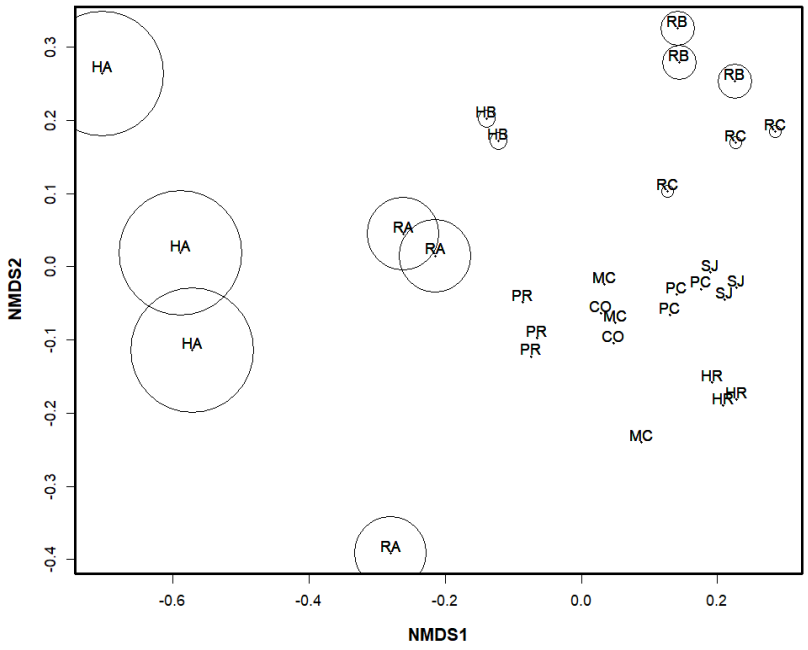
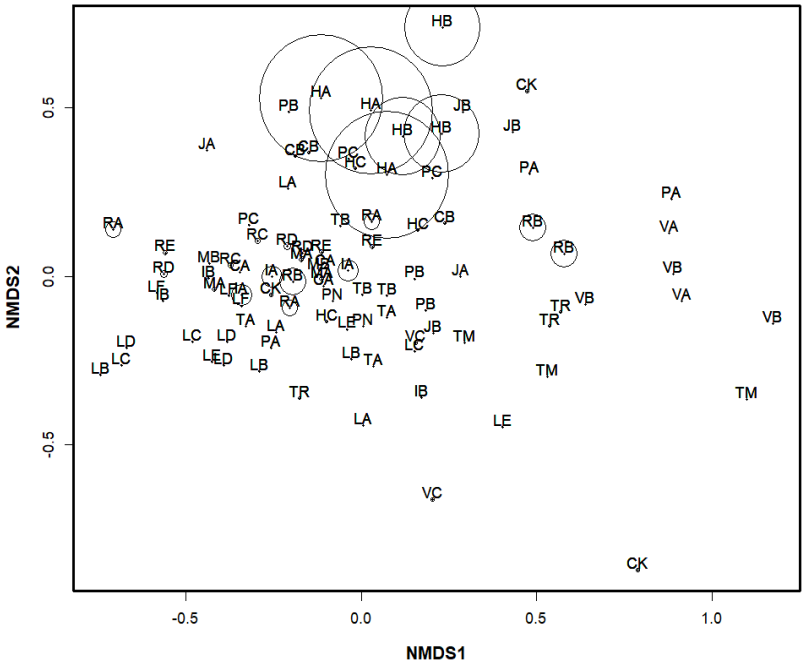


Fig. 5.13 NMDS bubble plot of the 12-site dataset using square root transformed data based on Bray-Curtis similarity of Archaeal abundances. It shows OTU counts across all sites with OTUs occurring at less than 1% abundance in all samples removed. Stress value is 0.12. Bubble sizes represent the pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

For the larger dataset (**Fig. 5.14**), removal of rare OTUs produced modest changes. With no filtering (**Fig. 5.14 A**), samples clustered with limited site-level discrimination. Filtering below 0.1 % (**Fig. 5.14 B**) did not change the ordination appreciably, and the 1 % cut-off (**Fig. 5.15**) gave a small increase in separation for high-copper sites while the overall structure remained similar.

A All ASVs Retained



B Pruned (<0.1%)

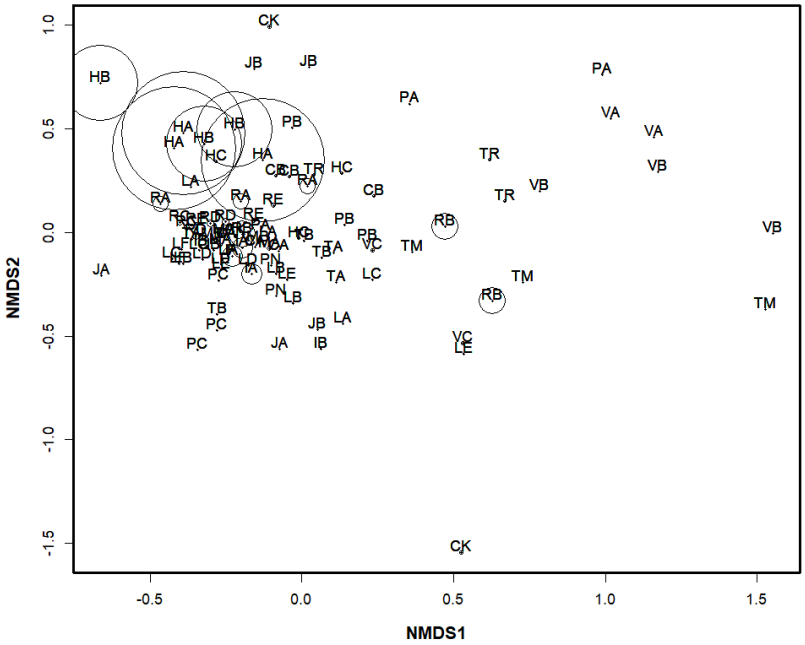


Fig. 5.14 NMDS bubble plots of the 34-site dataset using square root transformed data based on Bray-Curtis similarity of Archaeal abundances. It compares OTU counts across all sites under two conditions: **A)** without removing rare OTUs and **B)** with OTUs occurring at less than 0.1% abundance in all samples removed. Stress values are 0.19 and 0.19, respectively. Bubble sizes represent the pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

Pruned (<1%)

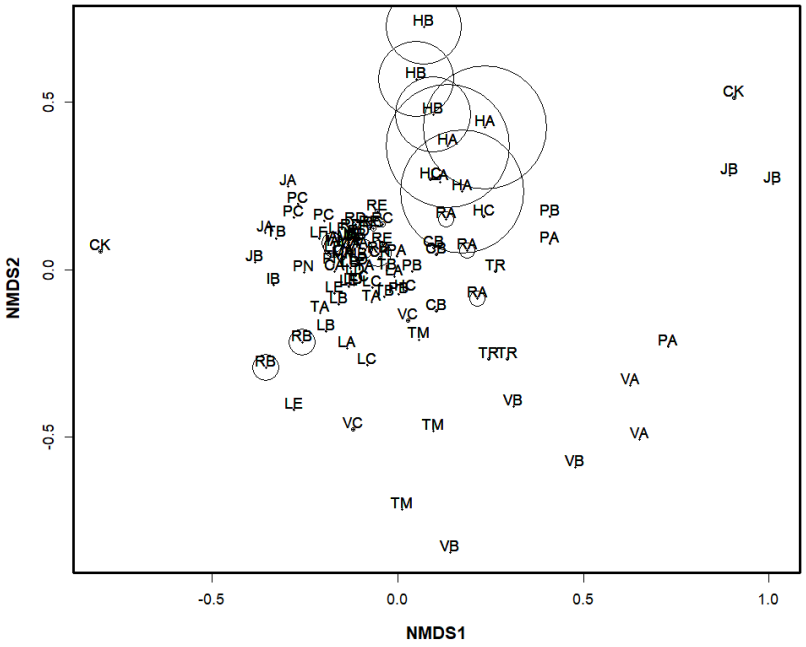


Fig. 5.15 NMDS bubble plot of the 34-site dataset using square root transformed data based on Bray-Curtis similarity of Archaeal abundances. It shows OTU counts across all sites with OTUs occurring at less than 1% abundance in all samples removed. Stress value is 0.16. Bubble sizes represent the pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

ANOSIM results (**Fig. 5.16**) showed that, in the 12-site dataset, R-values increased when OTUs were pruned at 1 % compared with 0.1 % or no pruning. In the 34-site dataset, R-value differences among thresholds were minor, and the 1 % cut-off maintained site separation.

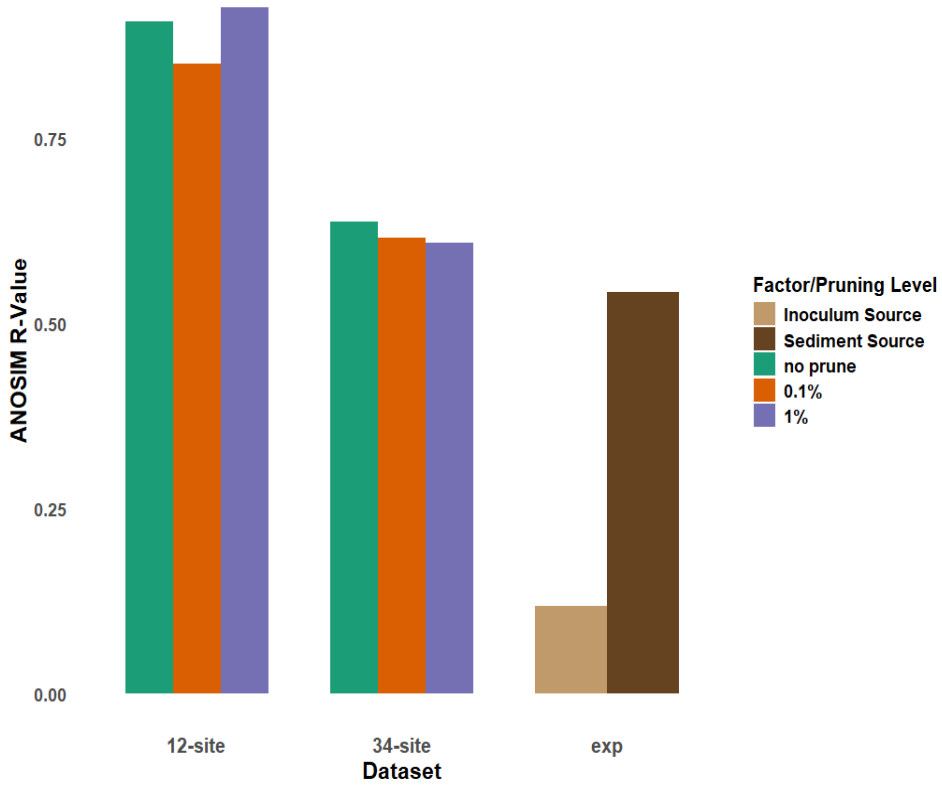


Fig. 5.16 ANOSIM R-values for 12-site, 34-site and experimental datasets generated using the ARF primer under three pruning thresholds (no pruning, 0.1% and 1%). The experimental dataset includes inoculum source and sediment source factors. The results illustrate the degree of dissimilarity between sites, with R statistics quantifying the strength of clustering. R-values > 0.75 indicate very strong differences between groups and all results are statistically significant ($p \leq 0.0001$).

5.3.6 Evaluating Site Grouping Consistency

One rationale for comparing Bray-Curtis with UniFrac was to explore whether phylogenetic relationships might intensify or mask pollution effects among archaeal lineages. In this dataset, the Bray-Curtis MDS plot (**Fig. 5.17 A**) distinguished heavily polluted sites (e.g., HA, RA, HB) on the left, with cleaner sites grouped more to the right. By contrast, the UniFrac MDS plot (**Fig. 5.17 B**) showed a slightly less pronounced separation.

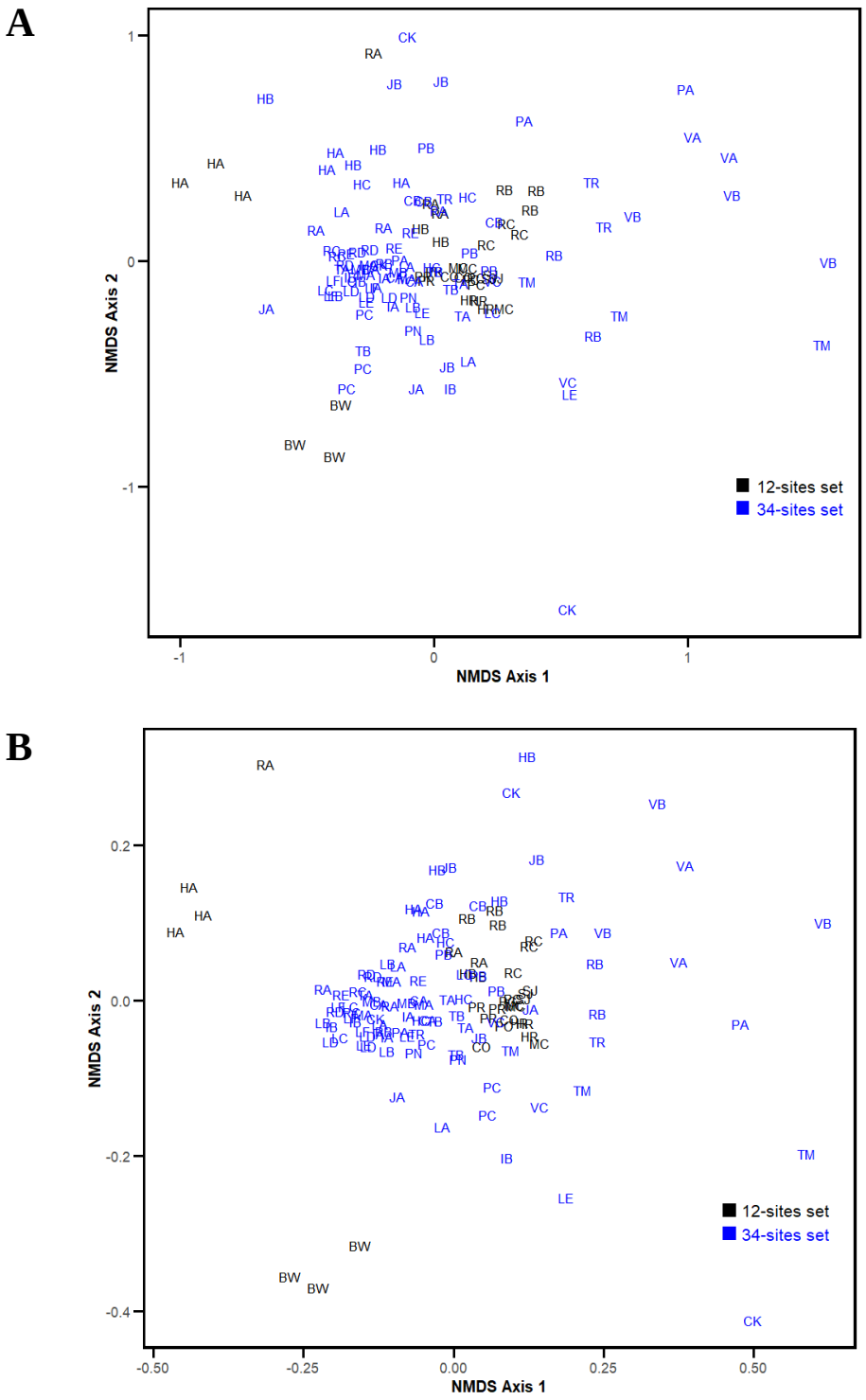


Fig. 5.17 Comparison of **A)** Bray-Curtis and **B)** UniFrac NMDS similarity matrices of square root transformed plots. derived from 12 and 34 site datasets, each two-letter sample label represents site, colours represent the different datasets. No OTUs were pruned. Data processed using LotuS2 in conjunction with the KSPG database. Site codes as in **Table 2.1**.

5.3.7 Comparison of Clustering Patterns Between NMDS and PCoA Analyses

To compare the effectiveness of different ordination methods in illustrating archaeal community composition across pollution gradients, both Non-metric Multidimensional Scaling (NMDS) and Principal Coordinates Analysis (PCoA) were employed. In the NMDS plot (**Fig. 5.18 A**), samples were distributed along both axes, with polluted and cleaner sites occupying different areas of the ordination. The PCoA plot (**Fig. 5.18 B**) grouped samples mainly into two clusters corresponding to polluted and clean sites. The hierarchical clustering dendrogram (**Fig. 5.20 A**) showed that replicates from the same site clustered together. Cleaner sites (PC, SJ, CO, MC) and HR formed one cluster, RB and RC grouped with PR, and the more polluted sites HB and RA formed separate clusters; HA and Breydon Water (BW) each formed individual clusters.

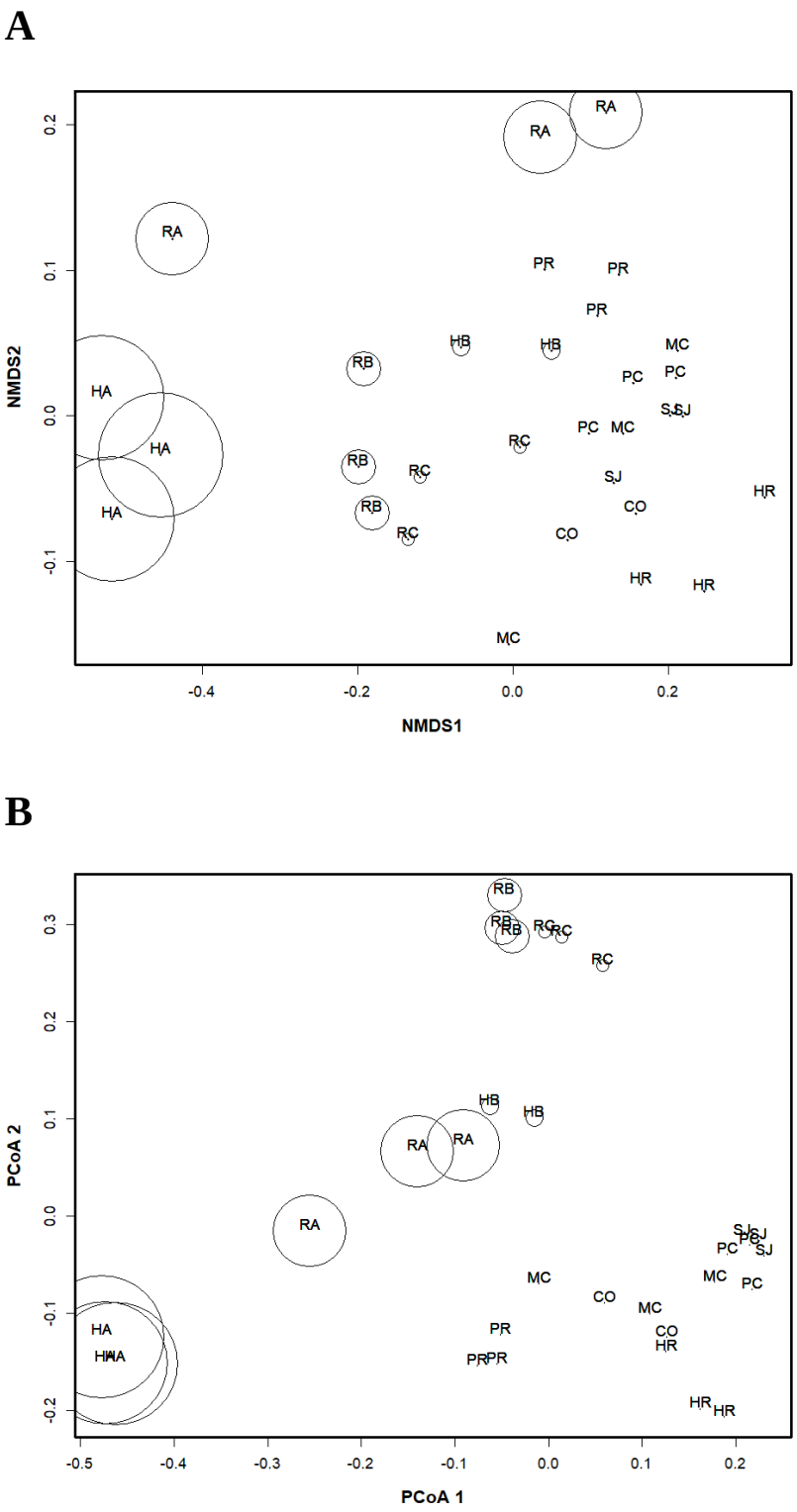


Fig. 5.18 Comparison of **A)** NMDS (Non-metric Multidimensional Scaling) and **B)** PCoA (Principal Coordinates Analysis) of square root transformed plots, with bubble sizes representing PWCu levels to evaluate clustering patterns among sites, derived from 12-site dataset. A 1% OTU pruning threshold was applied in these analyses. Site codes as in **Table 2.1**.

For the 34-site dataset, PCoA ordinations (**Fig. 5.19 A** and **B**) separated the most polluted sites from cleaner ones, and those highly polluted sites were omitted from subsequent analyses focused on other environmental variables. In hierarchical clustering of the 34-site dataset (**Fig. 5.20 B**), sites from the same estuary often clustered together, indicating similarity in archaeal community composition within estuaries. Replicate samples grouped consistently across Bray–Curtis, UniFrac, NMDS and hierarchical clustering, supporting separate analyses for the 12-site and 34-site datasets.

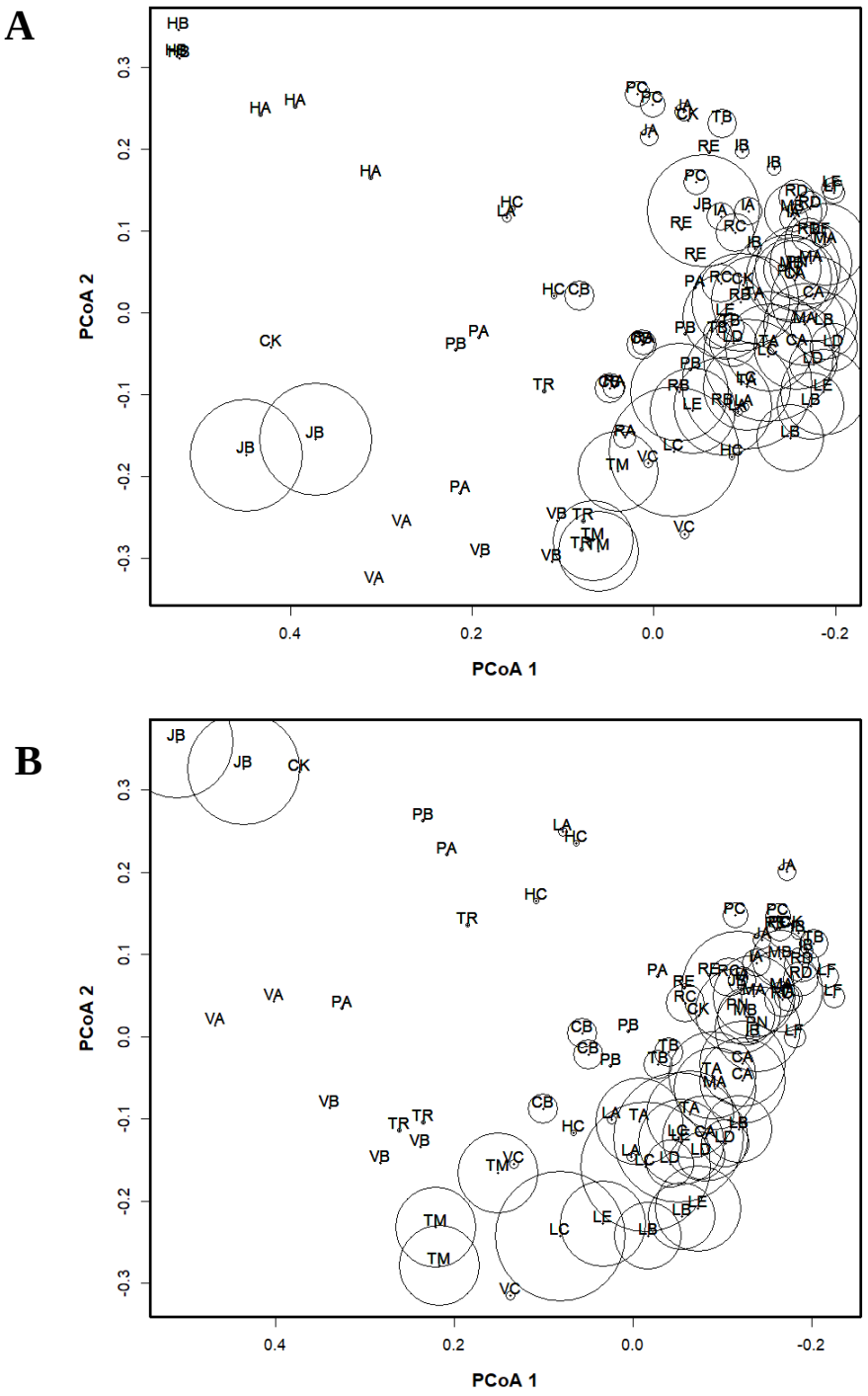


Fig. 5.19 Comparison of PCoA ordination of square root transformed plots, with bubble sizes representing LT63 levels to evaluate clustering patterns among sites, derived from 34-site dataset, **A)** without sites exception and **B)** with the exception of Hayle (A and B) and Restronguet Creek (A and B). A 1% OTU pruning threshold was applied in these analyses. Site codes as in **Table 2.1**.

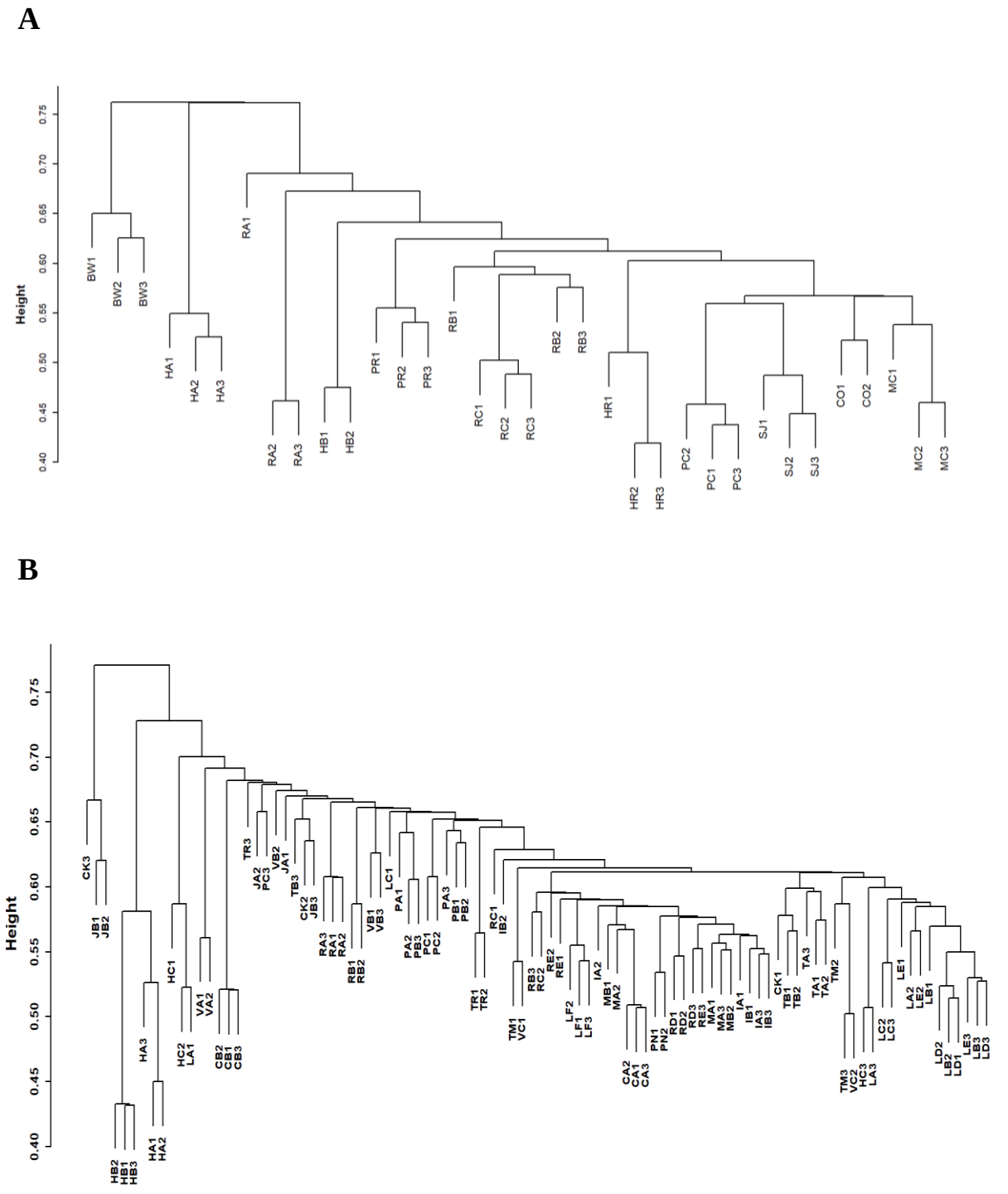


Fig. 5.20 Single linkage hierarchical clustering dendrogram of **A)** 12-Site and **B)** 34-site datasets. Branch heights represent the degree of dissimilarity between site clusters. Bray-Curtis similarity matrices were used, without OTU pruning.

5.3.8 The relationships between ecological pattern and environmental variables

The relationships between environmental variables and microbial communities were assessed using the method of Clarke and Ainsworth (1993), known as BIOENV and based on Spearman's rank correlation. The 12-site dataset revealed that porewater copper (PWCu) had the strongest association with community composition (**Fig. 5.21**). Among individual variables, PWCu alone approached the explanatory power of the best multivariable combination: PWCu, Simultaneously-Extracted Metals (SEMZn), total organic carbon (TOC), dissolved organic carbon (DOC) and porewater organic carbon-bound copper (PW_OC_Cu). Other factors, including sediment-associated copper (EqpCu), TOC, DOC, salinity, sediment grain size (LT63 and D50) and acid-volatile sulphide (AVS), exhibited non-significant correlations.

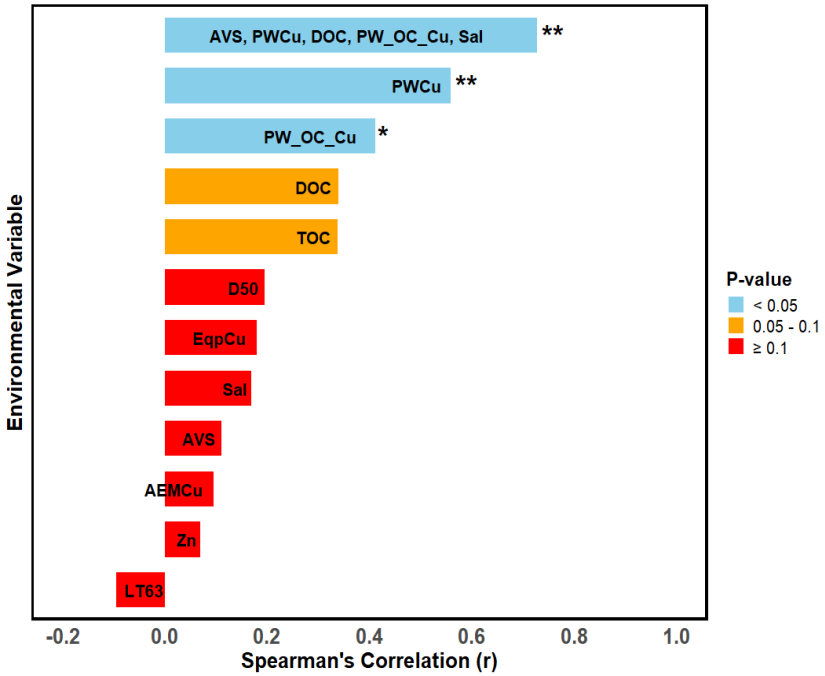
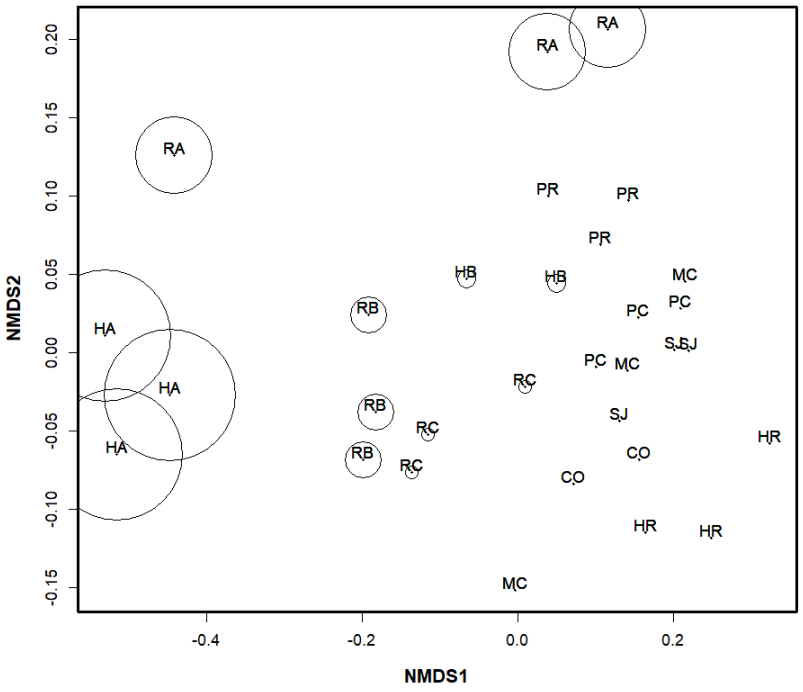


Fig. 5.21 Spearman's rank correlation (Rho) values based on the BIOENV analysis between environmental variables and archaeal community composition for the 12-sites dataset. Bar colours indicate statistical significance: blue; most significant correlations ($p < 0.05$); orange; moderate significance ($0.05 < p < 0.1$); and red; non-significant ($p \geq 0.1$). AEMCu (Acid-Extractable Copper), LT63 (<63 µm fines), AVS (Acid Volatile Sulphide), PWCu (Porewater Cu), TOC (Total Organic Carbon), DOC (Dissolved Organic Carbon), PW_OC_Cu (DOC normalised PWCu), EqPCu (Equilibrium Partitioning Cu), D50 (Median Grain Size), Sal (Salinity), and pH.

The NMDS bubble plots corroborated these BIOENV findings by illustrating how sites with higher PWCu and PW_OC_Cu form distinct clusters (**Fig. 5.22 A** and **B**), echoing the strong correlations highlighted in the statistical analysis. Conversely, variables such as TOC, AVS (**Fig. 5.23 A** and **B**), LT63 and PWZn did not show clear patterns in their bubble plots (**Fig.** (**Fig. 5.24 A** and **B**), aligning with their non-significant results in the BIOENV analysis.

A PWCu



B PW_OC_Cu

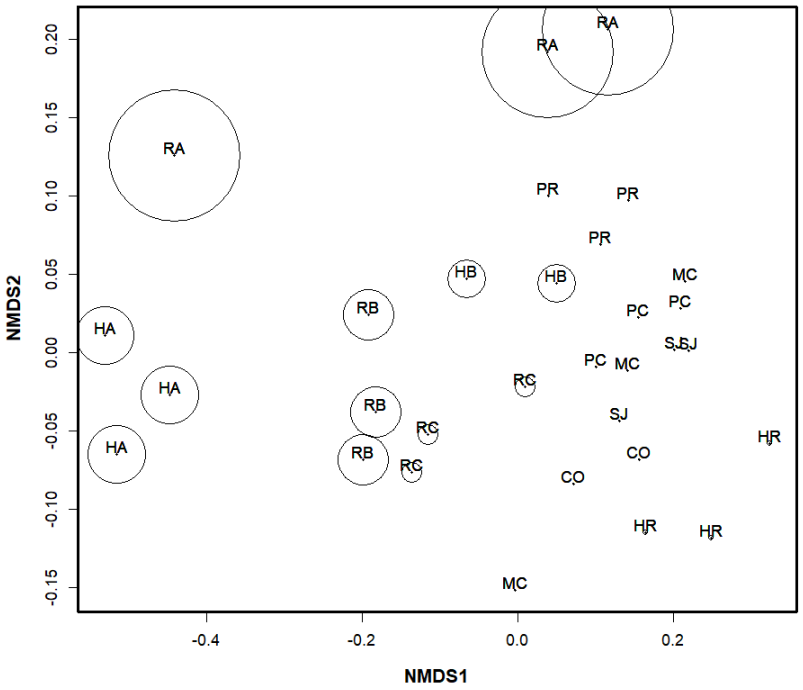
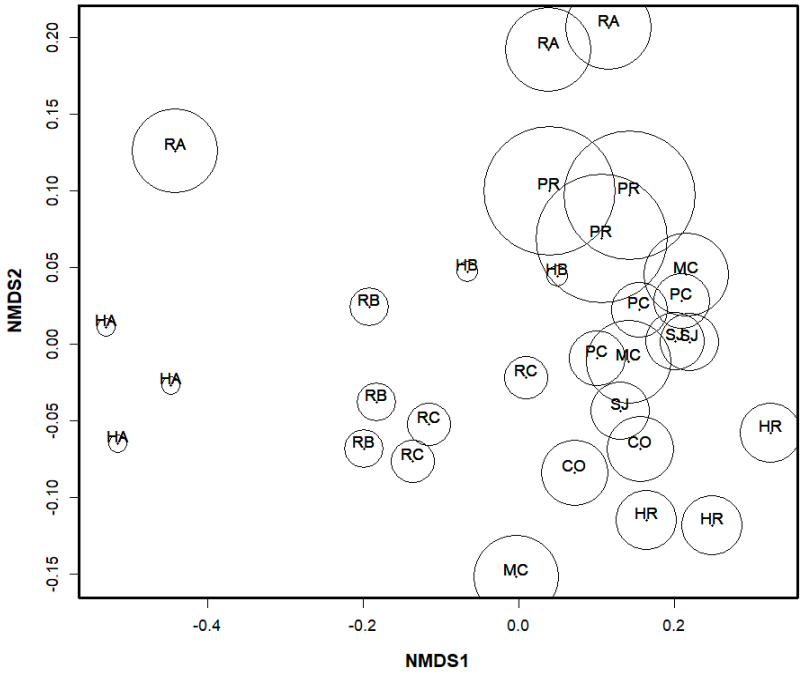


Fig. 5.22 NMDS plots of square root transformed data of Archaeal abundances illustrating the spatial patterns of the 16S Archaeal community structure derived from the 12-site dataset. **A)** shows the relationship with PWCu and **B)** with PW_OC_Cu. Bray-Curtis similarity matrices were used, without OTU pruning. Site codes as in **Table 2.1**.

A TOC



B AVS

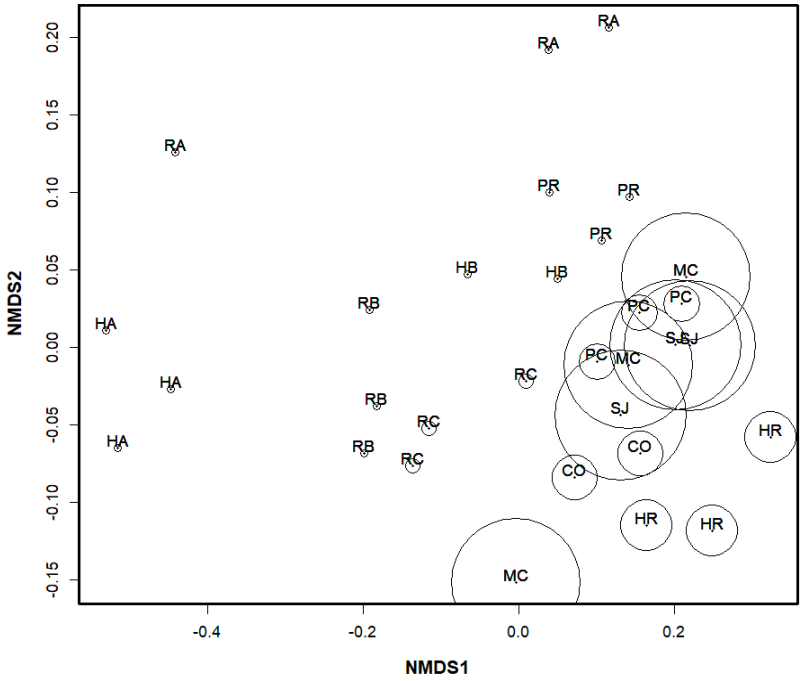
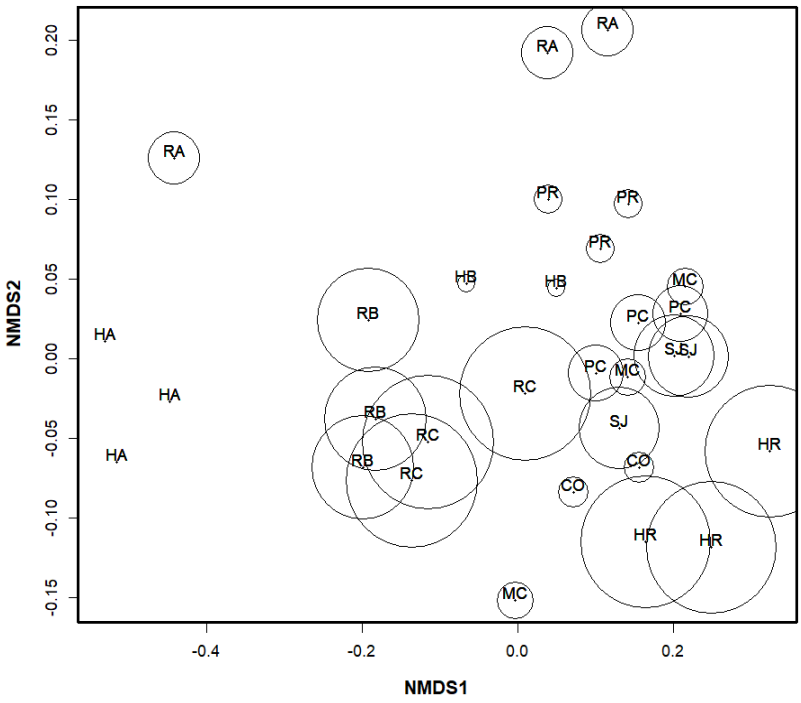


Fig. 5.23 NMDS plots of square root transformed Archaeal abundances from the 12-site dataset showing relationships with environmental variables. **A)** shows the relationship with TOC and **B)** with AVS. Bray-Curtis similarity matrices were used, without OTU pruning. Site codes as in **Table 2.1**.

A LT63



B PWZn

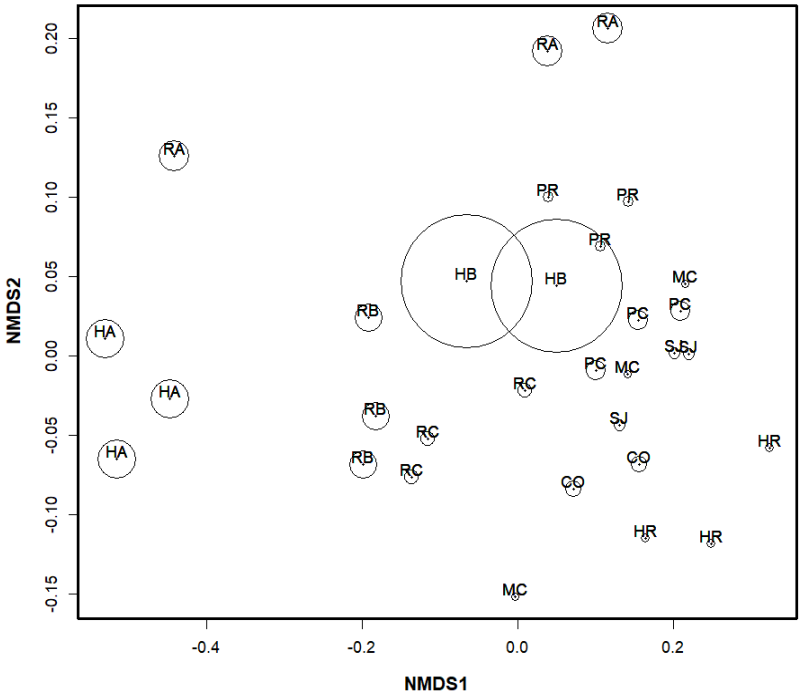


Fig. 5.24 NMDS plots of square root transformed Archaeal abundances from the 12-site dataset showing relationships with **A)** LT63 and **B)** PWZn. Bray-Curtis similarity matrices were used, without OTU pruning. Site codes as in **Table 2.1**.

In contrast, the BIOENV analysis for the 34-site dataset (**Fig. 5.25 A**) indicated that porewater copper (PWCu), the proportion of fine particles (LT63), median grain size (D50) and salinity collectively offered the strongest explanation for archaeal community composition. Excluding the most heavily polluted sites (HA, HB, RA and RB) rendered PWCu non-significant, while LT63 and salinity remained strong correlates, (**Fig. 5.25 B**).

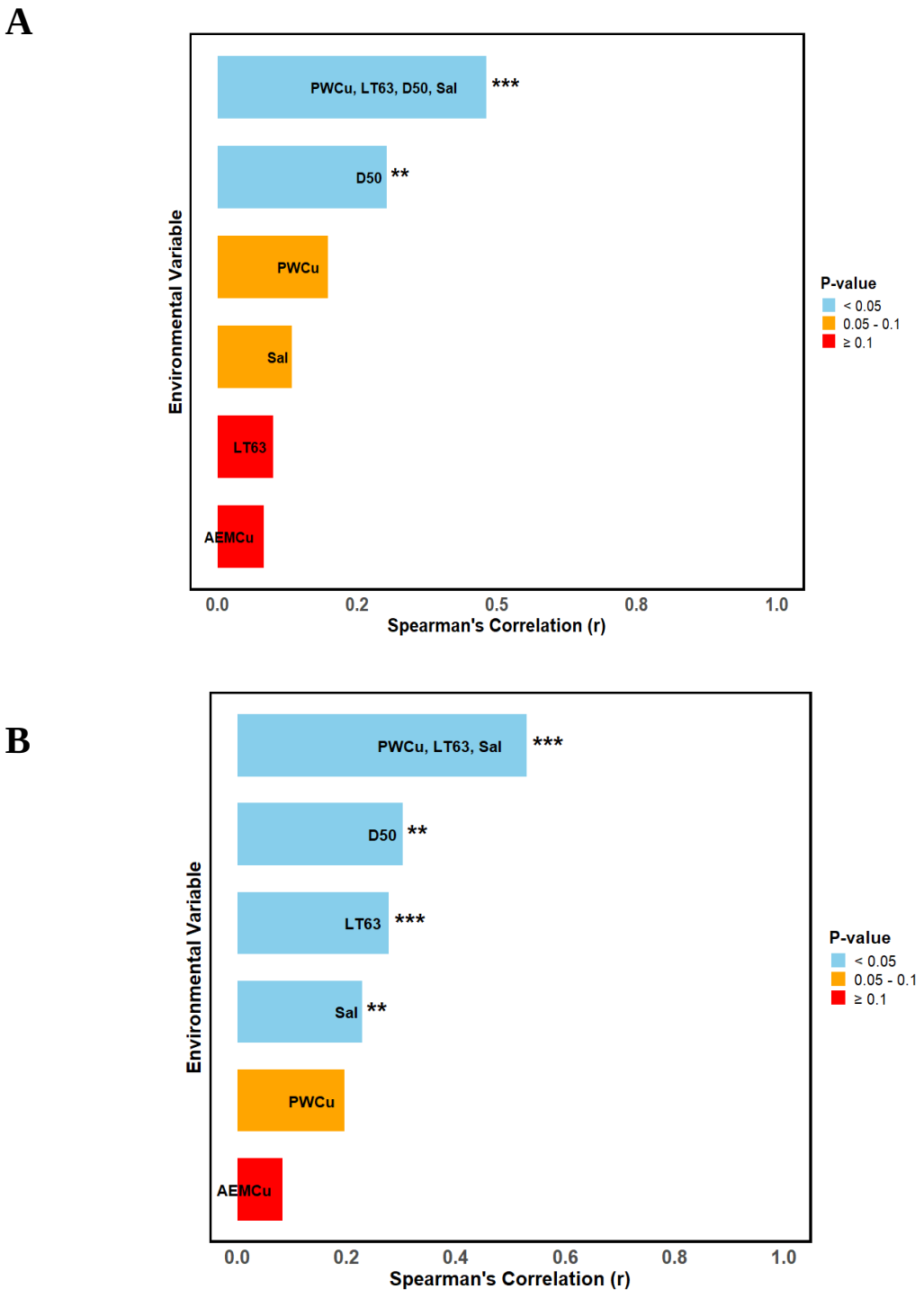
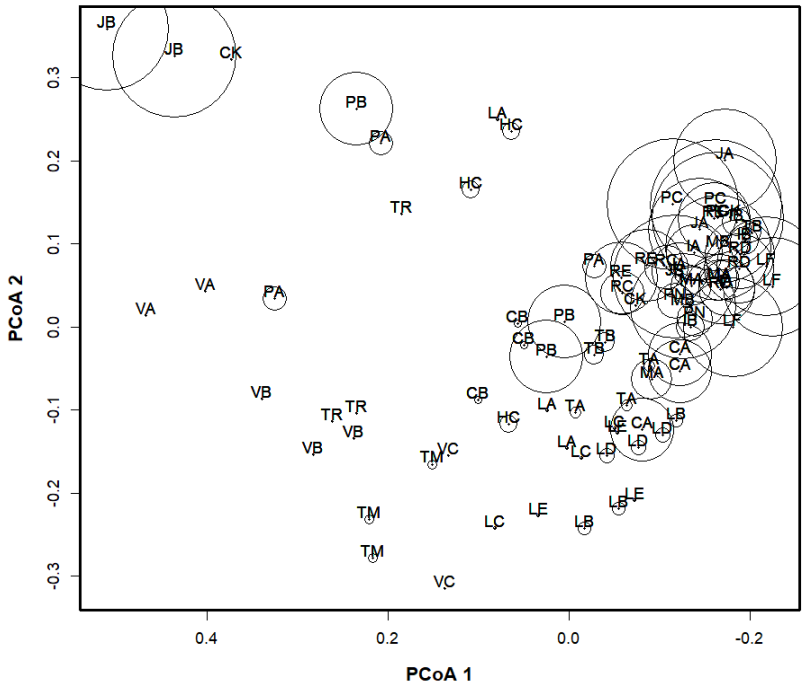


Fig. 5.25 Spearman's rank correlation (Rho) values based on the BIOENV analysis between environmental variables and archaeal community composition. **A)** shows the results for the 34-sites dataset, and **B)** for the 34-sites dataset excluding highly polluted sites (HA, HB, RA and RB). Bar colours indicate statistical significance: blue; most significant correlations ($p < 0.05$); orange; moderate significance ($0.05 < p < 0.1$); and red; non-significant ($p \geq 0.1$). PWCu (Porewater Cu), AEMCu (Acid-Extractable Cu), D50 (Median Grain Size), LT63 (<63 µm Fines), Sal (Salinity).

The (PCoA) analysis (**Section 5.3.7**) and PWCu in (**Fig. 5.27 B**) highlighted a clearer separation between the most and least contaminated sites. The bubble plots shown, when highly polluted sites (HA, HB, RA and RB) were excluded from the analysis, Salinity (**Fig. 5.26 A**) and LT63 (**Fig. 5.26 B**) displayed the best patterns. D50 (**Fig. 5.27 A**) remained did not exhibit clear patterns. Principal Components Analysis (PCA), detailed in **Chapter 2**, further underscores D50 and fine particles (LT63) importance through its prominent vector.

A Salinity



B LT63

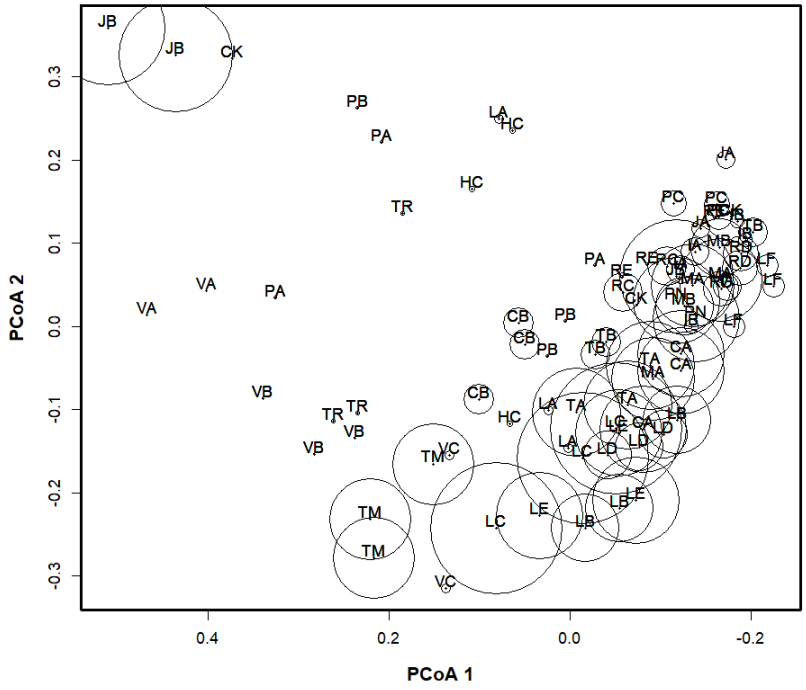
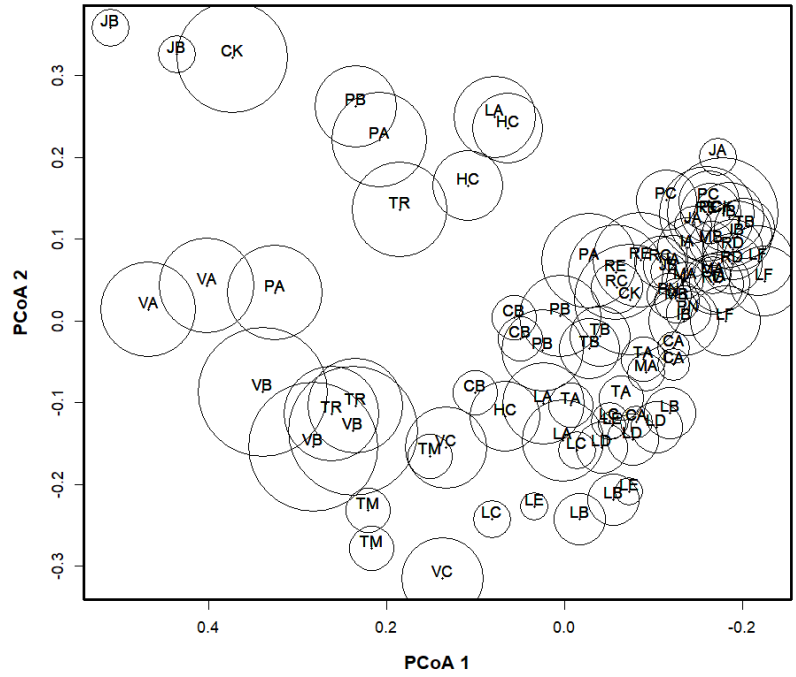


Fig. 5.26 PCoA plots of square root transformed data of Archaeal abundances illustrating the spatial patterns of the 16S Archaeal community structure derived from the 34-site dataset in relation to key environmental variables. **A)** shows the relationship with salinity and **B)** with LT63. Bray-Curtis similarity matrices were used, with a 1% OTU pruning threshold. The most heavily polluted sites (HA, HB, RA and RB) were excluded from both plots. Site codes as in **Table 2.1**.

A D50



B PWCu

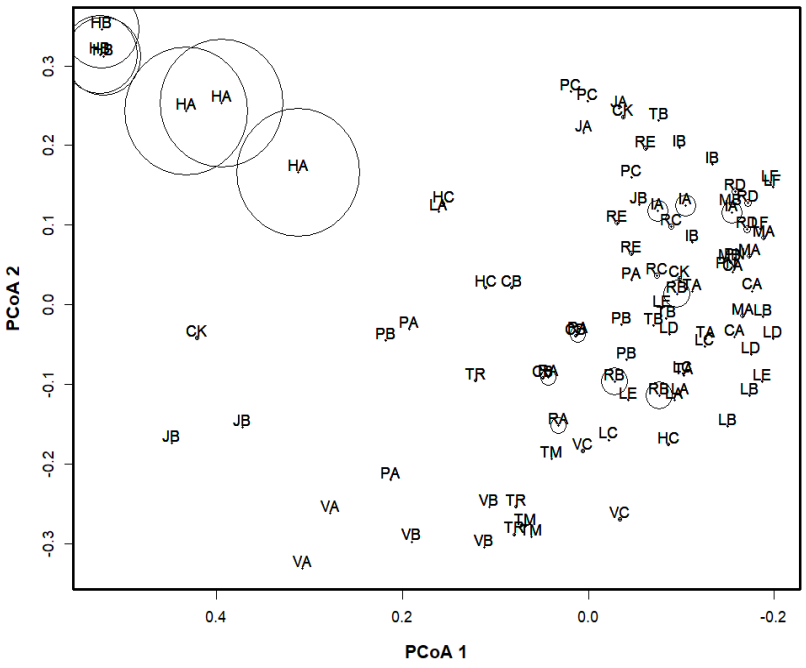


Fig. 5.27 PCoA plots of square root transformed Archaeal abundances from the 34-site dataset showing relationships with environmental variables. **A)** shows the relationship with D50, excluding the most heavily polluted sites, while **B)** shows the relationship with PWCu including all sites. Bray-Curtis similarity matrices were used, with a 1% OTU pruning threshold. Site codes as in **Table 2.1**.

Across both datasets, the clearest pattern was the distinctiveness of the highly contaminated sites in Hayle and Restronguet Creek. This difference was strongly correlated with elevated porewater copper levels in the 12-site data set but falls below statistical significance in the site dataset when environmental variables were examined individually.

5.3.9 Identifying Copper-Tolerant and Sensitive taxa

To identify phyla that were tolerant or sensitive to copper pollution, OTU counts were aggregated to the phylum level. Average abundances in clean sites (ava) and polluted sites (avb) were calculated for each phylum and the fold change between polluted and clean sites was determined along with p-values. Phyla and OTUs that did not show statistically significant differences ($p > 0.05$) were not included in this analysis, as they did not provide evidence of a meaningful association with copper levels. The heatmaps in **Fig. 5.28** illustrate the top ten phyla showing differences across both datasets; 12 sites (**A**) and 34 sites (**B**).

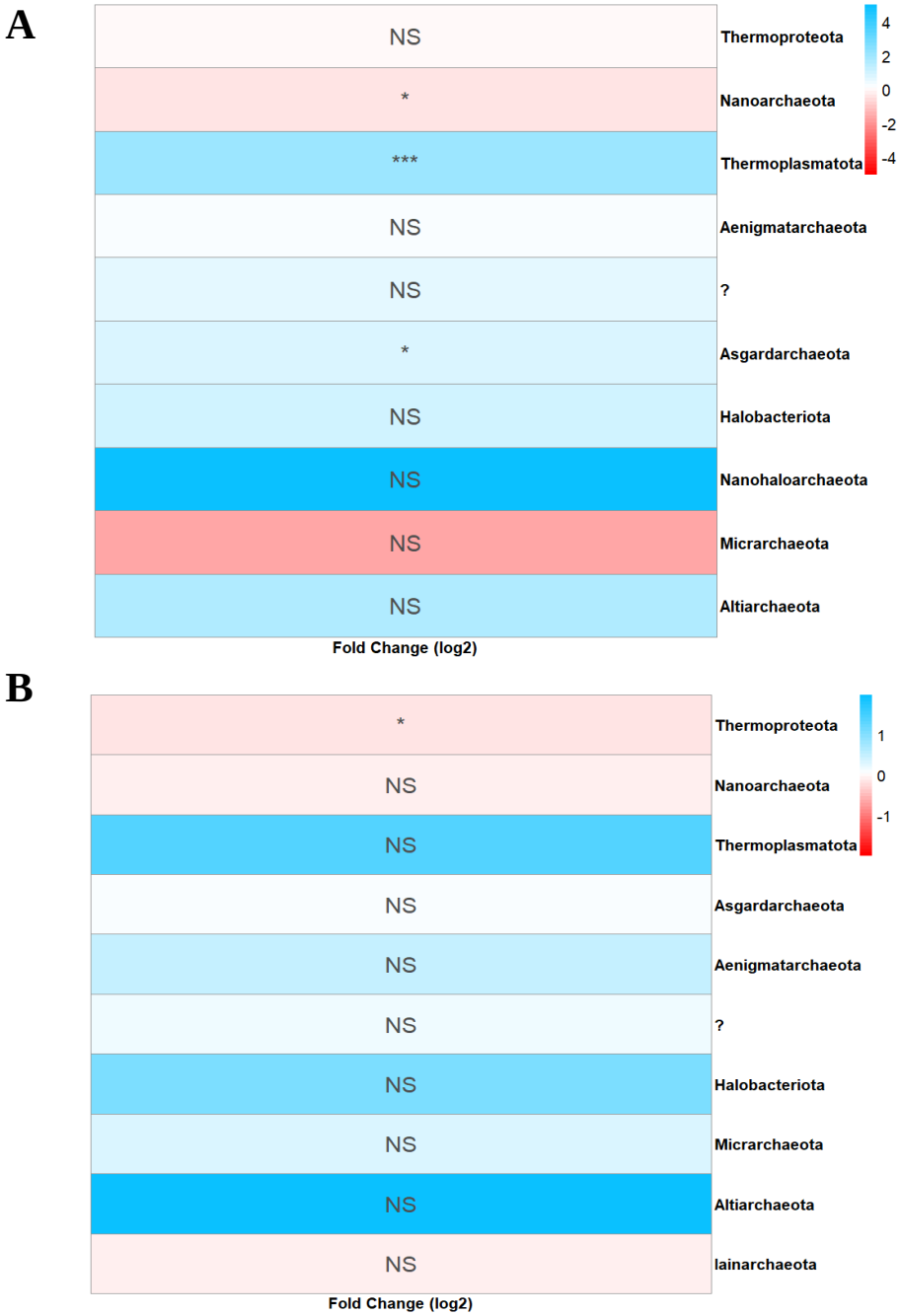


Fig. 5.28 Heatmap illustrating the top 10 phyla exhibiting significant differences in abundance between polluted and clean sites across **A)** 12-site and **B)** 34-site datasets, along with their p-values and raw fold changes. The phyla were sorted based on their average SIMPER contribution to dissimilarity between groups. Statistical significance is indicated by asterisks (*$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$), with "NS" denoting non-significant results. Darker colour intensities represent higher raw fold changes, blue shades indicate positive fold changes (more abundant in clean sites), white represents minimal or no change and red shades indicate negative fold changes (more abundant in polluted sites). Raw fold changes [$\log_2(ava^2/avb^2)$] revert the square-root-transformed averages to the original abundance scale. Taxonomic assignments were made using the KSPG database. '?' means unresolved taxonomy by LCA in KSGP.

In the 12-site dataset, three phyla showed statistically significant differences in abundance between clean and polluted sites. Thermoplasmatota and Asgardarchaeota were more abundant in clean sites, whereas Nanoarchaeota showed greater abundance in polluted sites. Phyla such as Thermoproteota and Aenigmatarchaeota did not exhibit statistically significant differences. In comparison, in the 34-site dataset, only Thermoproteota demonstrated a significant difference in abundance between high and low copper sites, showed increase abundance in polluted environments (**Fig. 5.28 B**).

5.3.10 Identifying Copper-Tolerant and Sensitive Taxa at the OTU Level

In both the 12-site and 34-site datasets (**Fig. 5.29**), certain OTUs displayed statistically significant differences in abundance between clean and polluted environments. In the 12-site dataset, OTU4, OTU6, OTU8, OTU9, OTU12 and OTU17 were consistently more abundant in clean sediments. By contrast, the 34-site dataset reinforced the copper sensitivity of OTU3, OTU4 and OTU12, which again showed increased abundance in clean sites. All OTUs belong to the phylum Thermoproteota, except OTUs 9, 12 and 17, which were Thermoplasmatota.

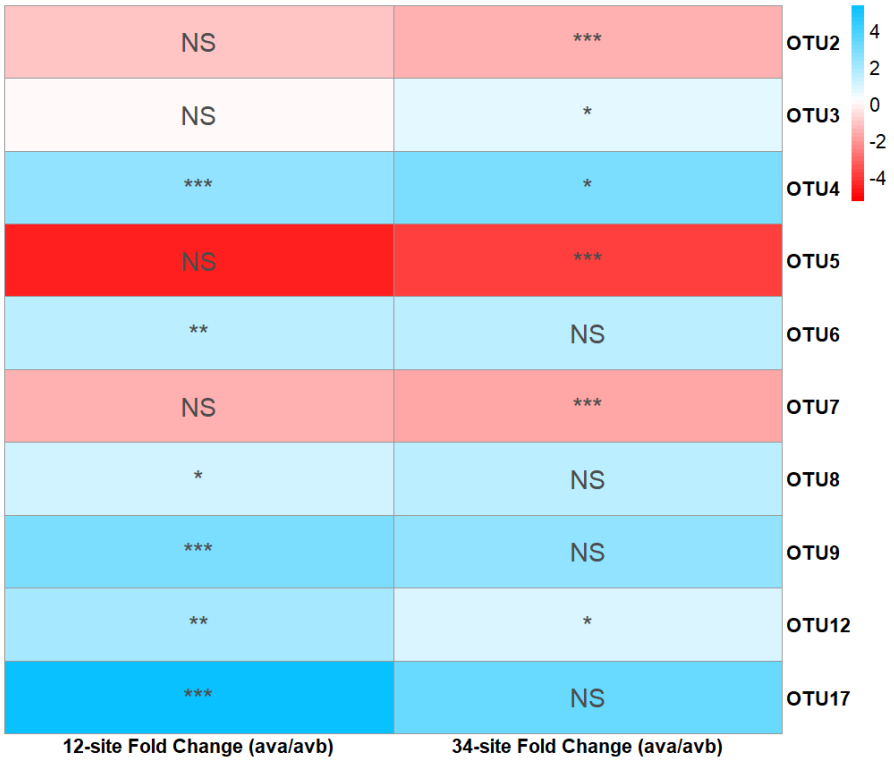


Fig. 5.29 Heatmap illustrating the top 10 OTUs exhibiting significant differences in the 12-site and/or 34-site analyses, along with their p-values and raw fold changes. The OTUs were selected and numerically sorted based on their average SIMPER contribution to the dissimilarity between groups. Statistical significance is indicated by asterisks ($p \leq 0.05$; *$p \leq 0.01$; **$p \leq 0.001$), with "NS" denoting non-significant results. Darker colour intensities represent higher raw fold changes, blue shades indicate positive fold changes (more abundant in clean sites), white represents minimal or no change and red shades indicate negative fold changes (more abundant in polluted sites). Raw fold changes [$\log_2(ava^2/avb^2)$] revert the square-root-transformed averages to the original abundance scale. Taxonomic assignments were made using the KSPG database. All OTUs belong to phylum Thermoproteota, except OTUs 9, 12 and 17 which were Thermoplasmatota.

5.4 Discussion

Archaea are fundamental components of microbial communities in estuarine and marine sediments (Petro et al., 2017), yet their taxonomic breadth has frequently been underestimated due to limited primer coverage. By employing recently published, broad-coverage primer set, this study captured archaeal lineages across multiple phyla, including Nanoarchaeota (formerly Woesearchaeota) which previous methods often failed to detect (Liu et al., 2021). Here, Nanoarchaeota accounted for about 40% of reads and 80% of OTUs, matching the dominance of Nanoarchaeota and Thermoproteota reported by Zhang et al. (2024a) in similar benthic

habitats. Although this proportion was higher than the 10.5% Woesearchaeota observed by Sun et al. (2023) such discrepancies may stem from differences in primer specificity, local environmental conditions, or taxonomic classifications (Salmaso et al., 2022). Their ultrasmall genomes and obligate syntrophic partnerships with hydrogen-consuming methanogens enable Nanoarchaeota to thrive in anoxic, metal-rich sediments, so a rise in their relative abundance is considered a useful bioindicator of persistent hypoxia and redox stress (Liu et al., 2018, Huang et al., 2021, Liu et al., 2021). A similar depth-linked enrichment of Woesearchaeota in progressively anoxic horizons has been reported from Pacific abyssal sediments (Peoples et al., 2019), further supporting their diagnostic potential.

Across 12 and 34 estuarine sites, porewater copper ranged from about 2 to 431 µg/L, while sediment extractable copper reached up to 1798 µg/g, creating a wider gradient than many previous studies (Voica et al., 2016, Zou et al., 2020, Coppo et al., 2023). The markedly flatter tail of the 34-site archaeal accumulation curve suggests that expanding site coverage is especially effective for this domain, whose richness appears to saturate more quickly once moderate spatial replication is achieved. Variation among replicates was low relative to variation among sites, indicating that local environmental differences primarily drove changes in archaeal community composition. Although archaeal communities can be quite diverse in moderate or near-pristine conditions (Chen et al., 2022, Zhao et al., 2022), their richness was lower than that of bacterial assemblages. While archaeal OTU counts were lower overall, it still showed marked shifts in response to copper contamination, suggesting that certain archaeal groups may be sensitive indicators of metal stress. In the 12-site subset, archaeal diversity showed a clear decline once PWCu surpassed roughly 20 µg/L, whereas in the larger 34-site dataset, copper's impact appears less pronounced and some moderately contaminated sites (e.g., Percuil B and C) exhibit diversity similar to more polluted locations. Highly contaminated sites do showed differences in community composition in both analyses, though in the 34-site dataset only HA and HB stood out as having distinctly altered archaeal assemblages. These patterns suggest that severe copper stress can reshape archaeal communities, though these effects may diminish or become confounded when environmental heterogeneity was greater.

Comparing archaeal and bacterial responses to copper contamination in the 12-site subset indicated that both domains show significant negative correlations between diversity and PWCu, but archaea exhibit a stronger correlation (rho = -0.64 vs. -0.23 for bacteria). This could

imply that archaea were more sensitive to moderate copper loads, possibly due to lower functional redundancy or different detoxification pathways (Sandaa et al., 1999, Yu et al., 2024). However, in the 34-site dataset, the correlation for archaea (rho = -0.25) was weaker and it was more comparable to the bacterial pattern (rho = -0.16), suggesting that under broader environmental gradients, factors like salinity or sediment texture can obscure copper's direct effects. Consequently, while the 12-site data suggest archaeal diversity can markedly decline once PWCu surpasses 20 µg/L levels that only moderately affect bacterial communities, this threshold was less apparent in the larger and more heterogeneous 34-site analysis. Still, these findings collectively indicate that archaea can serve as an early-warning indicator of metal stress in relatively uniform estuaries, but not necessarily across all settings (Chen et al., 2024).

Detecting the ecological consequences of heavy metal pollution, particularly copper, on microbial communities remains a central challenge in environmental monitoring (Bååth, 1989). In aquatic environments, the complexity of natural gradients complicates efforts to isolate the effects of pollutants, including heavy metals (Grant, 2010). Although earlier moderate-contamination studies (Chen et al., 2019) and extreme-environment research (Korzhenkov et al., 2019) suggest high copper loading can markedly alter archaeal communities at concentrations less impactful to bacteria, this pattern arises mainly in the 12-site subset here. In the 34-site dataset, both archaea and bacteria responded similarly to copper, shaped by multiple factors. This result nevertheless indicated that under certain localized conditions, such as those encountered in heavily polluted Hayle or Restronguet Creek sites, archaea may serve as sensitive indicators of metal stress.

Within that smaller subset, heavily contaminated sites exhibit markedly reduced OTU richness and a higher relative abundance of certain putative copper-tolerant OTUs belonging to Thermoproteota (previously classified as separate lineages of Crenarchaeota, Thaumarchaeota and Bathyarchaeota in older NCBI-based schemes; (Rinke et al., 2021)) or Nanoarchaeota (encompassing Woesearchaeota in older classification). By contrast, the more heterogeneous 34-site dataset revealed a weaker copper diversity relationship, with sediment grain size (D50), salinity and organic content also influencing archaeal composition. These findings align with Zou et al. (2020), who observed that multiple environmental gradients can mask metal impacts, illustrating how copper's effect on archaeal assemblages becomes less pronounced when other variables vary extensively. In comparing to other studies, many still use older taxonomy (e.g.,

Bathyarchaeota, Thaumarchaeota) instead of the updated GTDB designations (Thermoproteota).

Within contaminated sites, particular archaeal taxa showed increased abundance, indicating the presence of copper-tolerant lineages that could serve as bioindicators of metal contamination (Besaury et al., 2014, Voica et al., 2016, Gupta et al., 2021). Although Thermoproteota's resilience under metal stress (Carlier et al., 2020, Zheng et al., 2022) was supported particularly in the 34-site dataset, Nanoarchaeota appears more prevalent at heavily polluted locations mainly in the 12-site analysis. This discrepancy likely reflects different sampling seasons between the two datasets, rather than a site or context-specific response. Both Thermoplasmatota and Asgardarchaeota tend to be more abundant in cleaner environments, aligning with observations of Thermoplasmatota declines under heavy pollution (Di Cesare et al., 2020, Schenk et al., 2022) and Asgardarchaeota's preference for less disturbed conditions (Cai et al., 2021). However, these lineage-level relationships exhibited modest variation between the 12- and 34-site datasets. Notably, in one dataset, the top seven OTUs belonged to Thermoproteota, reflecting potential differences in copper sensitivity even within a single phylum. Overall, these findings highlight the value of specific taxa as putative indicators of severe copper pollution, although the ability to confidently identify these lineages remains constrained by incomplete reference databases and evolving classification systems (Krzmarzick et al., 2018, Grant et al., 2023). These lineage-specific findings expand upon broader mechanistic reviews such as (Voica et al., 2016). by pinpointing specific phyla and OTUs that respond to severe copper pollution.

Analytical scale, whether focusing on a smaller, relatively uniform subset of sites or a broader, more varied set, strongly influences how pollution impacts were detected. For instance, the 12-site dataset (concentrated mainly on Fal and Hayle estuaries) revealed a more pronounced copper effect because fewer confounding variables (e.g., salinity or sediment texture) overshadowed its influence, whereas the larger 34-site dataset encompasses greater environmental heterogeneity, reducing copper's apparent role. NMDS and PCoA thus highlighted copper-driven changes more distinctly in the 12-site analysis, whereas the 34-site ordinations emphasize how salinity, D50 and organic matter can also shape archaeal distributions. Although filtering rare OTUs can be debatable (Ramette, 2007, Navas et al., 2021, Nikodemova et al., 2023), removing very low-abundance taxa helps focus on dominant

groups such as Thermoproteota and Nanoarchaeota, particularly as rare taxa can inflate richness counts but may be transient or minimally adapted to environmental extremes (Yenni et al., 2017). Nonetheless, recent Arctic marine sediment analyses indicate that some rare archaea remain functionally limited, whereas others actively shape community assembly, underscoring that low-abundance populations do not necessarily lack ecological impact (Sun et al., 2023). Understanding how these minor lineages persist was therefore crucial, since not all rare organisms contribute equally to resilience under stress (Yenni et al., 2017).

This analysis showed that archaeal communities were notably sensitive to copper pollution, especially in relatively uniform environments where confounding variables were fewer. The reduction in archaeal diversity, coupled with the presence of putative copper-tolerant groups, suggests the potential for archaeal-based bioindicators, although environmental heterogeneity and the need for robust taxonomic frameworks advise caution. Overall, these findings confirm that archaeal communities exhibit a measurable response to high copper levels, often matching or surpassing the thresholds reported for other microbial indicators. They also suggest that archaea may respond more sharply than bacterial communities in some instances. Consequently, integrating archaeal data into wider monitoring and management strategies could enhance the detection and mitigation of metal pollution in estuarine ecosystems.

Chapter 6:

Examining Eukaryotic Primer Biases and Taxon-specific Responses to Environmental Variables in Contaminated Estuaries

6.1 Introduction

Environmental monitoring in marine ecosystems has traditionally relied on macrofauna, which require large sample sizes due to their body dimensions (Kendall and Widdicombe, 1999). The processing and identification of macrofaunal samples are labour-intensive, expensive and demand specialist taxonomic expertise. In estuarine habitats, the relatively limited number of macrofaunal species reduces the sensitivity of these methods in detecting subtle pollution impacts (Warwick et al., 1991). To address these drawbacks, meiobenthic groups, especially nematodes, have been proposed as valuable alternatives. Morphological analyses of meiofauna have long provided critical insights into pollution effects (Coull, 1992, Semprucci et al., 2015) and nematodes, in particular, serve as effective bioindicators due to their diverse ecological strategies, small body size and rapid turnover rates in response to metal stress (Heip et al., 1985, Schratzberger et al., 2006).

Several authors have argued that meiofauna offer a more practical option than macrofauna for ecological assessments, requiring smaller sample volumes and often exhibiting higher diversity in low-salinity environments (Moore and Bett, 1989, Coull, 1992). However, their processing was more complex, involving extra steps such as flotation, elutriation and sieving (Hummon, 1981, Giere, 2009). While nematodes pose identification challenges comparable to macrofauna, other meiofaunal taxa such as harpacticoid copepods and soft-bodied Platyhelminthes, demand even greater taxonomic expertise (Balsamo et al., 2020). Additionally, hard-shelled groups (e.g., foraminiferans and ostracods) require specialised handling because they cannot be separated using standard flotation (Giere, 2009, Martin et al.,

2021). Despite these limitations, evidence from traditional morphology-based studies suggests that metal pollution can influence various meiofaunal taxa, yet the full extent of these impacts remains inconclusive across different estuarine systems.

Smaller eukaryotes that pass through a 63 µm sieve which covering both diminutive metazoans and single-celled microbes, have received comparatively little attention in morphological surveys. By contrast, diatoms have been studied extensively using standard surface-sampling approaches based on their mobility and tendency to grow on sediment surfaces (Eaton and Moss, 1966). This emphasis on specific, easily sampled groups highlights the gaps in current scientific understanding of many other unicellular or microscopic lineages that might be equally sensitive to heavy metal stress. Current evidence does not clearly distinguish how smaller microbial eukaryotes respond to metals, revealing a substantial gap in the literature.

Metabarcoding has been proposed as an alternative to traditional morphological methods because it requires no specialised taxonomic knowledge, potentially reduces the time spent on sample processing and enables comprehensive surveys of meiofauna, microfauna and eukaryotic microbes (Gielings et al., 2021). By amplifying standardised genetic markers using universal primers, metabarcoding can capture a broad range of taxa in a single analysis, thereby overcoming many of the biases associated with manual identification. In doing so, metabarcoding offers a means to assess both macro- and microeukaryotic communities, potentially clarifying whether heavy metals exert a consistent influence across different organismal size classes. Numerous metabarcoding studies have explored how environmental variables, including pollution, shape eukaryotic community composition across multiple sites e.g. (Mazurkiewicz et al., 2024, Múrria et al., 2024)**.** However, many rely on limited site coverage (e.g., Dewi et al. (2024) tested only three sites and Kalu et al. (2023) only two), or primarily examine salinity and redox gradients instead of explicit metal contamination (Brannock et al., 2016, Zhao et al., 2020, Chen et al., 2022, Zeng et al., 2023) and some use only one primer set, which complicates efforts to disentangle the full effects of heavy metals (Kalu et al., 2023, Dewi et al., 2024).

Despite its capacity to capture wide-ranging biodiversity, eukaryotic metabarcoding remains susceptible to primer-driven biases that can skew amplification among target taxa (Creer et al., 2016, Reynolds et al., 2022). Unlike prokaryotic primers, which have undergone extensive

refinement, eukaryote-targeted primer sets often under- or over-represent particular lineages, raising concerns about incomplete or distorted ecological snapshots (Bik et al., 2012). Additionally, reference databases were incomplete or inconsistently curated, complicating accurate taxonomic assignments and potentially underestimating rare or poorly described taxa. Several recent studies, including Schoenle et al. (2021) and Li et al. (2024), have underscored how database gaps limit the resolution of metabarcoding data, while Kalu et al. (2023) highlighted the constraints of relying on a single primer set for complex eukaryotic assemblages. Such technical hurdles have limited the ability to confirm whether observed patterns genuinely reflected metal stress or were merely artifacts of primer bias and database deficiencies.

Morphological surveys by (Somerfield et al., 1994a) indicate that nematode communities in Restronguet, Mylor, Pill, St Just and Percuil Rivers differ from each other, but it remains uncertain whether these differences are driven by pollution or variability of other environmental variables. (Millward, 1995) showed that Percuil River nematode communities resemble those in Helford River while differing markedly from those in Restronguet Creek, suggesting no significant pollution impact in Percuil despite elevated copper concentrations. This would be consistent with data on the pollution tolerance of nematode communities, which showed increased tolerance to copper in the Hayle Estuary, Restronguet, St. Just and Cowlands Creeks, whereas copper tolerance of nematodes from Percuil River was similar to those from the uncontaminated Breydon Water (Norfolk) and Kingsbridge estuaries (Millward and Grant, 1995). By contrast, harpacticoid crustacea at the same sites revealed no clear pollution effects (Somerfield et al., 1994a), reinforcing the complexity of disentangling impacts of contamination from other environmental gradients.

Building on these morphological findings, this chapter employs a broader metabarcoding approach covering a wider array of eukaryotic groups with multiple primer sets and a substantially larger number of sites than most previous investigations. It utilises whole sediment samples without flotation or elutriation, targets a well-established pollution gradient and compares metabarcoding outcomes with established morphological and biochemical data. By incorporating 12-site and 34-site datasets that together span 2-431 µg/L porewater Cu and 4-1798 µg/g extractable Cu, this study achieves up to a 90-fold difference in porewater copper levels far exceeding the relatively mild gradients common in earlier eukaryotic surveys.

Multiple primer sets (ITS, G18S, TAR, NEM, DM568 and JB3) enable detection of diverse taxa, from microfauna to small macrofauna, providing a rare opportunity to evaluate whether copper effects remain consistent across varied eukaryotic lineages. Furthermore, by directly extracting DNA from whole sediment without time consuming flotation or elutriation, this approach captures both microbial and metazoan eukaryotes in a single workflow, thereby addressing the common limitations of incomplete coverage and smaller sample sizes seen in prior investigations. This combination of a large metal gradient, multiple primers and coverage of a large number of sites aims to resolve whether heavy metals truly drive consistent shifts in eukaryotic communities and the extent to which salinity, sediment composition and other natural drivers overshadow pollution effects.

6.2 Methods

6.2.1 Sampling and DNA extraction

Three replicate samples were obtained from intertidal areas within estuaries in Southwest England, across 12 locations as detailed in (Udochi, 2020) and 34 locations as outlined in **Chapter 2**, along with two sites from Breydon Water, Norfolk. Further details of sampling procedures, the specifics of sample collection and the exact geographical coordinates were also provided in **Chapter 2**. The DNA extraction followed the protocols described in **Chapter 2**, using DNeasy® PowerSoil® Pro kit (Qiagen, Hilden, Germany) extraction kits, with methodological adjustments specified therein.

6.2.2 PCR and sequencing

The amplicon-based approach targeted various regions of eukaryotic genes using a diverse selection of primers, each identified by its forward primer name or an abbreviated version for simplicity. The ITS primers (ITS1f12/ITS2) targeted the ITS region, while primers G18S (G18S4/22R), TAR (TAReuk454FWD1/TAReukREV3) and NEM (NEM/18Sr2b) were designed for the 18S region. The DM568 primers (DM568F/RM3R) focused on the 28S region and the JB3 primers (JB3adjusted/JB5) amplified the COI gene. Each primer set included a forward and reverse primer, both modified to include an Illumina adaptor, an 8-base barcode, a 0-7 base length heterogeneity spacer and the gene-specific primer sequence (Fadrosh et al.,

2014). These primers were synthesised by Sigma Aldrich Company limited, with more detailed specifications of the primer sequences provided in **Chapter 3**.

The polymerase chain reaction (PCR) was carried out in a 20 µl reaction volume comprising 10 µl of Phusion™ Flash master mix (Thermo Scientific, UK), 1 µl of forward primer, 1 µl of reverse primer (final primer concentration: 100 µM), 7 µl of ultrapure sterile water (MilliQ water) and 1 µl of DNA template. This template DNA was added at concentrations ranging from nearly zero up to 9.3 ng/µl, depending on the individual sample, as described in **Chapter 2**.

The amplification of the DNA templates was carried out in a Veriti™ HID 96-Well Thermal Cycler, 0.2mL system (Applied Biosystems, UK) in either 0.2mL x 96 well plates or 0.2 ml PCR tubes, depending upon the number of samples be amplified. The amplification consisted of an initial denaturation at 98°C for 10 minutes, followed by 35 cycles of denaturation at 98°C for 30 seconds; annealing temperatures were varied depending on the extraction methods and ranged between 53 - 70°C for 30 seconds and extension at 72°C for 30 seconds. A final extension step was performed at 72°C for 5 minutes, after which samples were held at 4°C. All cycling conditions, including the specific annealing temperatures and cycle numbers, were optimised to ensure the best possible amplification yield for each primer set, as further explained in **Chapter 3**. For the COI JB3 primer pair, a two-step PCR protocol was implemented (conditions detailed in **Chapter 3**). Primers were supplied by SigmaAldrich Company limited.

The PCR products were verified by gel electrophoresis, as described in **Chapter 2**. After visual inspection, bands of the approximately expected size were confirmed for each primer set: ITS (variable length), G18S (400 bp), TAR (420 bp), NEM (500 bp), DM568 (500 bp) and JB3 (370 bp), plus an additional ~145 bp from Illumina adaptors, barcodes and spacers. Quantification was performed using the StepOne™ Real-Time PCR system (**Chapter 2**) and any reactions yielding no visible band on the gel or a concentration of 0 ng/µL were excluded. The samples that exhibited amplicons were purified using the Aline Biosciences PCRClean DX kit (Aline Biosciences, Woburn, USA), in accordance with the manufacturer's instructions except for adjustments in the ratio of bead suspension to PCR product. The ratio employed for purification was 1.8:1 in the first run, which included only the initial ITS batch and 1:1 in

subsequent runs. The second ITS batch and the first batches of the other five primers were included in the second run, while the final batches of all primers were processed in the third run, optimizing the removal of primer dimers and adapter contamination. These specific ratio modifications were drawn from the experience gained in similar clean-up protocols (Quail et al., 2009) and were also described in **Chapter 2**. Quantification of the purified PCR products proceeded as in **Chapter 2**, allowing precise normalisation and ensuring equivalent concentrations of amplicons in the final pooled library. Each primer was processed in two separate batches corresponding to two datasets: the first batch represented the 12-site dataset and the second batch represented the 34-site dataset.

Sequencing of the purified eukaryotic amplicons was undertaken at the Earlham Institute in Norwich, UK, using a pre-made library pool on a single lane of Illumina MiSeq with 300 bp paired-end sequencing in June 2021. Subsequent sequencing runs on the NovaSeq 6000 flow cell with 250 bp paired-end reads were conducted in February 2022 and February 2023, as detailed earlier, with the ITS primer included in the first run and subsequent primers processed in later runs.

The number of reads recovered for each sample ranged from 2 to 1M reads for the ITS primer pair; 121K to 2.8M reads for the G18S primer pair; 998 to 3M reads for the TAR primer pair; 119K to 543K reads for the NEM primer pair; 7.7K to 481K reads for the DM568 primer pair; and 3 to 3M reads for the JB3 primer pair.

Samples with low sequencing read counts (ranging from 2 to 4981 reads) were excluded from the analysis. This removal of low count samples maintained the integrity of the dataset and enhanced the accuracy of subsequent analyses. The excluded samples were (12_ITS_PR1) from the 12-site set; and (ITS_CK3), (ITS_HA1), (ITS_JB1), (ITS_JB2), (ITS_LB1), (ITS_LF1), (ITS_MA1), (ITS_RB1), (ITS_RD1), (ITS_TB1), (ITS_VA1), (TAR_CB2), (TAR_VC2), (JB3_LB1), (JB3_LD2), (JB3_VB1), (JB3_VB2) and (JB3_VC2) from the 34-site set. All thresholds were set below 5000 reads, except for JB3, which was set below 50 due to its lower average read count. The JB3 dataset for the 34-site samples included only two replicates per site.

Bioinformatic analysis was carried out using the LotuS2 pipeline (Özkurt et al., 2022). Amplicon Sequence Variants (ASVs) were generated using the UNOISE algorithm, which models sequencing errors to correct them and recover true biological sequences. Taxonomic assignment for eukaryotes was conducted using Eukaryome version 1.7 (www.eukaryome.org) (Tedersoo et al., 2024), which specifically covered the 18S, ITS and 28S markers. COI assignments employed MIDORI-Longest (www.reference-midori.info) (Leray et al., 2018). Detailed methods were described in **Chapter 2**. Section **6.3.2** included a summary of sequencing reads per amplicon, visualised in **Fig. 6.4**. The distribution of reads across eukaryote phyla differs substantially between primer pairs. Ordination analyses were conducted using both Bray-Curtis and Jaccard similarity indices, as described in detail in **Chapter 2**, to evaluate community patterns based on relative abundance and presence-absence data, respectively.

To provide a comparator from similar environmental conditions without PCR bias, the eukaryote 18S sequences from the marine and freshwater sediment samples in the Karst RNAseq dataset (Karst et al., 2018) were classified using USEARCH local matches to the Eukaryome database version 1.8 (www.eukaryome.org) and the Lowest Common Ancestor (LCA) approach using the LCA utility from the LotuS2 pipeline with default similarity thresholds for each taxonomic level (Yarza et al., 2014, Özkurt et al., 2022).

6.2.3 Linking Primer-Taxon Combinations to BIOENV Results

To explore how environmental variables shape the composition of different taxonomic groups, the BIOENV procedure (Clarke and Ainsworth, 1993) was used across multiple primer sets. Each variable was initially tested individually, followed by the selection of the best overall relationship among the measured variables, as described in **Chapter 2**. Data for nine main taxonomic groups, as detailed in **Section 6.3.1**, were assessed against pore water copper (PWCu), acid-extractable metal copper (AEMCu), fine sediment fraction (<63 µm; LT63), median grain size (D50), salinity (Sal), zinc (Zn), acid-volatile sulphides (AVS), total organic carbon (TOC), dissolved organic carbon (DOC), equilibrium partitioning copper (EqpCu) and pore water organic carbon-normalised copper (PW_OC_Cu). While all parameters were available for the 12-site dataset, only a dataset was used for the 34-site dataset (PWCu, AEMCu, LT63, D50 and Sal).

Table 6.1 Main eukaryotic taxonomic groups detected by the six primer pairs, summarising the environmental drivers, responding taxonomic groups and associated primers for the 12- and 34-site datasets.

Taxonomic Group	Primer Pair
All Eukaryotes	ITS, G18S, TAR, DM568, JB3
Stramenopila	ITS, G18S
Fungi	ITS, G18S
Alveolata	TAR
Annelida	TAR
Arthropoda	G18S, DM568
Nematoda	G18S, TAR, DM568, JB3
Platyhelminthes	TAR
Mollusca	G18S

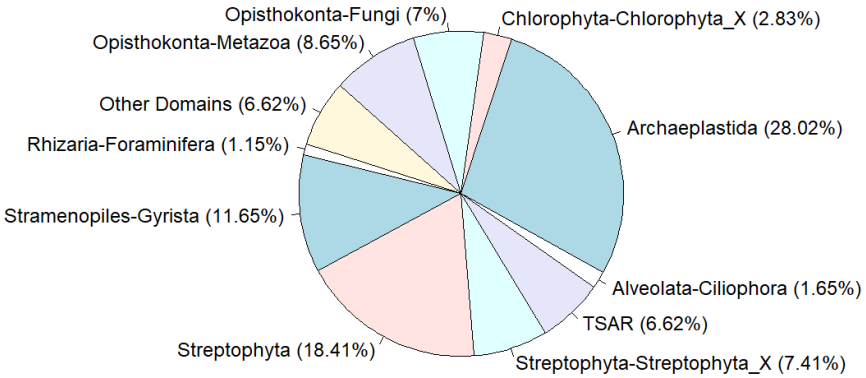
6.3 Results

6.3.1 Taxonomic composition

The taxonomic composition of eukaryotic communities was analysed using multiple primers targeting 18S, ITS and 28S rRNA genes. While each primer set provided valuable insights into community structure, significant biases were observed.

To compare the metabarcoding results against broader ecological patterns, RNAseq data from Karst et al. (2018) and species distributions from the World Register of Marine Species (WoRMS) (www.marinespecies.org) were used to establish expectations for community composition. The RNAseq data, derived from sediment samples, indicated that Archaeplastida account for 28%, followed by Streptophyta at 18%, Stramenopiles-Gyrista at 12%, Metazoa at 9% and Fungi at 7% (**Fig. 6.1**). Within metazoans, the most abundant phyla were Arthropoda (45%), Platyhelminthes (19%), Nematoda (14%) and Annelida (13%), with Mollusca composing a smaller fraction (3%).

A



B

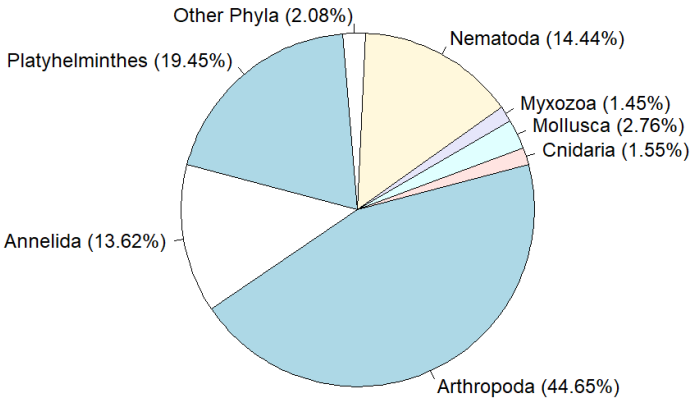
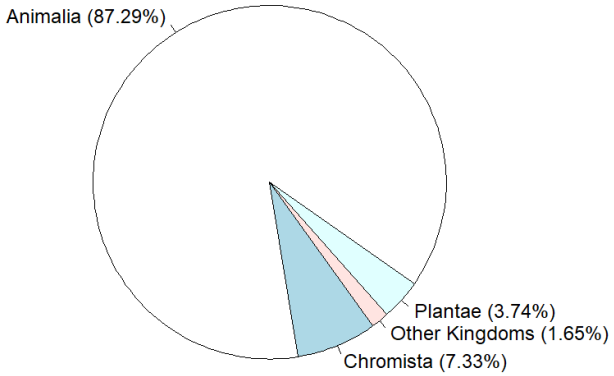


Fig. 6.1 Proportional distribution of the top 10 eukaryote **A)** and metazoan **B)** phyla identified in RNAseq data (Karst et al., 2018). Phyla representing less than 1% combined were grouped under 'Other'. Percentages were calculated based on relative abundance data to illustrate the dominant phyla within the dataset.

In contrast, WoRMS data, which reflect global species richness, showed that Arthropoda (28%) and Mollusca (25%) dominated marine metazoan diversity, followed by Chordata, Annelida and Platyhelminthes (**Fig. 6.2**).

A



B

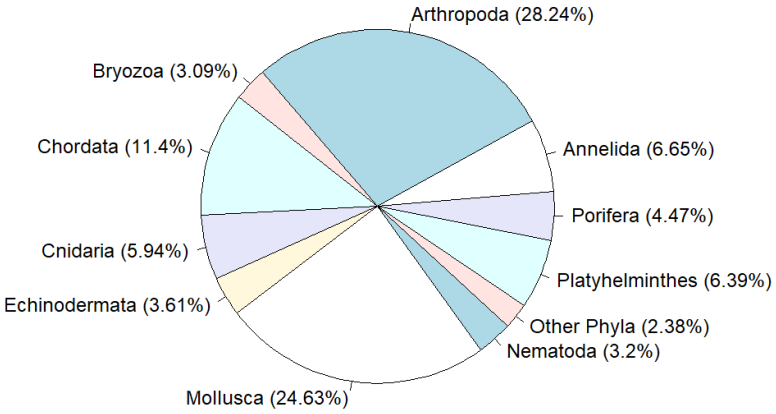


Fig. 6.2 Proportional distribution of the top 10 kingdoms **A)** and phyla **B)** identified by the World Register of Marine Species (WoRMS) (www.marinespecies.org). Phyla with a combined contribution of less than 1% were grouped under 'Other.' Percentages are based on relative abundance data to highlight the dominant kingdoms and phyla within the dataset.

6.3.2 Primer Performance and Observed Biases

This section evaluated how six primer sets influenced the detection of major eukaryotic groups and highlighted the implications of these biases for interpreting community composition. Metabarcoding results revealed a dominance of Metazoa in data generated using G18S, TAR,

NEM and DM568 (**Fig. 6.3 A**). By contrast, the ITS primer set captured a higher proportion of Stramenopila (such as diatoms), although a notable fraction of reads remained unassigned. JB3 (COI-based) yielded broad amplification across eukaryotic lineages but lacked consistent higher-level taxonomic resolution. Specialised primers provided narrower coverage of the most abundant kingdoms and phyla (**Fig. 6.3 A** and **B**), representing about 30% of the reads. DM568 displayed a predominance of Metazoa, especially arthropods, whereas JB3 consistently showed minimal detections across all major groups.

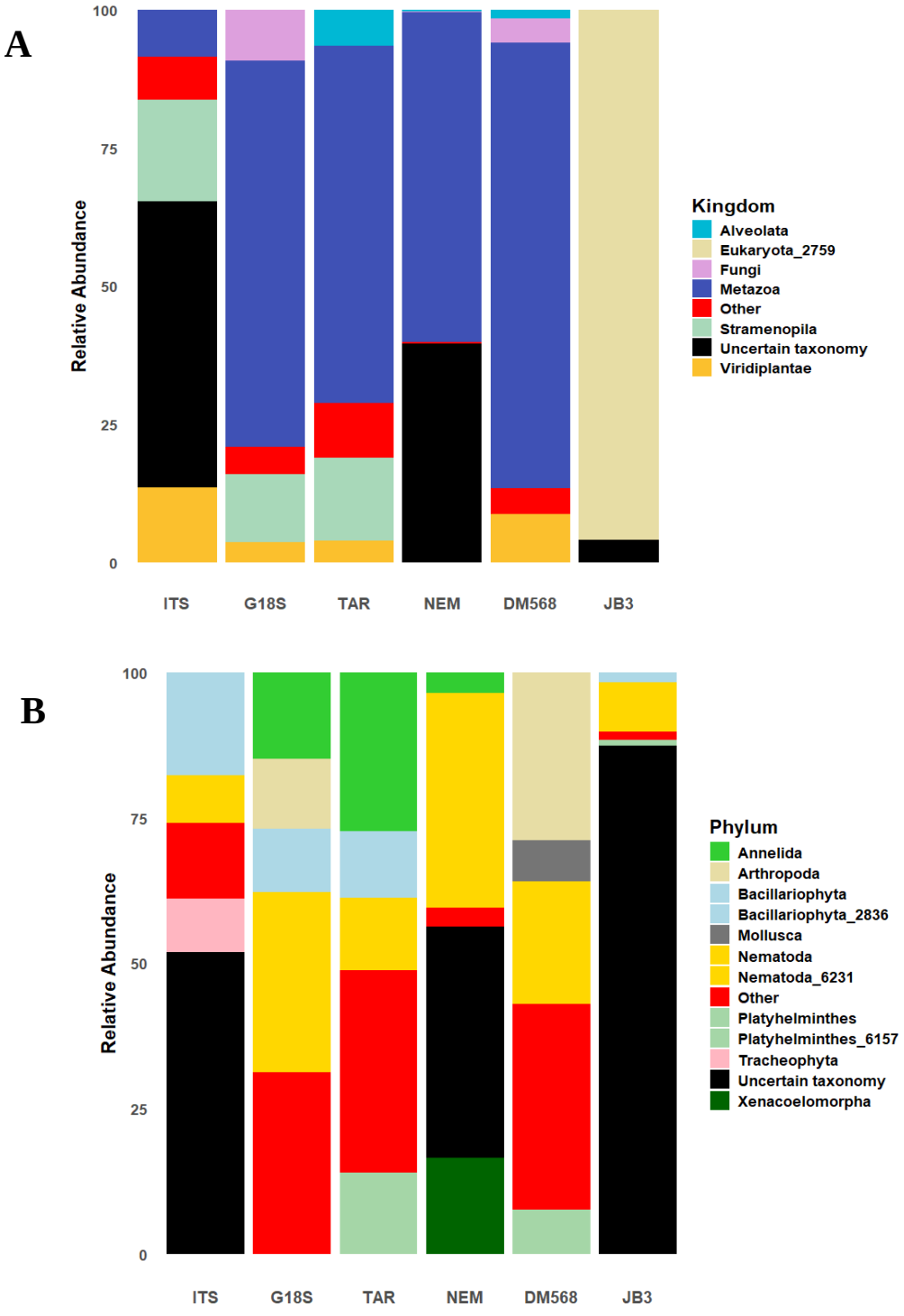


Fig. 6.3 Relative abundance of the five most abundant **A)** kingdoms and **B)** phyla, expressed as a percentage of the total reads. Data were processed using the LotuS2/Eukaryome pipeline (except for COI JB3 using the MIDORI-Longest database). The G18S, ITS and TAR data are derived from all datasets; DM568 from 34-sites and exp; NEM from pilot and 12-sites and JB3 from pilot, 12 and 34 sites. Taxon names with a number following them represent mitochondrial sequences. 'Uncertain taxonomy' refers to assignments that remain unresolved at the kingdom or phylum level.

The heatmap in **Fig. 6.4** illustrated how each primer detected and represented major eukaryotic kingdoms and metazoan phyla. Colour intensities represented the percentage of total reads for each taxon-primer combination, with G18S showed robust amplification of arthropods and fungi, TAR favoured alveolates and ITS captured stramenopiles (e.g., diatoms) but underrepresented metazoans. NEM was highly specific to nematodes, while DM568 detected various metazoan groups at lower overall levels and JB3 consistently showed minimal signals across all groups.

Notably, **Fig. 6.4** showed that G18S and TAR offered broad coverage, detecting diverse taxa such as nematodes, annelids and alveolates. In contrast, ITS captured Stramenopila and fungi with higher specificity but left a fraction of reads unassigned. Highly focused primers like NEM provided finer resolution for nematodes but overlooked other sensitive groups.

The importance of aligning local ASV results with global references was further illustrated by comparing primer-based proportions to data from sources like WoRMS (**Fig. 6.2**), which indicated roughly 25% Arthropoda and 25% Mollusca worldwide.

For arthropods (28% in WoRMS), G18S was closest at about 22%, whereas DM568 showed 15% and 11% (NEM) and only 3% with JB3, but neither TAR nor ITS featured arthropods among their top five phyla. Molluscs appeared at 3% under NEM but were absent from the other top-five lists, compared with the 25% expected globally. Annelids, approximately 7% worldwide, G18S was closest at about 6%, whereas NEM showed 21%. Nematodes account for only 3.2% in WoRMS, yet they were remarkably more prevalent across different primer sets, reaching 8% in ITS, 14% in G18S, 5% in TAR, 25% with NEM, 24% in DM568 and 9% with JB3.

Fig. 6.4 The red colour scale shows the percentage of total reads for each taxon-primer combination, with darker red indicating higher percentages. Numeric labels represent the ASV counts across six primers for the top five eukaryotic kingdoms (in bold) and major metazoan phyla. Taxonomic assignments were determined using the Eukaryome database. In the JB3 dataset, "0" indicates kingdoms or phyla not classified within the MIDORI-Longest taxonomy system. All data were processed using the LotuS2 pipeline.

6.3.3 Evaluation of Similarity Metrics in Eukaryotic Ordinations

High-throughput metabarcoding counts provide, at best, a semi-quantitative picture of community structure because amplification efficiency, rRNA copy number and primer bias can

decouple read numbers from actual biomass or individual counts (Leray & Knowlton 2017; Lamb et al. 2019). To test whether this uncertainty affects the observed patterns, the 18S data for the 12-site dataset were ordinated using both an abundance-weighted metric (Bray-Curtis) and a presence/absence metric (Jaccard), applying NMDS and PCoA (**Figs. 6.5** and **6.6**). Both ordination methods revealed a similar overall gradient, with sites exhibiting the highest porewater Cu concentrations positioned towards one side of the configuration, while lower-Cu sites clustered oppositely. Minor differences were noted between the two metrics: in the Jaccard PCoA (**Fig. 6.6 A**), the Percuil River (PR) sites appeared more detached from the other clean sites, positioned towards the side of the plot where highly contaminated sites also occur. Bray-Curtis ordination (**Fig. 6.6 B**) produced a slightly more cohesive separation along the contamination gradient. Given its capacity to incorporate relative abundance information while still reflecting between-site differences consistent with presence/absence data, Bray-Curtis was retained for the main analyses.

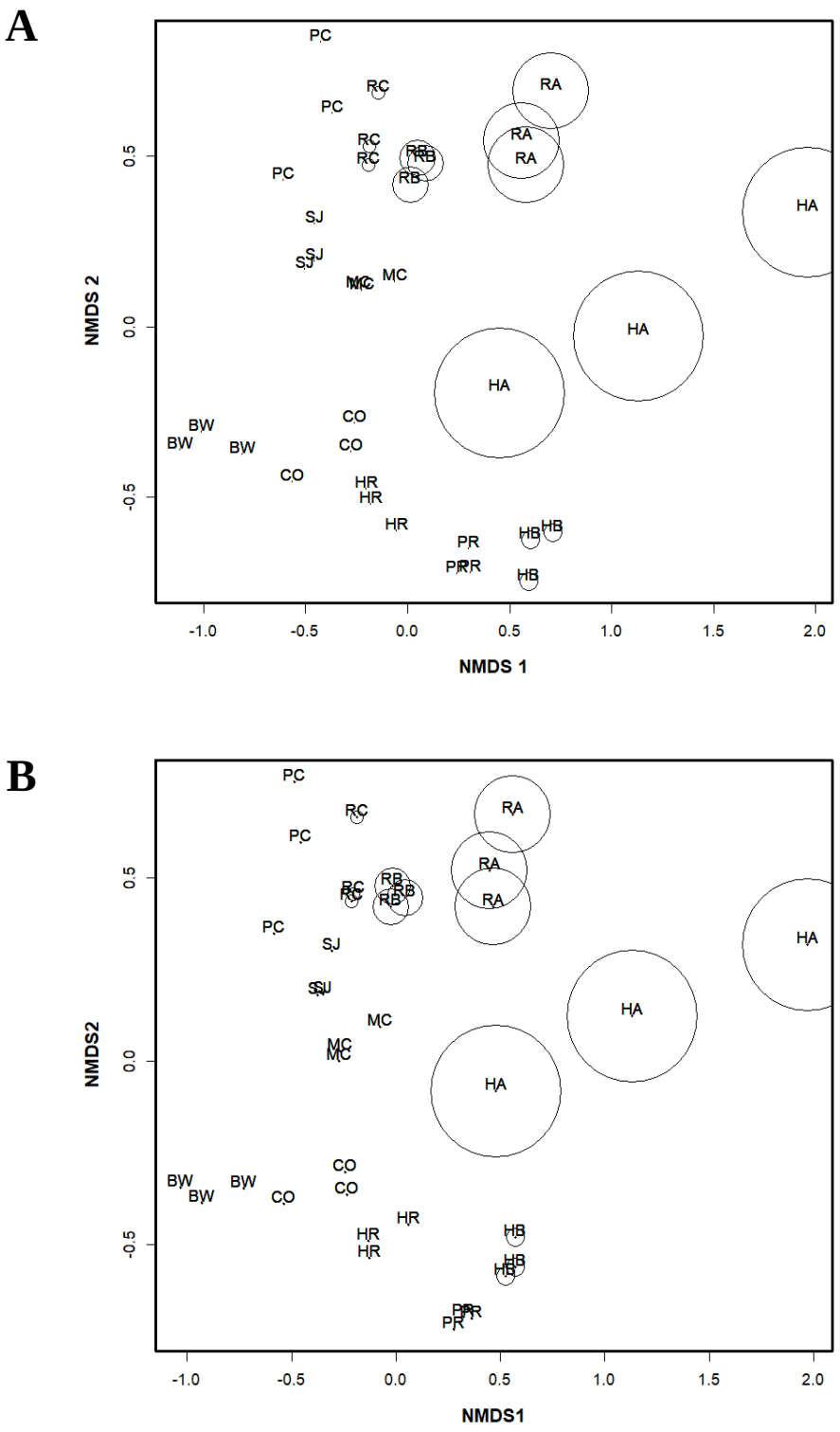


Fig. 6.5 Total Eukaryote NMDS bubble plots from G18S primers across the 12-site dataset, based on untransformed data using **A)** Jaccard and **B)** Bray-Curtis similarity metrics. Bubble sizes indicate pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

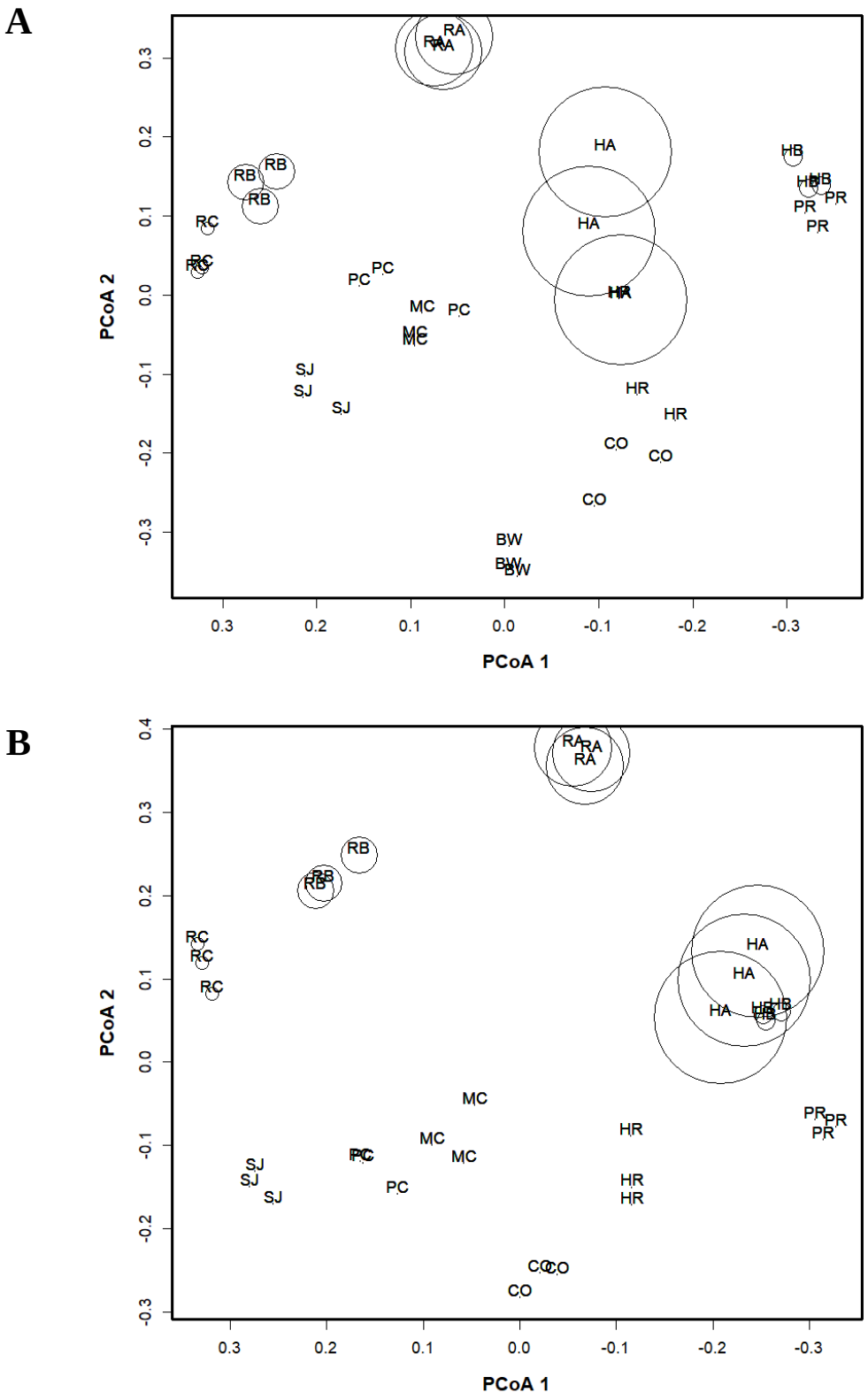


Fig. 6.6 Total Eukaryote PCoA bubble plots from G18S primers across the 12-site dataset, based on untransformed data using **A)** Jaccard and **B)** Bray-Curtis similarity metrics. Bubble sizes indicate pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

6.3.4 Comparison of Clustering Patterns Between NMDS and PCoA Analyses

Visual inspection of ordination plots (**Fig. 6.7**) indicated that PCoA produced more distinct clustering patterns than NMDS for both porewater copper (PWCu; **Fig. 6.7 A** and **B**) and acid-extractable copper (AEM; **Fig. 6.8 A** and **B**). In the 12-site dataset, PWCu was selected for ordination, while AEM was used for the 34-site dataset. NMDS plots showed similar spatial arrangements but with reduced separation between sites. Differences in the number of sites across datasets contributed to varying degrees of overlap in ordination space.

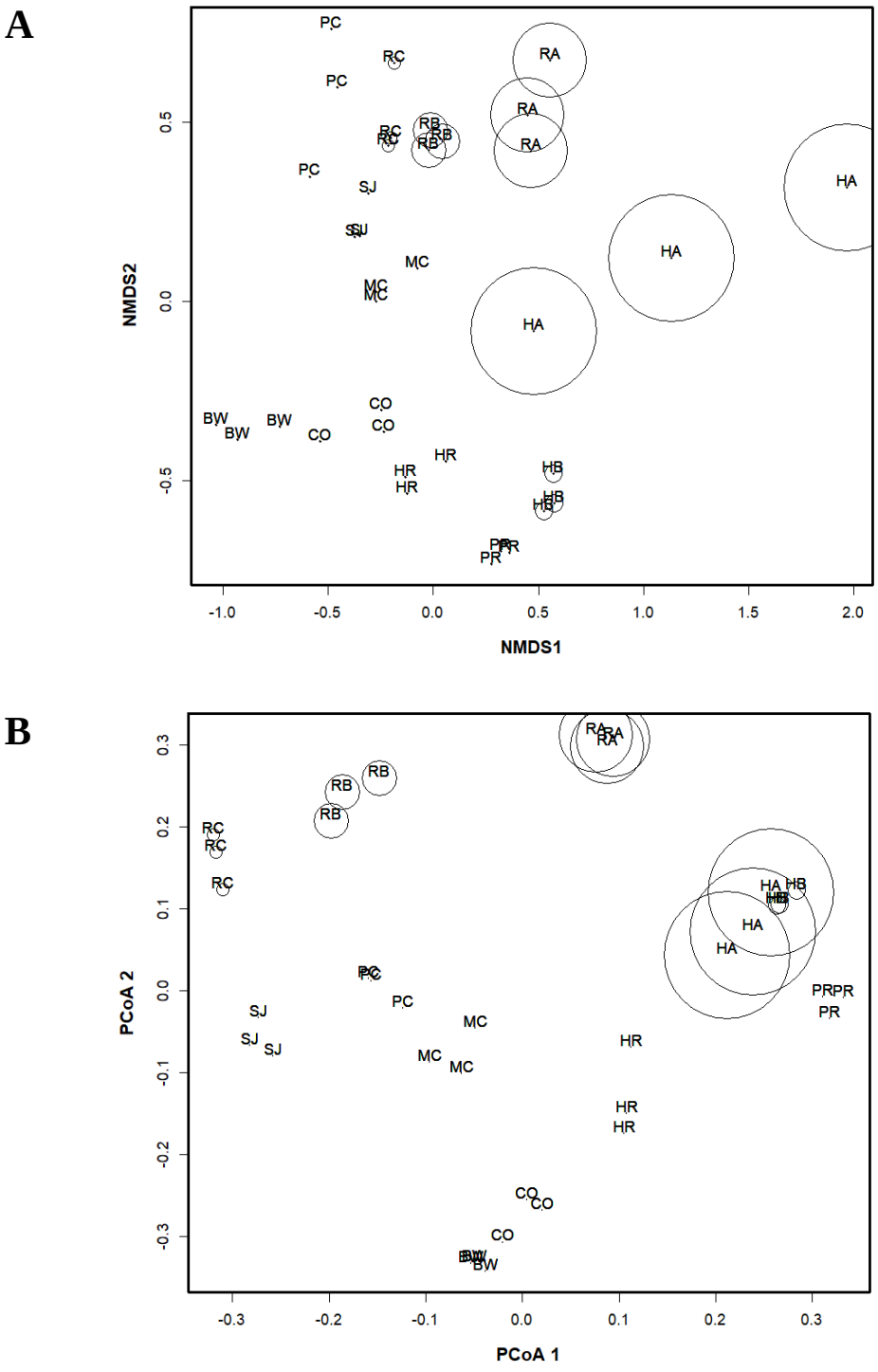


Fig. 6.7 Total Eukaryote bubble plots from G18S primers across 12 and 34 sites, based on square root transformed data using the G18S primer and Bray-Curtis similarity. The analysis includes **A)** NMDS and **B)** PCoA ordinations, with no ASV pruning. Bubble sizes indicate the pore water copper (Cu) concentrations at each site. **Plots A** and **B** highlight PWCu. Site codes as in **Table 2.1**.

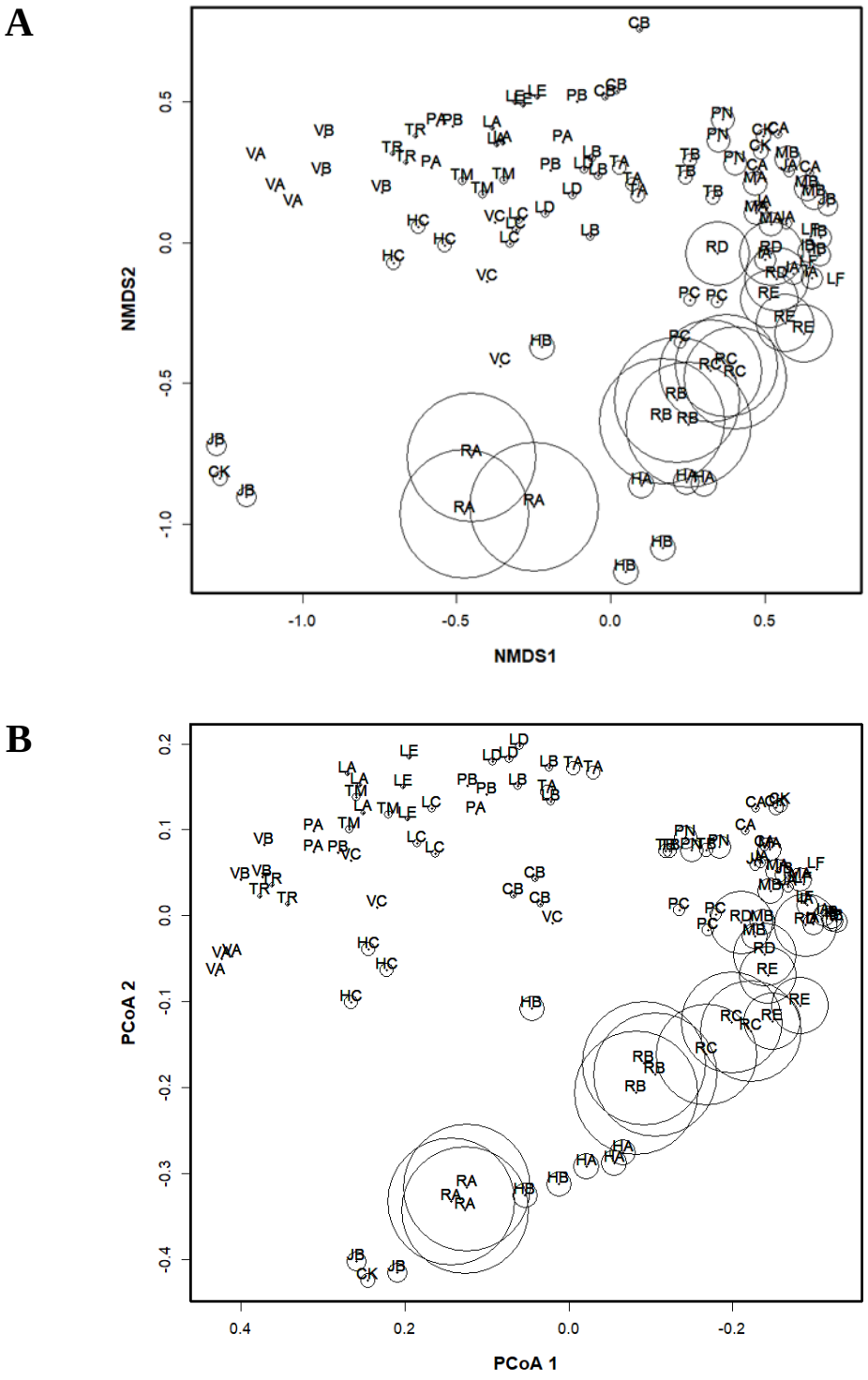


Fig. 6.8 Total Eukaryote bubble plots from G18S primers across 12 and 34 sites, based on square root transformed data using the G18S primer and Bray-Curtis similarity. **A)** NMDS and **B)** PCoA ordinations focus on AEM, with no ASV pruning. Bubble sizes indicate the pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

6.3.5 Effect of Rare ASV removal on analysis

Following NMDS and PCoA outputs that highlighted overall site differences (and suggested copper-driven gradients), this section evaluates whether retaining rare ASVs improves site-level resolution, datasets with and without pruning thresholds (0.1 % and 1 %) were compared. Pruning low-abundance ASVs reduced the number of retained variants but did not alter the overall grouping patterns observed in NMDS and PCoA ordinations. Subtle compositional differences between heavily impacted estuaries and less contaminated sites appeared more distinct in unpruned datasets, as shown in PCoA plots (**Figs. 6.9, 6.10** and **6.11**).

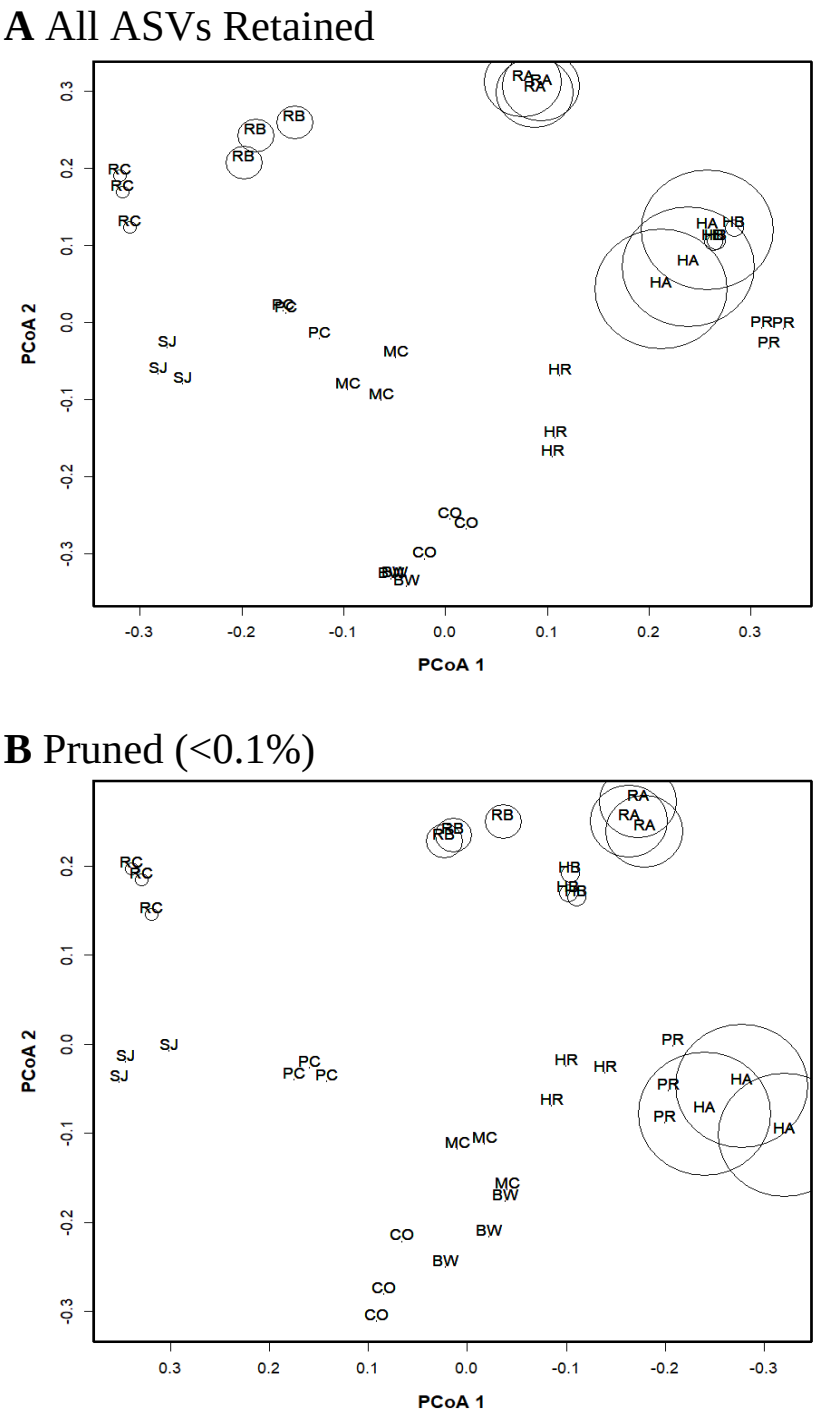
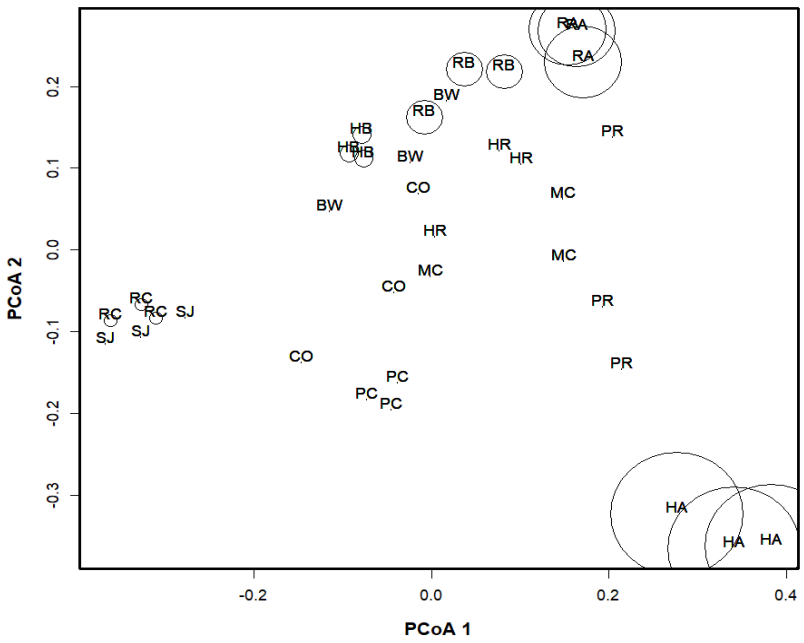


Fig. 6.9 PCoA bubble plots for total eukaryotes based on square root transformed data using the G18S primer and Bray-Curtis similarity. Comparisons are shown for 12 sites with pore water copper (PWCu). **A)** shows data with no ASV pruning and **B)** with 0.1% ASV pruning. Bubble sizes indicate the pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

A Pruned (<1%)



B All ASVs Retained

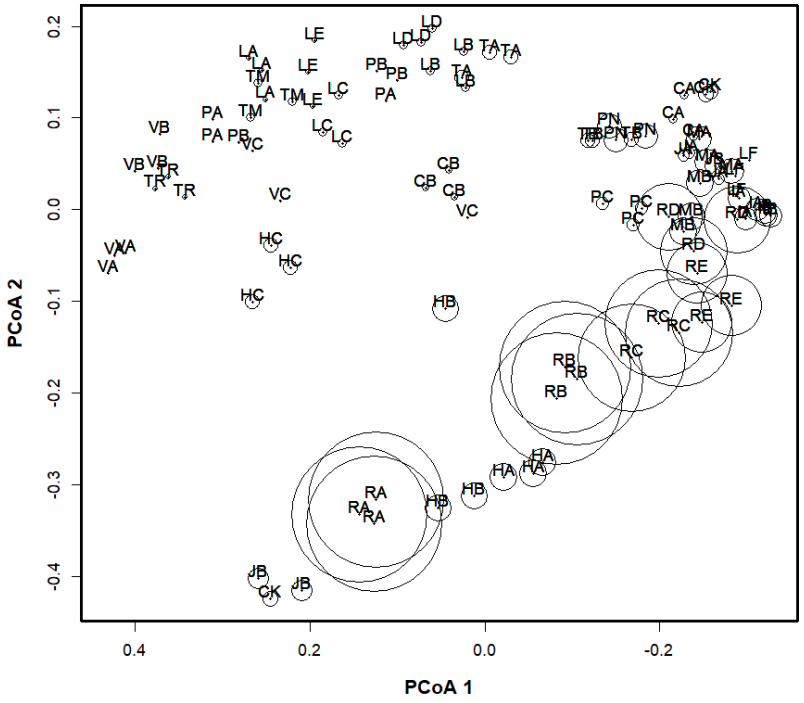
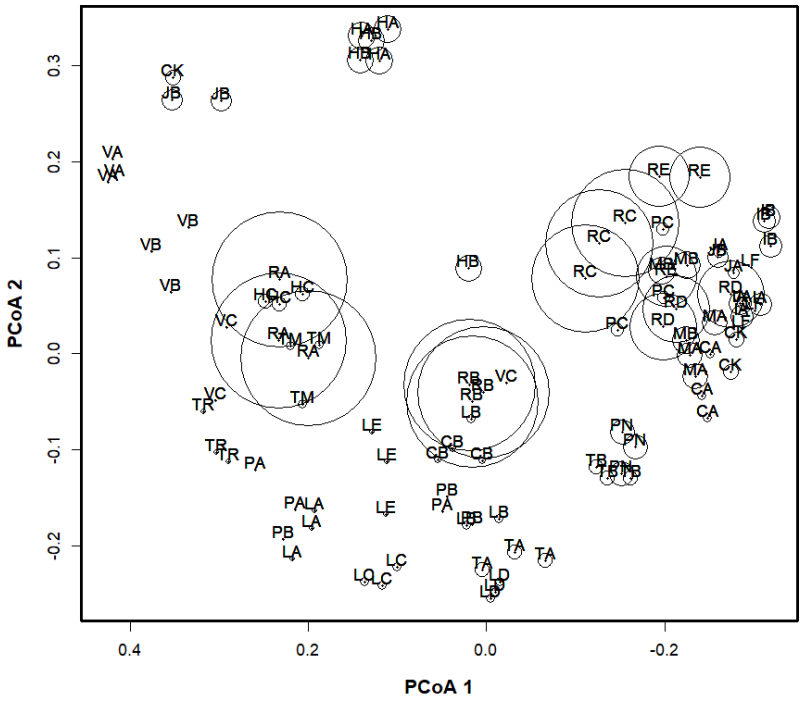


Fig. 6.10 PCoA bubble plots for total eukaryotes using the G18S primer and Bray-Curtis similarity. **A)** shows the 12-site dataset with 1% ASV pruning and **B)** the 34-site dataset with no pruning, focusing on acid-extractable metals (AEM). Bubble sizes indicate the pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

A Pruned (<0.1%)



B Pruned (<1%)

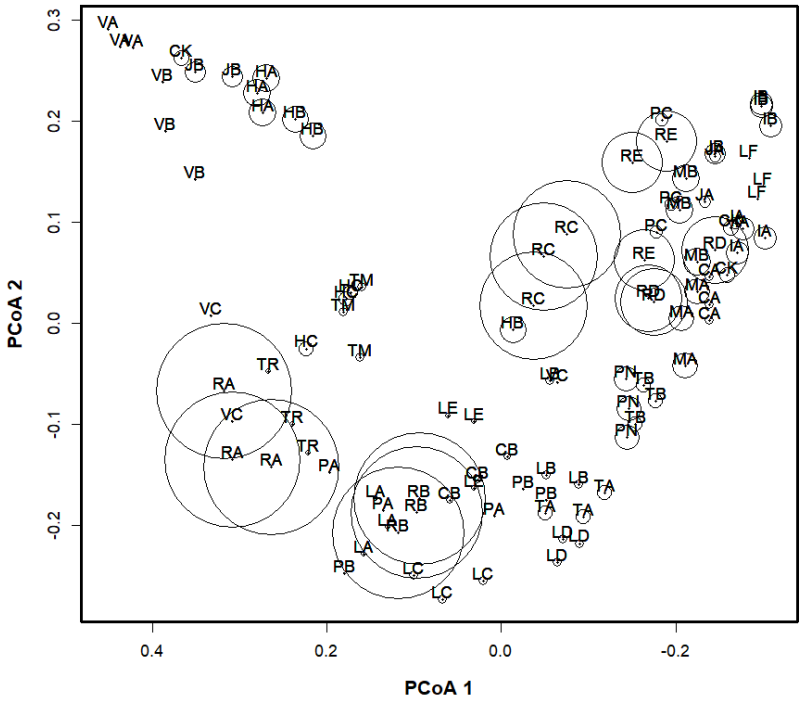


Fig. 6.11 PCoA bubble plots for total eukaryotes using the G18S primer and Bray-Curtis similarity. **A**) shows the 34-site dataset with 0.1% ASV pruning and **B**) with 1% ASV pruning, both focusing on acid-extractable metals (AEM). Bubble sizes indicate the pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

ANOSIM tests were then performed using site as the grouping factor (**Fig. 6.12)**. Including rare ASVs slightly increased within-site variability but preserved more site-level distinctions. ANOSIM R-values remained consistent across pruned and unpruned datasets in both the 12-site and 34-site analyses, indicating that replicate agreement and the ability to detect site differences were robust regardless of pruning.

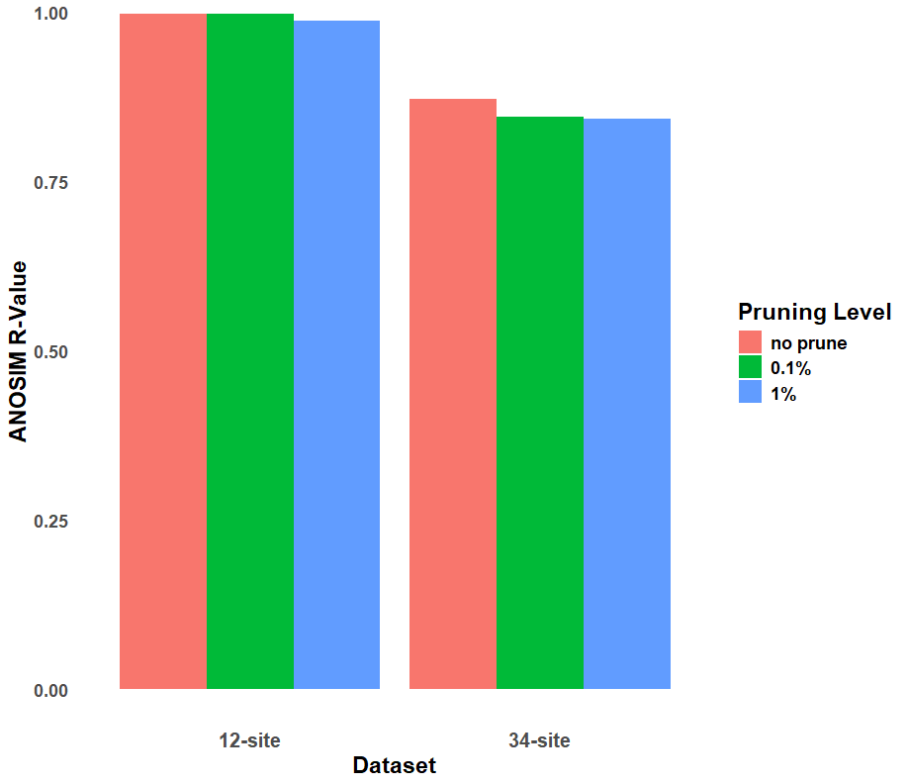


Fig. 6.12 ANOSIM R-values for 12-site and 34-site datasets generated using the G18S primer under three pruning thresholds (no pruning, 0.1% and 1%). The results illustrate the degree of dissimilarity between sites, with R statistics quantifying the strength of clustering by site. R-values > 0.75 indicate very strong differences between sites and all results are statistically significant ($p \leq 0.0001$).

6.3.6 Replication and Site Comparisons

Building on the ANOSIM results presented in **Fig. 6.12**, this section examined how consistently different primers detected community-level variability among sites, while also considering whether replicate variability masks or clarifies site-level patterns. To accomplish this, a combined approach utilised ordination visualisations.

Site-level patterns in relation to copper concentrations were visualised using PCoA bubble plots in **Fig. 6.13** and **Fig. 6.14 A**. These ordinations provided an overview of spatial clustering across sites. To quantify clustering consistency, ANOSIM R-values were calculated and compared across six primer datasets (ITS, G18S, TAR, NEM, DM568, JB3), as shown in **Fig. 6.15**. High R-values were observed for G18S, TAR and DM568 in the 34-site dataset. R-values were generally higher in the 12-site dataset. NEM produced a lower R-value and a non-significant result ($p = 0.1$) in the 12-site dataset. For ITS, G18S, TAR, DM568 and JB3, R-values exceeded 0.75 and were statistically significant ($p \leq 0.0001$). Replicate clustering patterns indicated that small-scale environmental variability did not prevent detection of broader site-level patterns. In the 12-site dataset, G18S ordinations including and excluding Breydon Water samples (**Figs. 6.14 A** and **B**) showed that these samples often clustered closer to Southwest sites compared to bacterial or archaeal patterns.

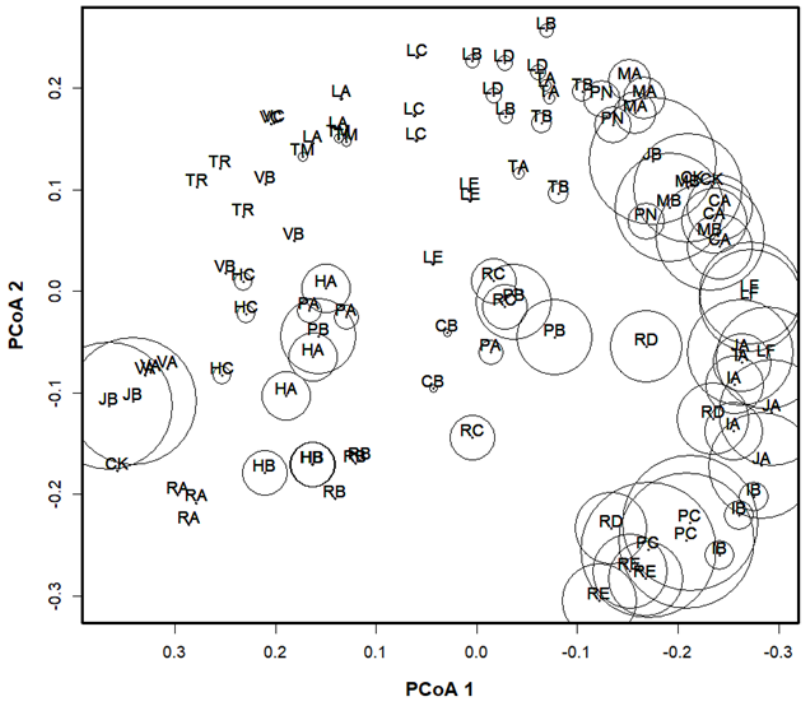


Fig. 6.13 Total Eukaryote bubble plot based on square root transformed data and Bray-Curtis similarity showing TAR primers across 34 sites in relation to salinity. The analysis employs PCoA and NMDS ordinations with no ASV pruning. Bubble sizes represent pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

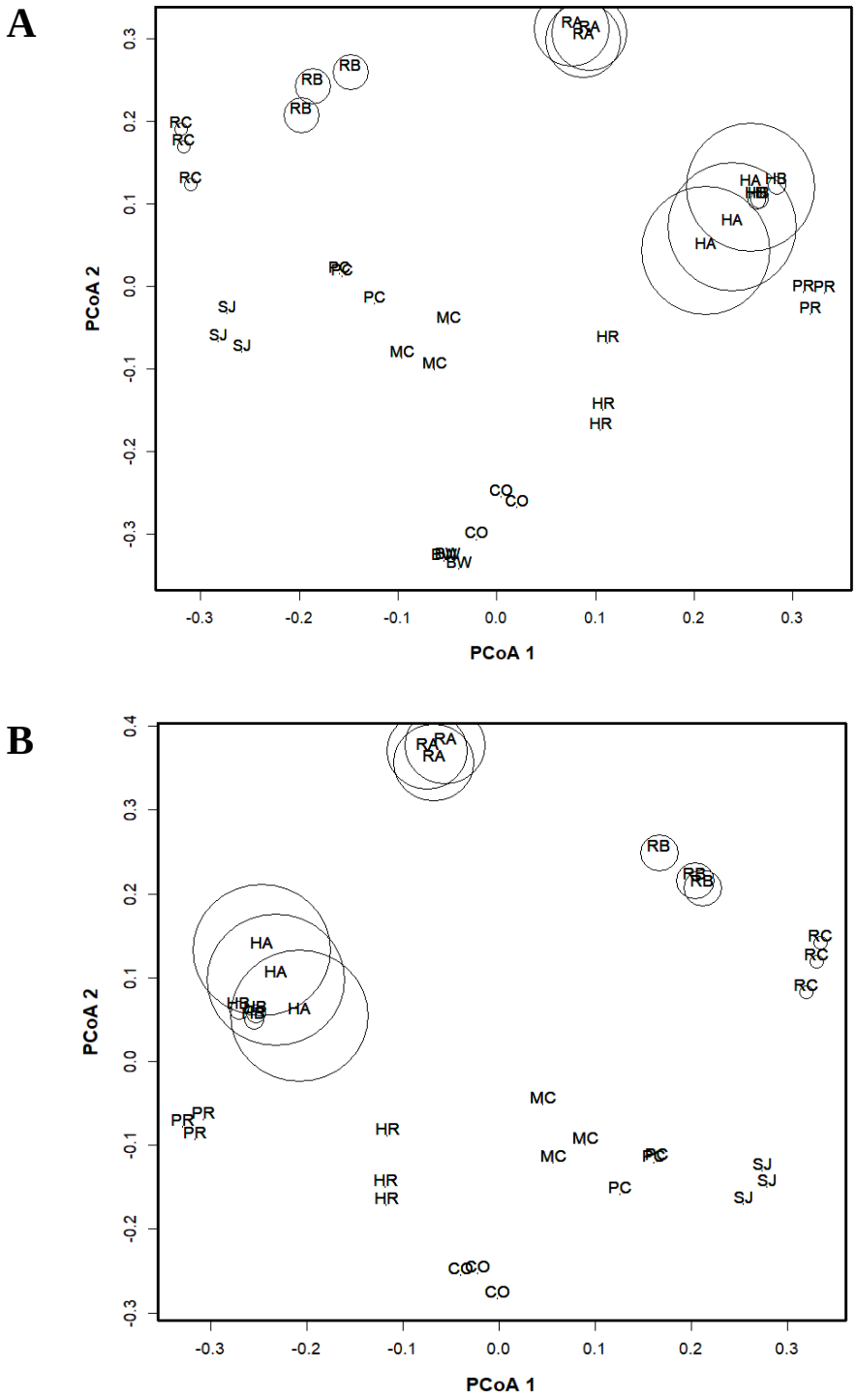


Fig. 6.14 Total Eukaryote bubble plots based on square root transformed data and Bray-Curtis similarity using G18S primers for the 12-site dataset with pore water copper (PWCu). **A)** shows plots including Breydon Water (BW) samples, while **B)** shows plots excluding BW samples. The analysis employs PCoA and NMDS ordinations with no ASV pruning. Bubble sizes represent pore water copper (Cu) concentrations at each site. Site codes as in **Table 2.1**.

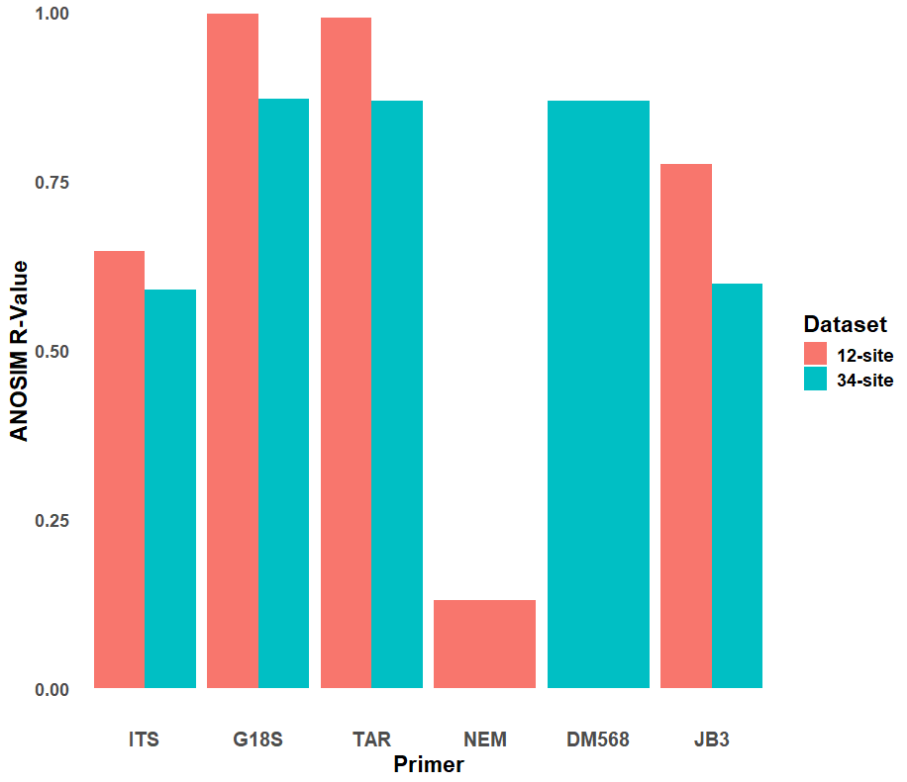


Fig. 6.15 ANOSIM R-values for six primer datasets (ITS, G18S, TAR, NEM, DM568 and JB3) comparing 12-site and 34-site datasets, with R values > 0.75 indicating very strong differences between groups and statistical significance confirmed for most primers ($p \leq 0.0001$), except NEM ($p = 0.1008$).

6.3.7 Eukaryotic Responses to Environmental Drivers

Does copper correlate strongly with community composition, or did other environmental factors play a more prominent role? An initial overview of BIOENV results (**Table 6.2**) indicated that pore-water copper (PWCu), the fraction of sediment finer than 63 µm (LT63) and salinity (Sal) emerge most frequently in top predictor sets, doing so 16, 11 and 15 times, respectively. Median grain size (D50) and equilibrium partitioning copper (EqpCu) showed fewer significant appearances and, in the case of EqpCu, no individual significance.

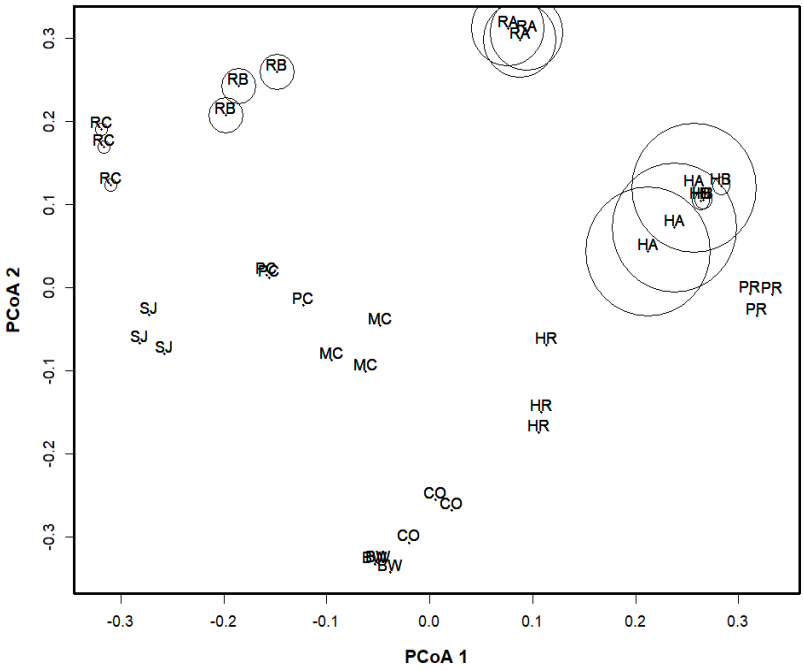
What does a closer look reveal about copper and other variables for each dataset and taxonomic group? For All Eukaryotes, the best environmental combination yields a relatively high correlation ($r = 0.61$) under G18S-12, whereas JB3-34 was weaker at 0.34. In preliminary analyses, a wide range of variables was tested; only those achieving significance ($p \leq 0.05$) and showed consistent associations were selected for final visualisation.

Table 6.2 Summary of environmental drivers, responding taxonomic groups and associated primers for the 12- and 34-site datasets. The table shows Spearman's rank correlation (Rho) values from BIOENV analysis, with individual variables in columns and the best combined r-values in the far-right column. Shaded cells highlight significance: **Gray** for significant and **pink** for non-significant variables. All significant r-values, including combined and individual, are in bold.

Taxonomic group	Primer pair-dataset	PWCu	AEMCu	LT63	D50	Sal	Zn	AVS	TOC	DOC	EqpCu	PW_OC_Cu	Combined overall
All Eukaryotes	ITS-12	-0.19	-0.16	0.04	0.17	0.03	-0.13	**0.38**	**0.31**	-0.05	-0.25	-0.17	**0.55**
	ITS-34	**0.19**	0.12	**0.20**	**0.24**	**0.26**							**0.46**
	G18S-12	**0.41**	0.08	0.01	**0.49**	0.06	0.15	-0.18	0.24	**0.35**	0.05	0.28	**0.61**
	G18S-34	0.17	**0.26**	**0.15**	**0.17**	**0.37**							**0.45**
	TAR-12	0.21	0.03	0.13	0.28	-0.06	0.04	0.01	0.22	0.10	-0.08	0.22	0.51
	TAR-34	0.19	0.15	**0.22**	**0.24**	**0.34**							**0.52**
	DM568-34	0.21	0.16	**0.22**	**0.28**	**0.31**							**0.52**
	JB3-12	-0.17	-0.09	-0.07	-0.08	0.03	-0.12	0.36	-0.02	-0.03	-0.19	-0.06	0.40
	JB3-34	0.15	0.12	0.07	0.09	**0.21**							0.34
Stramenopila	ITS-12	-0.05	-0.14	0.13	0.05	0.00	-0.14	**0.37**	0.26	0.08	-0.15	-0.02	0.57
	ITS-34	-0.04	0.01	**0.12**	0.11	**0.23**							**0.26**
	G18S-12	0.12	-0.06	0.15	0.26	-0.12	-0.06	0.06	-0.02	0.10	-0.14	0.10	0.37
	G18S-34	0.13	0.14	**0.14**	**0.18**	**0.43**							**0.51**
Fungi	ITS-12	0.07	0.04	0.11	**0.25**	0.04	0.00	0.05	0.04	-0.09	-0.01	-0.17	0.25
	ITS-34	**0.13**	**0.12**	-0.03	-0.01	-0.03							0.15
	G18S-12	0.15	0.11	0.09	0.29	0.18	-0.12	0.13	0.21	0.07	0.05	-0.02	0.39
	G18S-34	0.18	0.13	**0.19**	**0.23**	**0.23**							**0.40**
Alveolata	TAR-12	**0.42**	0.04	0.11	**0.43**	-0.15	0.13	-0.02	0.21	0.31	-0.04	0.34	0.60
	TAR-34	0.14	0.04	**0.27**	**0.29**	**0.23**							**0.49**
Annelida	TAR-12	**0.47**	-0.03	0.11	0.14	0.09	-0.05	-0.04	-0.14	0.31	0.05	0.24	0.54
	TAR-34	0.19	0.15	0.04	0.00	0.15							**0.33**
Arthropoda	G18S-12	-0.21	-0.06	-0.05	0.17	0.06	-0.20	0.03	-0.07	-0.11	-0.24	-0.03	0.21
	G18S-34	0.05	**0.22**	**0.11**	0.10	**0.27**							**0.28**
	DM568-34	0.08	0.06	**0.12**	**0.19**	**0.27**							**0.39**
Nematoda	G18S-12	**0.49**	0.12	0.02	**0.46**	-0.09	-0.11	-0.11	**0.34**	**0.41**	0.11	**0.43**	**0.69**
	G18S-34	0.17	**0.18**	**0.14**	**0.20**	**0.37**							**0.49**
	TAR-12	**0.48**	0.22	0.08	**0.31**	-0.07	**0.31**	-0.15	0.07	0.25	0.25	**0.48**	**0.59**
	TAR-34	**0.29**	**0.20**	**0.22**	**0.24**	**0.25**							**0.50**
	DM568-34	**0.34**	**0.23**	**0.22**	**0.29**	**0.25**							**0.55**
	JB3-12	-0.17	-0.09	-0.07	-0.08	0.03	-0.12	0.36	-0.02	-0.03	-0.19	-0.06	0.40
	JB3-34	0.01	0.05	0.00	0.00	**0.20**							0.01
Platyhelminthes	TAR-12	0.21	0.00	-0.14	0.23	-0.11	0.08	-0.11	0.25	0.04	-0.13	0.29	0.56
	TAR-34	0.13	0.00	**0.18**	**0.21**	**0.30**							**0.46**
Mollusca	G18S-12	**0.47**	-0.11	0.03	**0.75**	-0.20	0.27	-0.20	0.12	**0.46**	0.00	0.36	0.75
	G18S-34	0.00	-0.08	0.06	0.08	0.12							0.22
Total Individually Significant		10	6	14	18	16	1	1	3	2	0	2	
Total in Overall Function		16	5	11	0	15	3	1	3	2	0	2	

Does the 12-site or 34-site scale matter? Within the All Eukaryotes group, G18S data for the 12-site set correlate with both PWCu (0.41) (**Fig. 6.16 A**) and D50 (0.49). ITS, however, highlighted AVS (0.38) and TOC (0.31) (**Fig. 6.16 B**). In contrast, the 34-site sets exhibited recurring importance of Sal, appearing in ITS (**Fig. 6.17 A**), TAR (**Fig. 6.17 B**), DM568 (**Fig. 6.18 A**) and JB3 (**Fig. 6.18 B**). AEMCu (G18S4; **Fig. 6.19 A**) also aligned with 34-site data, while LT63 (TAR; **Fig. 6.19 B**) marked the prevalence of finer sediments.

A All eukaryotes- PWCu (G18S4)



B All eukaryotes - TOC (ITS)

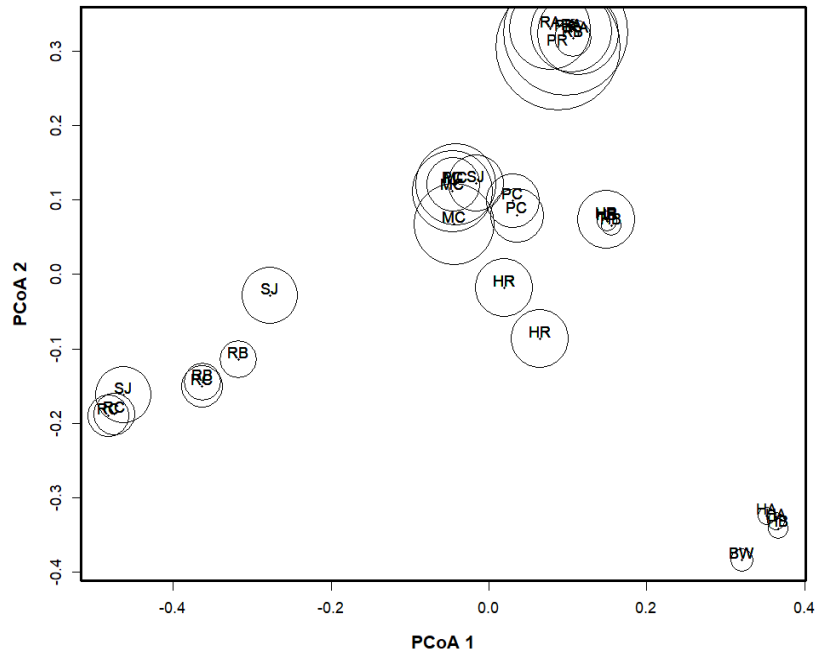
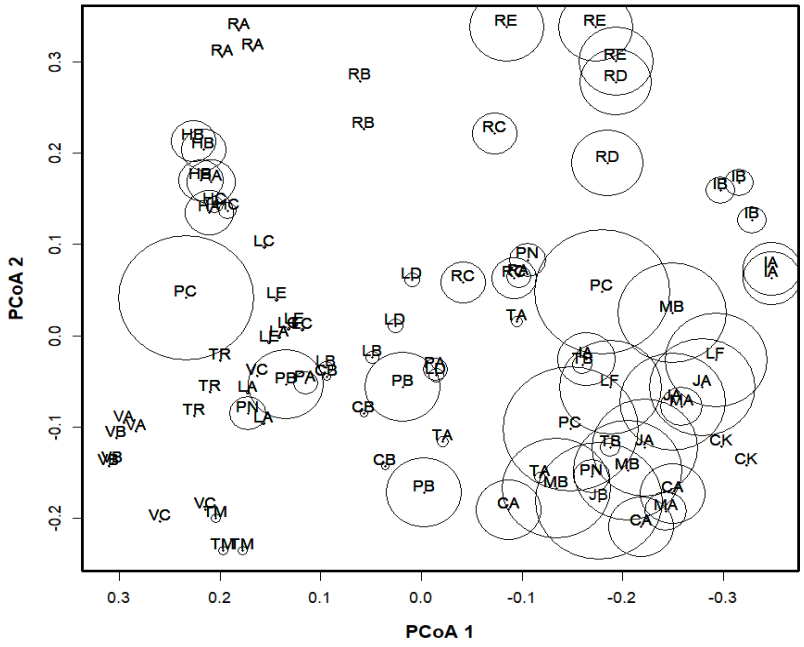


Fig. 6.16 Bubble plots illustrating the relationships between total eukaryotic communities across 12 sites to significant environmental variables identified in the BIOENV analysis: **A)** PWCu and G18S4 data and **B)** TOC and ITS data. The plots are based on PCoA ordination using Bray-Curtis dissimilarity, with no ASV pruning applied. Bubble sizes represent the intensity of the respective environmental variable at each site. Site codes as in **Table 2.1**.

A Salinity (ITS)



B Salinity (TAR)

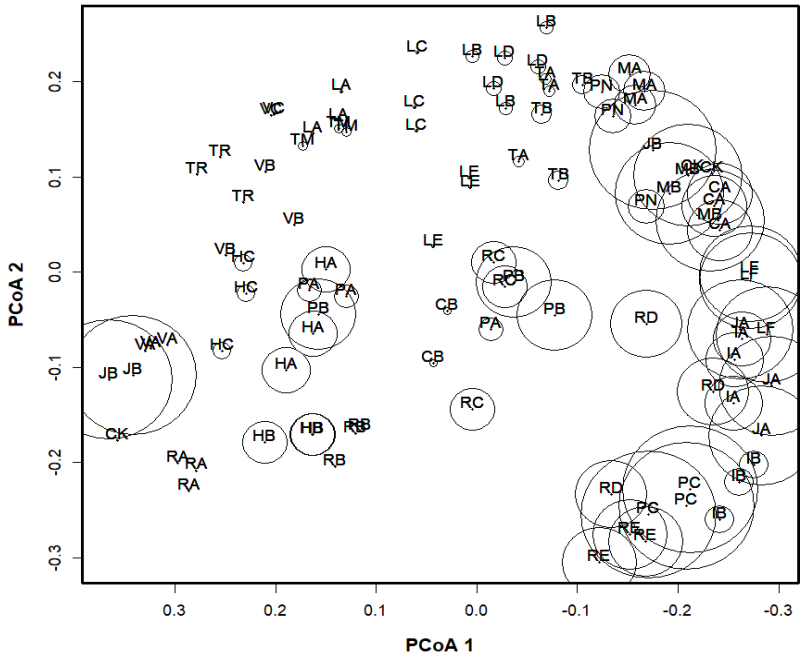
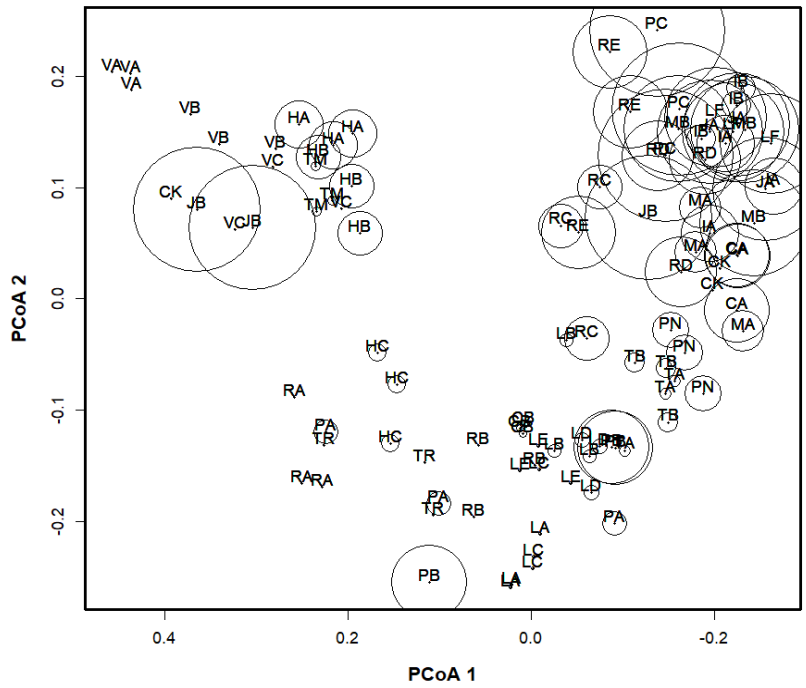


Fig. 6.17 Bubble plots illustrating the responses of total eukaryotic communities across 34 sites to significant environmental variables identified in the BIOENV analysis. **A)** shows salinity with ITS data, while **B)** shows salinity with TAR data. The plots are based on PCoA ordination using Bray-Curtis dissimilarity, with no ASV pruning applied. Bubble sizes represent the intensity of the respective environmental variable at each site. Site codes as in **Table 2.1**.

A Salinity (DM568)



B Salinity (JB3)

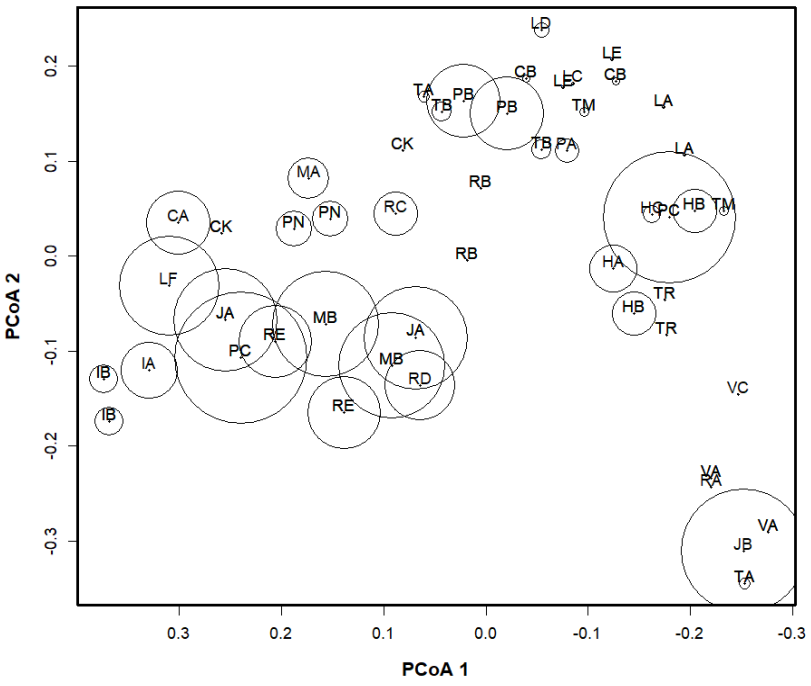
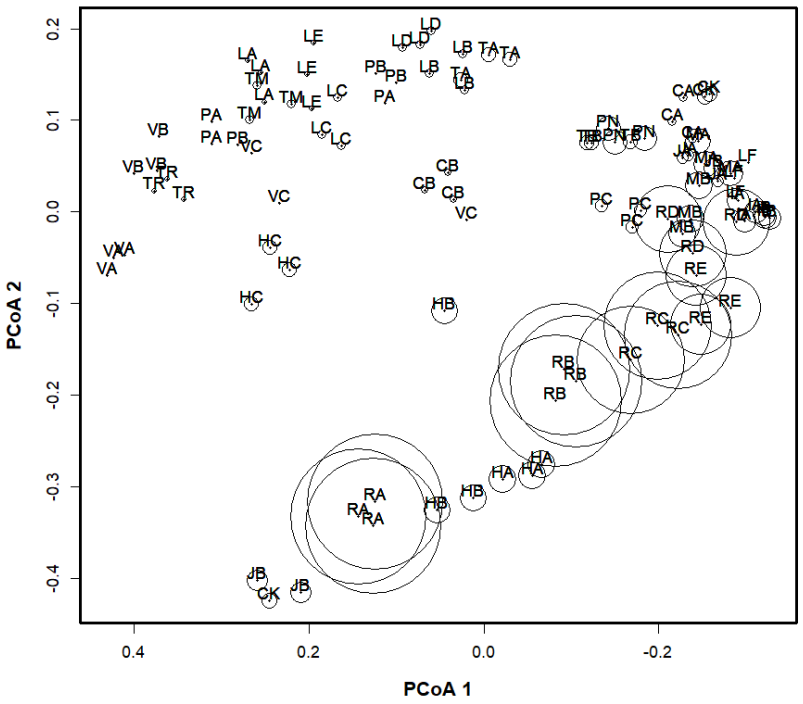


Fig. 6.18 Bubble plots illustrating the responses of total eukaryotic communities across 34 sites to significant environmental variables identified in the BIOENV analysis. **A)** shows salinity with DM568 data, while **B)** shows salinity with JB3 data. The plots are based on PCoA ordination using Bray-Curtis dissimilarity, with no ASV pruning applied. Bubble sizes represent the intensity of the respective environmental variable at each site. Site codes as in **Table 2.1**.

A AEMCu (G18S)



B LT63 (TAR)

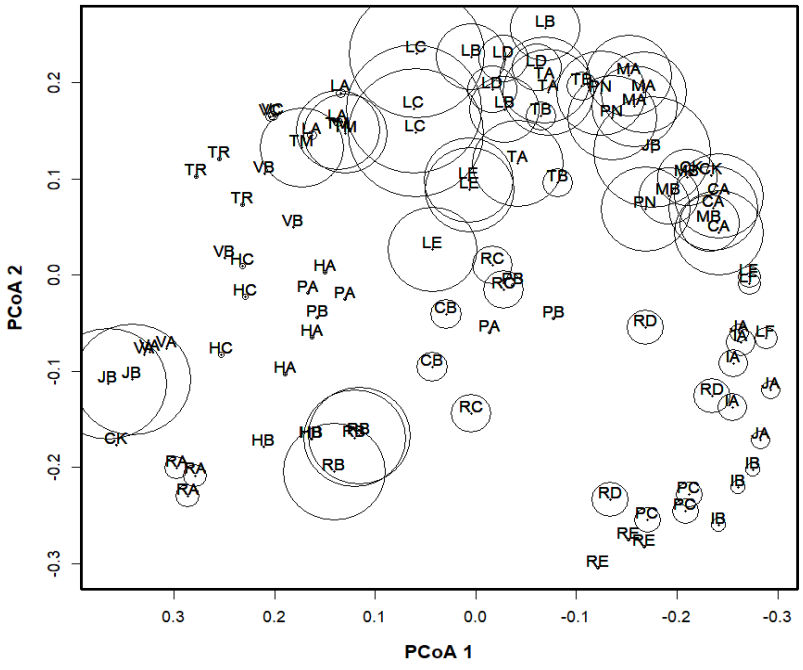


Fig. 6.19 Bubble plots illustrating the responses of total eukaryotic communities across 34 sites to significant environmental variables identified in the BIOENV analysis. **A)** shows AEMCu with G18S4 data, while **B)** shows LT63 with TAR data. The plots are based on PCoA ordination using Bray-Curtis dissimilarity, with no ASV pruning applied. Bubble sizes represent the intensity of the respective environmental variable at each site. Site codes as in **Table 2.1**.

What patterns emerged for specific taxonomic groups? Stramenopila showed moderate and high correlations under different primers, with ITS-12 aligning with AVS (0.37) and G18S-34 was associated with LT63, D50 and Sal as in (**Fig. 6.21 A**). Alveolata and Annelida both showed considerable relationships to PWCu in some primers. For instance, TAR-12 produced combined correlations of 0.60 (**Fig. 6.20 A**) and 0.54 (**Fig. 6.20 B**) for Alveolata and Annelida, respectively. Salinity (**Fig. 6.21 B**) and sediment texture (LT63, D50) may also coincide with their distributions, but whether these variables interacted with copper was not confirmed by the present data. Platyhelminthes in TAR-34 correlate with LT63, D50 and Sal (**Fig. 6.22**). In TAR-12, no single variable attained significance, but a combined correlation of 0.56 was observed. A correlation of 0.46 in TAR-34 likewise was recorded for the same variables.

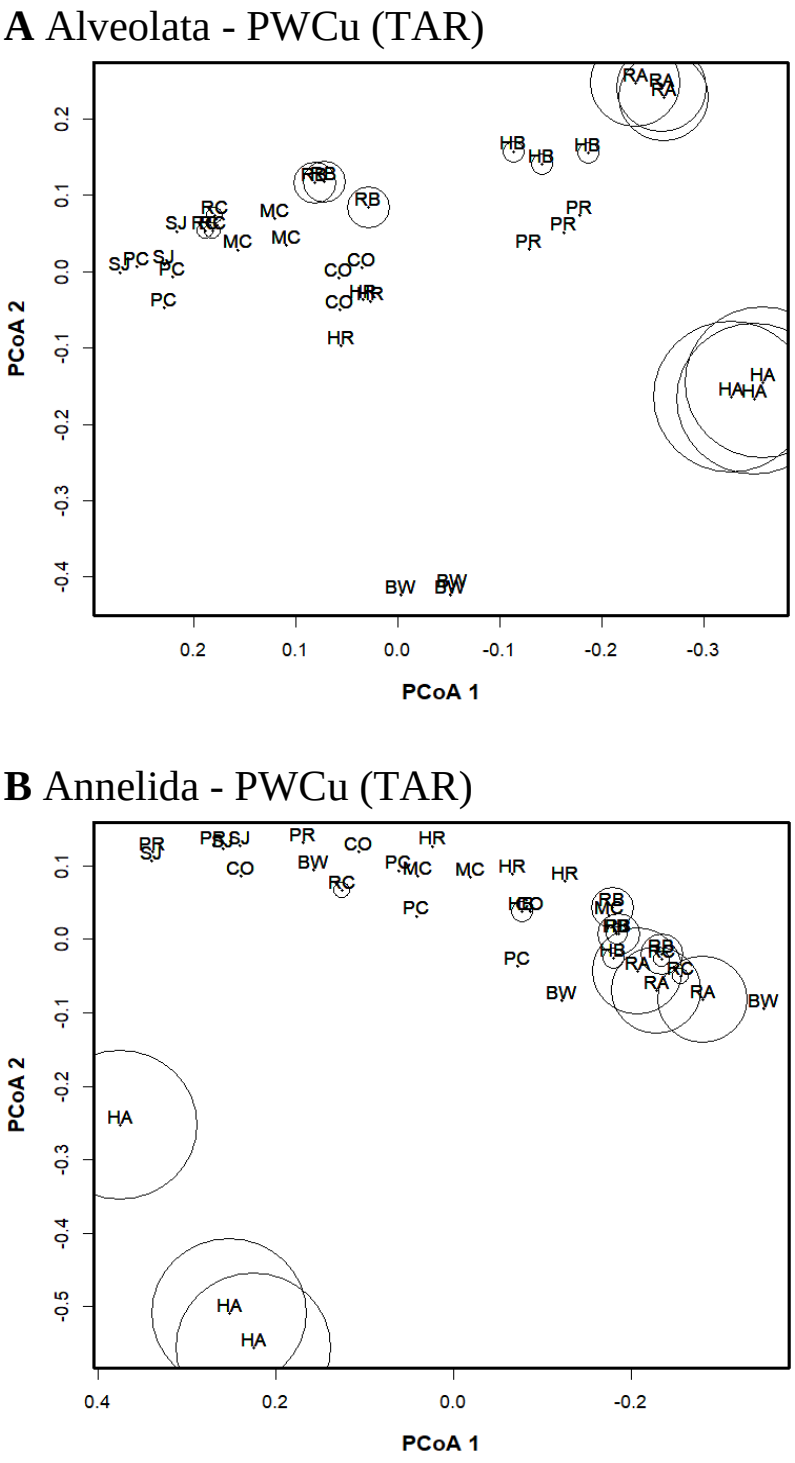
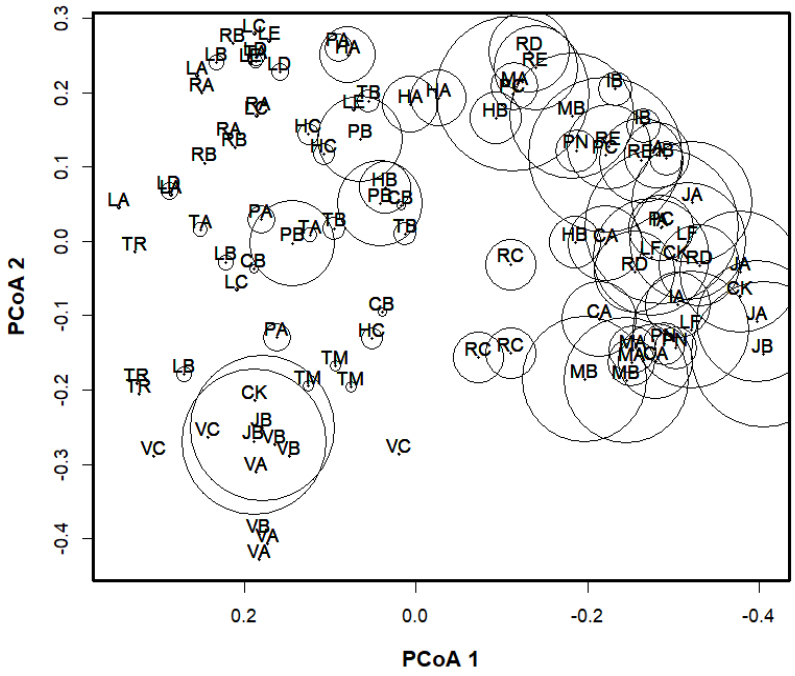


Fig. 6.20 Bubble plots illustrating the responses of the main metazoan and protistan phyla communities across 12 sites to key environmental variables identified as significant in the BIOENV analysis. **A)** shows Alveolata in relation to PWCu (TAR), while **B)** shows Annelida in relation to PWCu (TAR). The plots are based on PCoA ordination using Bray-Curtis dissimilarity, with no ASV pruning applied. Bubble sizes represent the intensity of the respective environmental variable at each site. Site codes as in **Table 2.1**.

A Stramenopiles - salinity (G18S4)



B Alveolata - salinity (TAR)

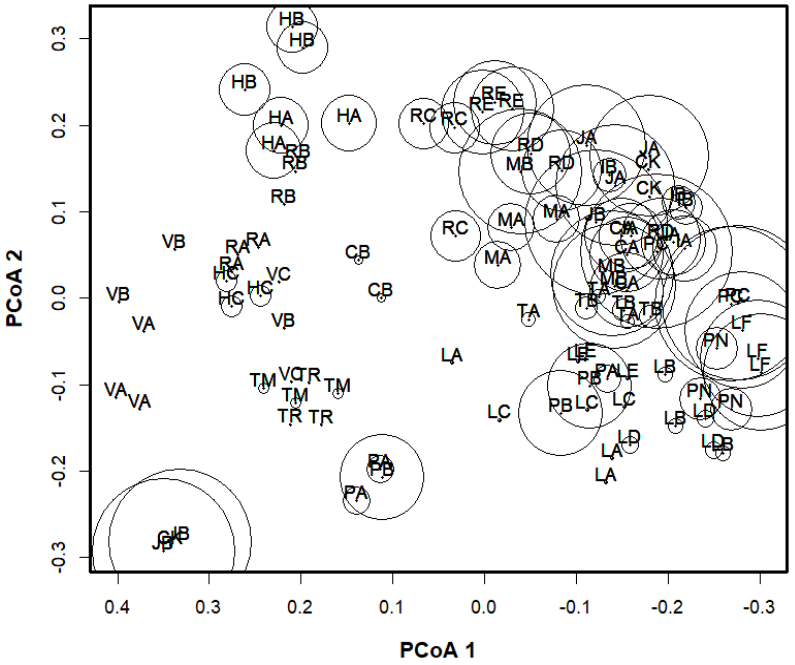


Fig. 6.21 Bubble plots illustrating the responses of key metazoan and protistan phyla across 34 sites to significant environmental variables identified in the BIOENV analysis. **A)** shows Stramenopiles in relation to salinity (G18S4), while **B)** shows Alveolata in relation to salinity (TAR). The plots are based on PCoA ordination using Bray-Curtis dissimilarity, with no ASV pruning applied. Bubble sizes represent the intensity of the respective environmental variable at each site. Site codes as in **Table 2.1**.

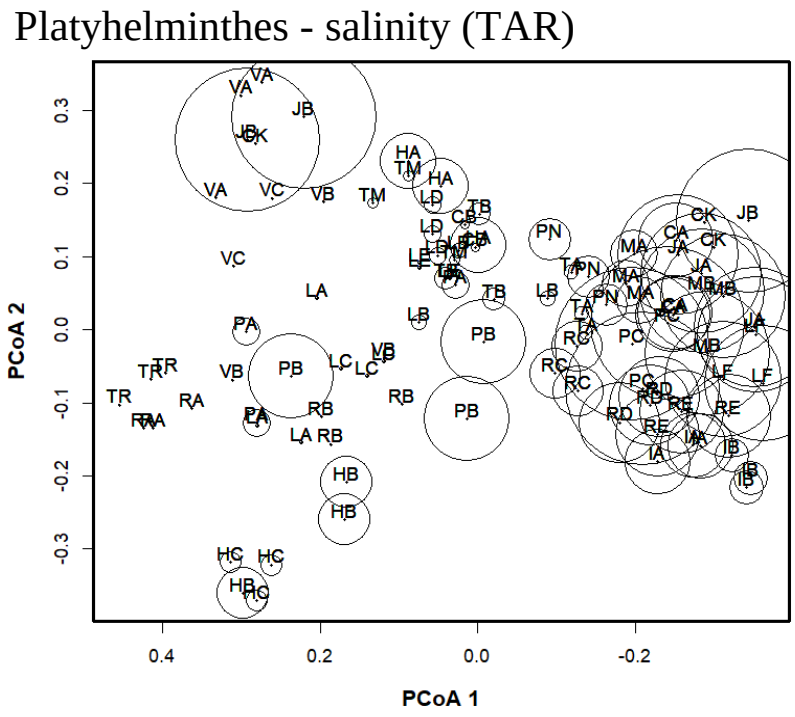


Fig. 6.22 Bubble plots illustrating the responses of key metazoan and protistan phyla across 34 sites to significant environmental variables identified in the BIOENV analysis. It shows Platyhelminthes in relation to salinity (TAR). The plots are based on PCoA ordination using Bray-Curtis dissimilarity, with no ASV pruning applied. Bubble sizes represent the intensity of the respective environmental variable at each site. Site codes as in **Table 2.1**.

Nematodes displayed comparatively clearer multi-parameter correlations. G18S-12 can reach combined correlations up to 0.69, featuring PWCu (**Fig. 6.23 A**), AEMCu, DOC and PW_OC_Cu (**Fig. 6.23 B**). In G18S4-34, D50 appeared prominently (**Fig. 6.24**). Mollusca under G18S-12 reached a combined correlation of (0.75) with PWCu and D50. By contrast, Fungi and Arthropoda exhibited moderate or inconsistent correlations with copper but were more strongly aligned with Sal and LT63 under DM568-34 (**Figs. 6.25 A** and **B**).

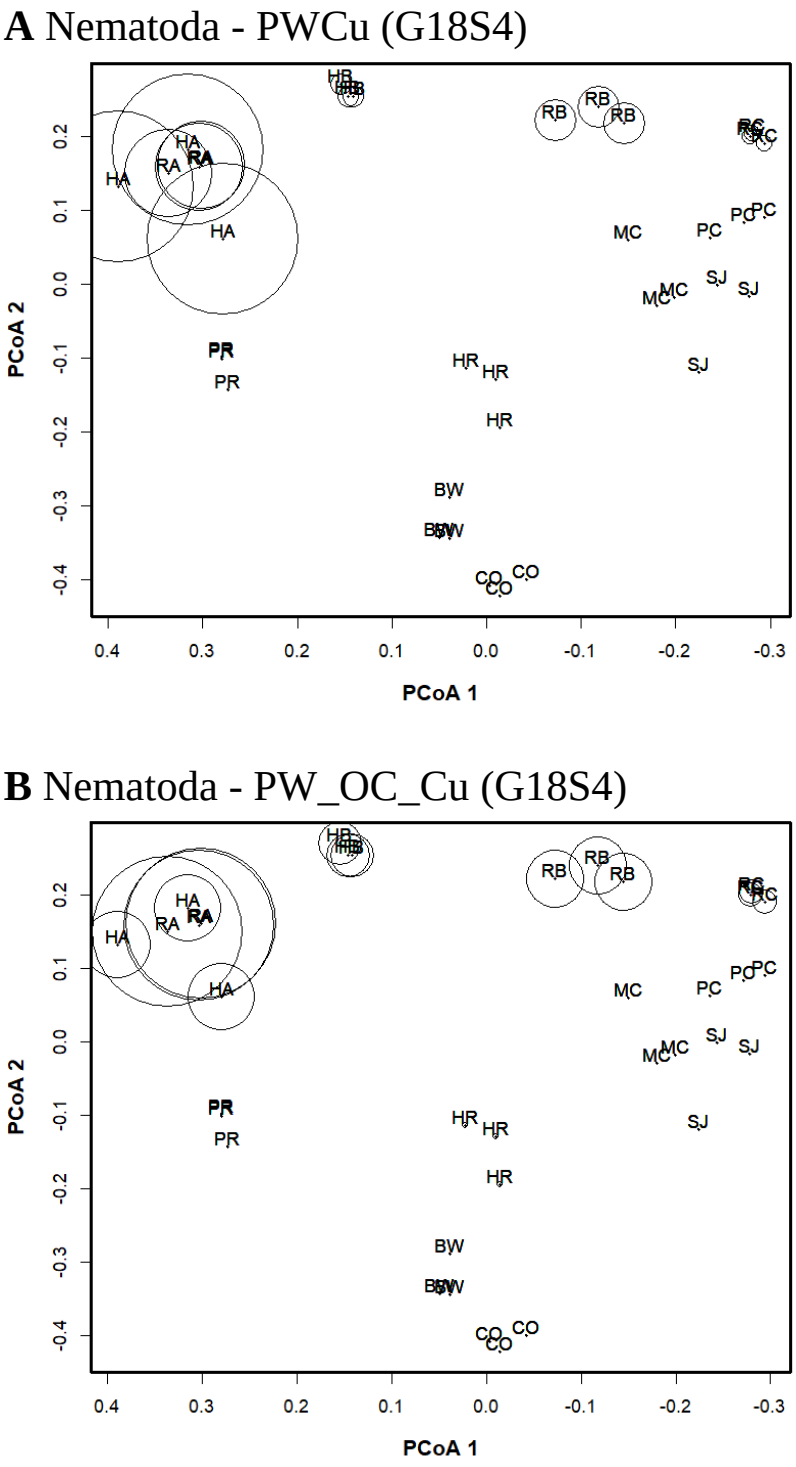


Fig. 6.23 Bubble plots illustrating the responses of the main metazoan and protistan phyla communities across 12 sites to key environmental variables identified as significant in the BIOENV analysis. **A)** shows Nematoda in relation to PWCu (G18S4), while **B)** shows Nematoda in relation to PW_OC_Cu (G18S4). The plots are based on PCoA ordination using Bray-Curtis dissimilarity, with no ASV pruning applied. Bubble sizes represent the intensity of the respective environmental variable at each site. Site codes as in **Table 2.1**.

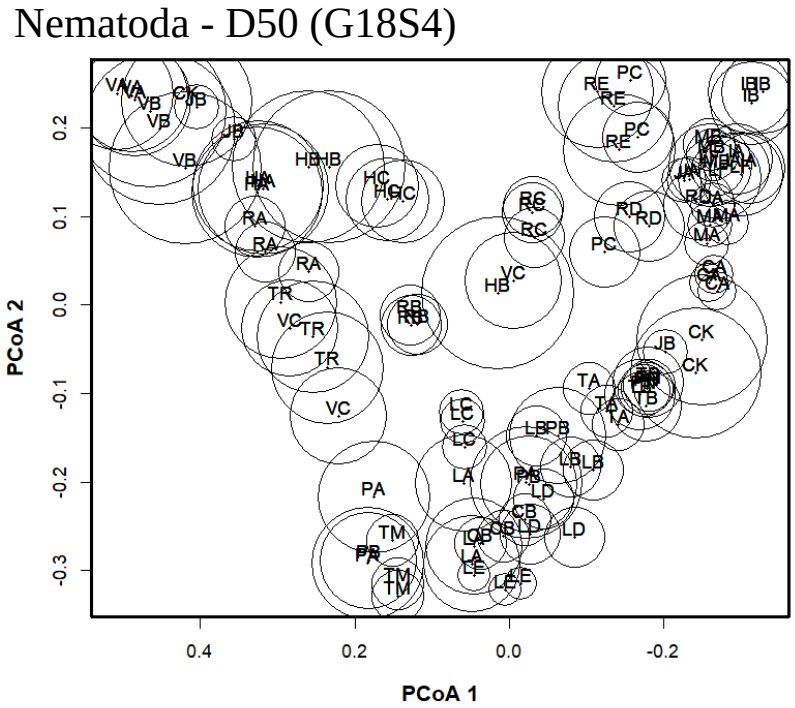


Fig. 6.24 Bubble plots illustrating the responses of key metazoan and protistan phyla across 34 sites to significant environmental variables identified in the BIOENV analysis. It shows Nematoda in relation to D50 (G18S4). The plots are based on PCoA ordination using Bray-Curtis dissimilarity, with no ASV pruning applied. Bubble sizes represent the intensity of the respective environmental variable at each site. Site codes as in **Table 2.1**.

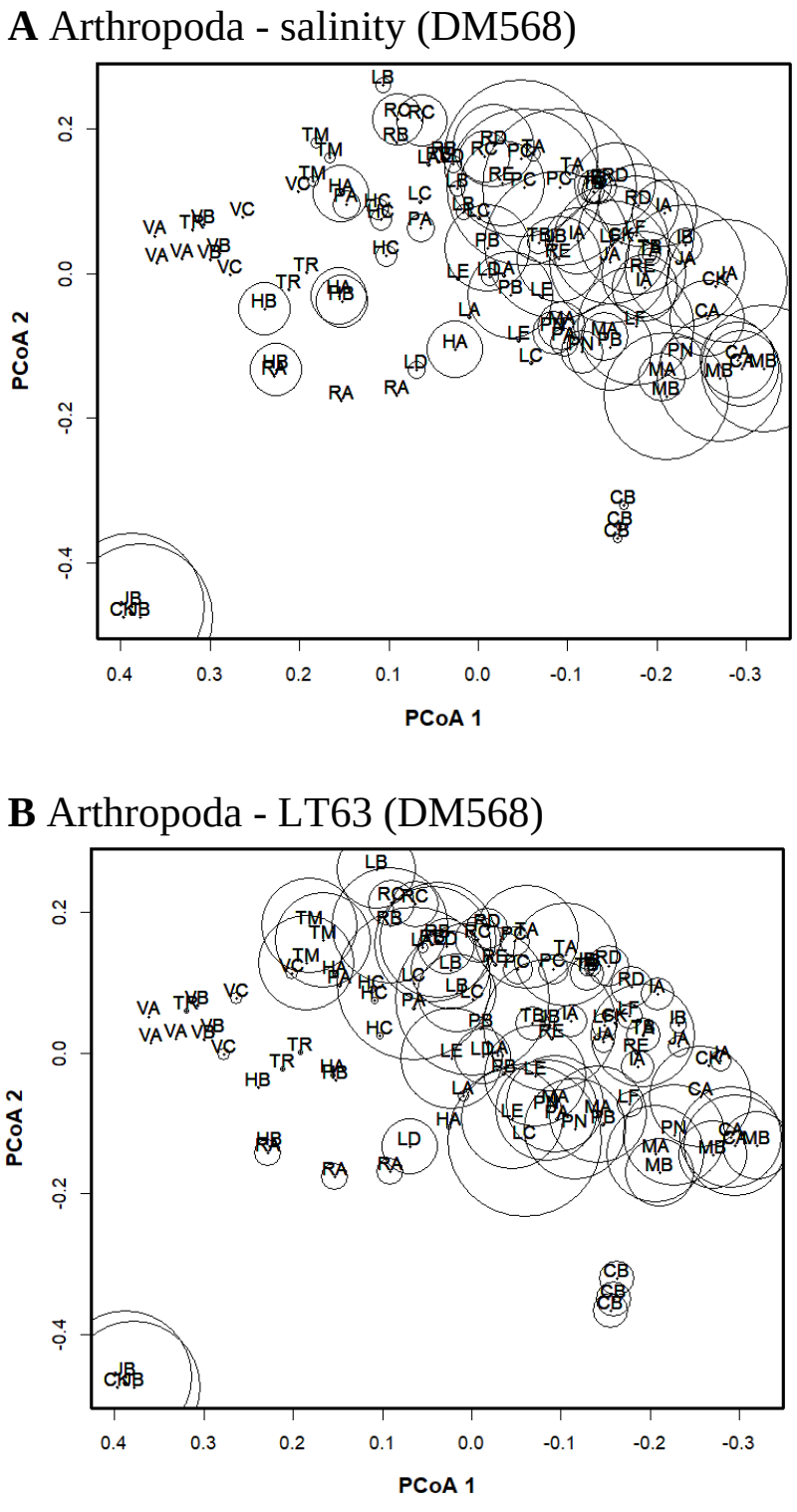


Fig. 6.25 Bubble plots illustrating the responses of key metazoan and protistan phyla across 34 sites to significant environmental variables identified in the BIOENV analysis. **A)** shows Arthropoda in relation to salinity (DM568), while **B)** shows Arthropoda in relation to LT63 (DM568). The plots are based on PCoA ordination using Bray-Curtis dissimilarity, with no ASV pruning applied. Bubble sizes represent the intensity of the respective environmental variable at each site. Site codes as in **Table 2.1**.

6.4 Discussion

Across 12 and 34 estuarine sites, porewater copper reached up to 431 µg/L and acid extractable sediment copper up to 1798 µg/g, exceeding the ranges covered by many earlier studies. This 90-fold span in copper levels was substantially larger than most previous surveys (Kalu et al., 2023, Dewi et al., 2024), and such an extensive gradient is rarely captured within a single estuarine study, allowing a clearer statistical separation of metal effects from background variability (Grant, 2002). Notably, replicate variation was much lower than site-to-site variation, indicating that local conditions strongly influenced eukaryotic assemblages. This pattern shows that the three within-site cores reflected fine-level patchiness, while the 12 or 34 distinct locations captured broader environmental variation. To broaden taxonomic coverage, this study used multiple primer sets (ITS, G18S, TAR, NEM, DM568 and JB3) which detected a wide range of taxa, including nematodes, annelids, arthropods, diatoms (Stramenopila) and alveolates. Because rRNA and mitochondrial barcodes occur in tens to thousands of copies per cell, taxa with high gene-copy numbers, particularly Metazoa, can dominate read pools even when biomass is similar (Pompanon et al., 2012, Aylagas et al., 2014, Bucklin et al., 2016). In line with **Chapter 3** findings, each primer demonstrated strengths for certain groups. The ITS primer pair yielded moderate read counts, yet fungal reads comprised only 5% of the total, aligning with previous observations (Schmidt et al., 2013, Harnelly et al., 2022), whereas G18S data were dominated by Metazoa, in line with Stoeck et al. (2010) and Tytgat et al. (2019). However, this predominance might partly reflects the high rRNA gene copy numbers in metazoan genomes, which can inflate read proportions relative to groups with fewer copies per genome (Pompanon et al., 2012, Aylagas et al., 2014, Bucklin et al., 2016)TAR amplified a diverse range of eukaryotes as indicated by Fonseca et al. (2022) and Maosa et al. (2024), though with slightly lower efficiency. The NEM was particularly effective in amplifying nematodes, consistent with the findings of Porazinska et al. (2009), Sapkota and Nicolaisen (2015). The DM568 was originally described as particularly adept at amplifying nematode species (Kounosu et al., 2019) and the data here confirm effective nematode recovery, yet larger read counts were recorded for arthropods and more ASVs emerged for stramenopiles, fungi and alveolates, reflecting the primer's broader 28S-based coverage (Machida and Knowlton, 2012, Kounosu et al., 2019). Lastly, the JB3 was anticipated to capture COI sequences from certain metazoan lineages, including arthropods, molluscs and platyhelminths, as described by Tytgat et al. (2019), but yielded minimal detections across the major groups in

this dataset, with a sizeable fraction of reads remaining unassigned and offering less taxonomic resolution than the other primers. Additionally, the generally broader coverage and more complete reference libraries available for 18S often enable deeper classification at genus or species level than 28S-based or COI-based approaches. While DM568 can recover diverse taxa, its reference database remains less extensive than that for G18S, which limits fine-scale taxonomic assignments. Similarly, JB3's poor performance in assigning reads at the Kingdom level underscores the importance of filling coverage gaps in current COI databases. However, taxonomic groupings should be interpreted with caution because the higher-level classification of eukaryotes remains incomplete and sometimes inconsistent; for instance, the SAR supergroup (Stramenopila, Alveolata and Rhizaria) is treated separately from Archaeplastida, which itself encompasses Rhodophyta, Glaucophyta and Chloroplastida, the latter uniting both Streptophyta and green algae (Mackiewicz and Gagat, 2014, Yazaki et al., 2022). Consequently, this multi-primer metabarcoding strategy reduced the "blind spots" often seen when fewer sites or primer sets were used (Borja and Dauer, 2008, Dewi et al., 2024). Nevertheless, no single primer set captured the full spectrum of lineages, underscoring ongoing concerns that primer biases and incomplete reference databases may overlook groups such as Mollusca or foraminiferans (Bik et al., 2012, Schoenle et al., 2021). Several broad eukaryotic metabarcoding studies (Fonseca et al., 2022, Pawlowski et al., 2024) have likewise shown that multiple markers were needed to represent diverse lineages, supporting the rationale for using multiple primers in this work. Nevertheless, some primer sets (particularly, ITS, NEM and JB3) exhibited higher replicate variability, possibly because large-bodied taxa can be patchily distributed in a 10 g sample or because certain groups show small-scale habitat heterogeneity, leading to occasional absences or presences in replicates (Thornton et al., 2011, Nichols et al., 2018). Conversely, primers that inflate read numbers for a few taxa may under-detect groups with poor binding or low gene copies, potentially masking richness (Elbrecht and Leese, 2015, Lamb et al., 2019). G18S and TAR primer sets showed more stable results among replicates, likely because they captured a broad range of smaller metazoans and microeukaryotes that occurred consistently across subsamples. Nevertheless, the use of multiple primers covering diverse taxa allows consistent detection of community shifts across contamination gradients, demonstrating that metabarcoding can robustly capture overall patterns despite each primer's inherent limitations.

These results build on the morphological tradition for meiofauna (Coull, 1992, Semprucci et al., 2015) by detecting smaller taxa that macrofaunal surveys often miss (Hummon, 1981, Giere, 2009). In 10-gram sediment samples like those used here, juvenile macrofauna can occupy the same size range as typical meiofaunal organisms, serving as "temporary meiofauna" and bridging the boundary between permanent meiobenthos and larger taxa (Warwick, 2018). This overlap underscores the need for integrative analyses across multiple size classes, since early life stages of macrofauna may pass through the same sieves used to capture smaller, permanently meiofaunal species. This strategy aligns with evidence that meiofauna may be especially sensitive to pollution (Moore and Bett, 1989, Coull, 1999). In contrast, morphological methods require substantial taxonomic expertise and large sample volumes (Warwick et al., 1991, Kendall and Widdicombe, 1999). Metabarcoding offers a faster, broader alternative (Gielings et al., 2021), although its success depends on expanding reference databases and mitigating primer biases (Baetscher et al., 2023). Additionally, morphological surveys often overlook soft-bodied platyhelminths or molluscs (Mitsi et al., 2019, Balsamo et al., 2020), while broader eukaryotic literature showed that alveolates and stramenopiles can also be underrepresented by single-marker methods (Burki et al., 2021, Marinchel et al., 2024). The reliance on easily observable taxa like shelled organisms and the shortage of expertise in identifying soft-bodied species have led to skewed entries in databases such as (WoRMS), further perpetuating incomplete coverage (Rosenberg, 2014, McClain et al., 2025). By extracting DNA directly from whole sediment (without sieving, flotation or elutriation), this study also reduced sample processing time and bypassed usual biases in separating macrofauna from meiofauna (Borja and Dauer, 2008). Although the unsieved approach maximises recovery of small metazoans and microbes, it also captures trace DNA from incidental macrofaunal fragments (Bik et al., 2012). These "by-catch" reads should therefore be interpreted with caution when the study focus is limited to meiofauna or microfauna (Klunder et al., 2019).

Bray-Curtis ordinations are retained as the primary depiction of eukaryotic community patterns because their abundance weighting gives finer resolution of site differences than presence-absence metrics (Weiss et al., 2017). However, because Bray-Curtis relies on read counts, it inherits the semi-quantitative limitations of metabarcoding abundances, including rRNA copy number variation, primer bias and PCR stochasticity (Leray and Knowlton, 2017, Lamb et al., 2019), as detailed in **Chapter 1**. A complementary Jaccard ordination, based solely on presence-absence data, yielded broadly similar site groupings, confirming that the observed

gradient is robust while also providing a conservative check against over-interpreting abundance magnitudes. This concordance indicates that Bray-Curtis still captures meaningful ecological differences when biases are systematic, as noted by Luo Luo et al. (2023). Interpreting both metrics together therefore combines the sensitivity of abundance weighting with the safeguard of binary data, aligning with recommendations from Deagle et al. (2009) Krehenwinkel et al. (2017).

A key goal was to distinguish copper-driven changes in eukaryotic communities from natural gradients such as salinity and sediment texture (LT63, D50). Many metabarcoding studies have found it difficult to separate metal impacts from other factors (Chen et al., 2022, Zeng et al., 2023), especially when only a limited number of sites have been sampled (Kalu et al., 2023, Dewi et al., 2024). Heavy metals pose a serious ecological threat in marine sediments due to their toxicity, persistence and bio accumulative potential (Chapman et al., 1998, Grant, 2010, Ardila et al., 2023). Yet confounding variables such as salinity, sediment texture and nutrient enrichment, often mask the direct effects of contamination (Grant, 2002, Gillan et al., 2005). In estuarine settings, changing physicochemical conditions can alter metal toxicity and produce heterogeneous responses (Ogilvie and Grant, 2008). This study showed that porewater copper (PWCu) strongly shaped community composition in the 12-site dataset, which included several sites with very high Cu. In the Fal and Hayle estuaries, known for elevated copper (Somerfield et al., 1994b), sites clustered distinctly in ordination plots, supporting earlier suggestions that copper can be a major ecological driver here (Ogilvie and Grant, 2008). Even so, salinity and grain size emerged as important covariates, aligning with studies that note multiple stressors in estuaries (Brannock et al., 2016, Li et al., 2024). Earlier morphological work in the same region (Somerfield et al., 1994c) attributed nematode differences to salinity and habitat factors, rather than metals alone. They employed multiple sites in each creek but did not replicate within a site, meaning they primarily assessed among-creek variation within the Fal system, while the present study encompasses more estuaries and additional replicates per site. By spanning a 90-fold range in porewater copper (2-431 µg/L), this study offers more conclusive evidence that copper influences benthic communities while acknowledging the role of other environmental variables. Similar to findings in wider marine metabarcoding efforts (Tagliabue et al., 2023, Geraldi et al., 2024), overlapping gradients can reduce the visibility of metal-specific effects. For instance, alveolates and some annelids correlated more with copper at higher concentrations (Marinchel et al., 2024, Mazurkiewicz et al., 2024), whereas certain nematode

or arthropod taxa showed stronger links to salinity or sediment factors (Derycke et al., 2010, Tytgat et al., 2019). Reliable indicators help avoid confusing metal-driven shifts with general environmental "noise" (Grant, 2002, Grant, 2010).

Nematodes, matching previous morphological studies that highlight their rapid turnover and metal sensitivity (Heip et al., 1985, Schratzberger et al., 2006), were abundant under higher copper concentrations but exhibited even greater species-level diversity with metabarcoding, which was consistent with findings that morphological counts may underestimate cryptic or rare taxa (Derycke et al., 2010, Tytgat et al., 2019). Alveolates and certain annelids appear strongly associated positively with copper, although these patterns may partly reflect salinity and grain-size effects. In contrast, arthropods, stramenopiles, platyhelminths and fungi display weaker or inconsistent relationships with copper. Such findings expand on past suggestions (Coull, 1992, Semprucci et al., 2015) that various meiofaunal taxa can act as bioindicators. They further extend the morphological observations of Somerfield et al. (1994c) and Millward and Grant (1995) to include diatoms (Stramenopila) and alveolates, organisms less accessible in traditional surveys. These results did not pinpoint a single species. Still, the observed patterns in both presence-absence and abundance across multiple taxa suggest that a multi-taxon approach may help capture metal-driven community changes, particularly under severe metal gradients (Borja and Dauer, 2008). Additionally, Other studies have noted that annelids often thrive in habitats shaped by glacial sedimentation, high turbidity, scouring and unstable sediments (Mazurkiewicz et al., 2024) while Warwick (2001) observed that opportunistic annelids were especially numerous at heavily polluted sites, which aligns with the high annelid representation observed here.

Site coverage played a central role in detecting pollution-related trends. The 12-site dataset revealed more pronounced copper relationships, possibly because these sites included extremely high Cu values. In contrast, the 34-site dataset featured greater salinity and sediment variability, sometimes partially obscuring copper effects. This pattern aligns with earlier remarks that moderate pollution or fewer sites can hamper clear interpretations (Dewi et al., 2024, Múrria et al., 2024). Tagliabue et al. (2023) and Nijland (2020) highlight how broad spatial coverage introduces additional environmental gradients, making single-stressor impacts more challenging to discern. Bacteria and Archaea data from **Chapters 4** and **5** of this thesis indicate a similar pattern, with clear separation at heavily contaminated sites (such as Hayle

and Restronguet) but reduced discrimination at moderate copper levels. Furhter analyses supported these trends indicating that eukaryote communities were the most distinctly separated domain in both datasets, with R-values of 0.997 for the 12-site data and 0.871 for the 34-site data, surpassing Bacteria (0.791, 0.707) and Archaea (0.927, 0.609) in each case. Comparisons with earlier morphological findings and whole-community toxicity assessments confirm this interpretation; Somerfield et al. (1994c) reported distinct, less diverse nematode assemblages in the Fal system's most metal-enriched areas and Millward and Grant (1995) showed that nematodes from chronically polluted Restronguet Creek exhibited higher copper tolerance than those from cleaner sites. The metabarcoding patterns observed here, especially in the highly contaminated subset, similarly indicate that copper-rich sediments correspond to shifts in community composition and elevated tolerance, reflecting the same metal-driven changes identified by these morphological and pollution-induced community tolerance studies. By using both datasets, the present work therefore extends earlier, narrower efforts into a broader spatial context and confirms that site coverage can determine whether strong copper signals emerge amid other environmental factors.

Applying ASVs for sequence analysis and testing different pruning thresholds revealed that retaining rare taxa can expose subtle pollution signals, illustrating how modern bioinformatics can detect minor lineages potentially overlooked by traditional morphotyping (Hummon, 1981, Giere, 2009). The results suggest that some morphological species may actually represent multiple ASVs, suggesting cryptic diversity or unresolved taxonomic boundaries (Derycke et al., 2010, Tytgat et al., 2019, De Luca et al., 2021). As noted by Bik et al. (2012), incomplete or uneven taxonomic references can hamper ASV classification and the many unassigned or partially identified reads in this study underscore that local databases remain insufficient for several lineages. Although this study relied on DNA metabarcoding, complementary RNA approaches are increasingly available. Primer-free sequencing of environmental rRNA (eRNA) can generate full-length 16S and 18S reads and thus profile the metabolically active community (Pochon et al., 2015, Karst et al., 2018). Both eRNA and shotgun metagenomics bypass primer bias but remain more costly and bioinformatically demanding than amplicon sequencing (Eloe-Fadrosh et al., 2016). For routine monitoring, DNA metabarcoding is therefore still the most pragmatic option, provided read counts are interpreted as semi-quantitative indicators (Pompanon et al., 2012, Deagle et al., 2019). Multi-primer eukaryotic metabarcoding can capture macro- and microfaunal assemblages under severe metal pollution, surpassing earlier

work limited by smaller contamination ranges or fewer sites (Borja and Dauer, 2008, Dewi et al., 2024). As databases improve and primer sets become more refined, metabarcoding could feasibly supplement or even replace morphological techniques for routine monitoring, given its efficiency, breadth of taxa and lower manpower requirements (Gielings et al., 2021). By extracting DNA directly from whole sediment (without flotation or elutriation), this study also bypassed usual biases in separating macrofauna from meiofauna (Borja and Dauer, 2008). Nevertheless, diverse lineages like foraminiferans may still go undetected unless specifically targeted (Pawlowski et al., 2024), underscoring that enhanced references and primer designs did not completely eliminate coverage gaps.

Future efforts could concentrate on highly contaminated and genuinely clean sites to sharpen copper-driven trends, while also measuring additional environmental factors (e.g. redox, sulphides) to show how metals interact with other estuarine variables (Brannock et al., 2016, Li et al., 2024). Although morphological surveys and multi-primer metabarcoding have each been used to track metal-driven shifts in benthic communities (Heip et al., 1985, Semprucci et al., 2015, Kalu et al., 2023), the present study offers a novel application in a distinct environment, showed that comprehensive coverage and robust sampling design can detect metal-related community changes.

Overall, this study demonstrates that high copper levels can significantly shape eukaryotic communities in estuarine sediments, while salinity and sediment factors also play influential roles. By spanning a 90-fold range in porewater copper and employing multiple primers, it advances current knowledge, which has often been restricted to milder gradients or few sites (Kalu et al., 2023, Dewi et al., 2024). Clearly, replicate variation remained lower than site-level differences, likely reflecting fine-grained sediment heterogeneity typical of mudflats. The three within-site cores captured local-scale patchiness, while the 12- and 34-site designs encompassed broader spatial turnover. However, sampling coverage and replication are never perfect, and limited cores per site can reduce power to detect subtle effects (Tagliabue et al., 2023). Unquantified differences in DNA-extraction efficiency may also bias relative read counts (Emilson et al., 2017), while PCR stochasticity introduces additional technical noise that particularly affects rare ASVs (Nichols et al., 2018). The mean richness detected here (~ 350-420 eukaryotic ASVs site-1 with G18S) matches values from temperate North-Sea muds but remains below those reported for tropical mangrove sediments, echoing

recognised latitude-linked diversity gradients (Brannock et al., 2016; Múrria et al., 2024). Nematodes, diatoms and certain alveolates consistently respond to copper, suggesting a multi-taxon indicator framework might be valuable for future assessments (Borja and Dauer, 2008). Nevertheless, some groups may still evade detection without specialised primers or curated databases and confounding gradients can persist even in multi-primer studies (Nijland, 2020, Pawlowski et al., 2024). Future work that pairs DNA metabarcoding with eRNA profiling of active rRNA and targeted metatranscriptomics could help separate metal impacts from natural variation (Pochon et al., 2015; Karst et al., 2018), while improved local references and thoughtful sampling strategies will further strengthen coastal pollution assessments using molecular tools (Pawlowski et al., 2018; Leese et al., 2018).

Chapter 7:

Discussion, Implications and Future Directions

7.1 Restatement of Aims and Context

This thesis set out to investigate how historic metal contamination, particularly copper, affects benthic microbial (bacterial, archaeal) and eukaryotic communities in southwestern England using metabarcoding. Past Cornish mining generated large volumes of copper, tin and arsenic-rich wastes (Barton, 1961), creating persistent contamination in estuaries such as Restronguet Creek and the Hayle (Bryan and Gibbs, 1983, Grant, 2010). Although these systems are now subject to minimal new industrial inputs, copper and other metals remain in sediments indefinitely (Alloway, 2012). Under GESAMP's framework, estuaries classed as "contaminated" do not automatically qualify as "polluted" unless there is demonstrable harm to biological communities (Bryan and Hummerstone, 1971, Grant, 2010). The overarching goal was thus to determine whether elevated copper levels correlate with measurable ecological impacts, ranging from shifts in bacterial and archaeal composition to changes in eukaryotic diversity while also refining methodological approaches for detecting these pollution effects.

7.2 Methodological Advances and Rationale

A key challenge was the heterogeneous nature of estuarine sediments, so triplicate 10 g cores spaced ~0.5 m apart were taken at each site to capture fine-scale sediment variability. Results from ANOSIM and ordinations showed replicates clustered much more tightly than samples from different sites, confirming that the design adequately sampled local variability, which can hamper DNA extraction and bias PCR-based analyses due to co-extracted humic and fulvic acids inhibiting polymerase activity and reducing amplicon yield (Tebbe and Vahjen, 1993). Analysis in **Chapter 2** outlined how tailored dilution strategies and optimised PCR protocols helped surmount these barriers, paving the way for accurate molecular assessments even under high metal loads. Building on these optimisations, the thesis employed multi-target

metabarcoding (16S, ITS, 18S, 28S and COI) to survey a broad range of organisms (**Chapters 3**, **4**, **5** and **6**). This approach addressed the historical limitation of single-marker methods, which can underrepresent subgroups within the intended barcode group and omit entire non-target groups, thereby underestimating overall benthic diversity (Tytgat et al., 2019, Fonseca et al., 2022). One key advantage of extracting DNA from whole sediment was that it captures the full microbial eukaryotic community in a single workflow, potentially preventing the loss of rare or fragile taxa that might occur during fractionation steps (Sapkota and Nicolaisen, 2015). In addition, direct sediment approaches can yield a more holistic snapshot of benthic biodiversity, as indicated by the high nematode proportions found in large-volume samples (Fais et al., 2020). To ensure that ordination patterns were not an artefact of abundance weighting, complementary presence-absence ordinations (Jaccard) were generated and found to mirror the Bray-Curtis results, reinforcing that the main gradients reported here were method-independent. In addition, sampling 12- and 34-site datasets expanded the contamination gradient from as low as 2 µg/L porewater Cu up to 431 µg/L, providing a spectrum over which we would expect both subtle and extreme effects of metals to occur (Bryan and Hummerstone, 1971, Bryan and Gibbs, 1983). Across both domains, accumulation analyses showed that increasing the number of estuarine sites yielded a greater return in detected diversity than simply sequencing existing sites more deeply; this highlighted spatial heterogeneity as the principal constraint on recovering the full microbial complement of metal-impacted sediments. One key advantage of extracting DNA from whole sediment is that it captures the full microbial eukaryotic community in a single workflow, potentially preventing the loss of rare or fragile taxa that might occur during fractionation steps (Sapkota and Nicolaisen, 2015). Moreover, direct sediment approaches can yield a more holistic snapshot of benthic biodiversity, as indicated by the high nematode proportions found in large-volume samples (Fais et al., 2020).

7.3 Thresholds and Community Shifts in Bacteria

Analysis in Chapter 4 focused on bacterial assemblages and demonstrated that porewater copper exerted a discernible influence on community structure. Bacterial alpha-diversity remained relatively stable until roughly 20 µg/L porewater Cu, consistent with earlier thresholds reported by Ogilvie and Grant (2008). Beyond this point, compositional changes intensified and community composition at heavily polluted sites (HA, HB, RA, RB, RC and

RD), some of which had porewater Cu levels above 100 µg/L, was very distinctive. In spite of that, bacteria retained sufficient functional redundancy that alpha-diversity declines were modest compared with those of macrofauna in severely polluted marine systems (Olsgard and Gray, 1995). These patterns reinforced the notion that bacterial communities can tolerate moderate copper loads before showed pronounced reorganisation. However, confounding factors such as salinity, sediment grain size and organic matter may mask copper-driven shifts in community composition at intermediate contamination levels (Grant, 2002).

7.4 Archaeal Sensitivity and Bioindicator Potential

The findings in Chapter 5 revealed that archaea often exhibit stronger or earlier responses to copper contamination than bacteria in smaller, more uniform subsets of sites. For instance, in the 12-site analysis, archaeal richness declined sharply once porewater Cu exceeded 20 µg/L and showed a moderate but significant negative association with PWCu (rho = -0.64) whereas the corresponding bacterial correlation was weak (rho = -0.23) yet still statistically significant. This difference may reflect lower functional redundancy among archaea or their distinct detoxification pathways (Sandaa et al., 1999, Yu et al., 2024). The broader 34-site dataset, however, introduced additional heterogeneity (e.g., salinity, sediment texture) that diluted the copper signal, mirroring the challenges in estuaries with multiple overlapping stressors (Grant, 2010, Zou et al., 2020). Within heavily contaminated sites, specific OTUs from Thermoproteota or Nanoarchaeota appeared more abundant, suggesting these particular lineages, rather than entire phyla, may be copper-tolerant indicators in environments where confounding factors were less prominent. Furthermore, Thermoproteota's resilience was observed in the 34-site dataset, whereas Nanoarchaeota occurred primarily at heavily polluted sites in the 12-site subset, a discrepancy likely arising from different sampling seasons rather than site-specific factors. Thermoproteota include thermophilic sulfur reducers equipped with copper efflux and sequestration mechanisms, supporting their tolerance to elevated copper levels (Jay et al., 2016). In contrast, Nanoarchaeota are obligate epibiotic symbionts that rely on host-mediated detoxification to persist in metal-rich biofilms, lacking intrinsic copper resistance (Wurch et al., 2016).

7.5 Eukaryotic Communities and Copper-Driven Changes

Turning to eukaryotic assemblages (**Chapter 6**), multi-primer metabarcoding (ITS, G18S, TAR, NEM, DM568, JB3) revealed taxa such as nematodes, annelids, arthropods, diatoms (Stramenopila) and alveolates showed clear compositional turnover along the (2-431 µg/L porewater Cu). In the 12-site subset featuring extremely high copper, eukaryote communities clustered distinctly in ordination plots, indicating strong metal-driven changes reminiscent of earlier morphological findings (Somerfield et al., 1994c, Millward and Grant, 1995). By contrast, the 34-site analysis encompassed greater salinity and sediment variation, partially obscuring moderate copper signals. Replicate variation remained low relative to site-level differences, reinforcing that local environmental conditions drive eukaryotic assemblages (Moore and Bett, 1989, Coull, 1992). Among the more copper-tolerant lineages, certain nematode and alveolate groups correlated significantly with elevated Cu, though confounding gradients could still overshadow these relationships in moderately polluted sites. Meanwhile, arthropods, stramenopiles, platyhelminthes and fungi displayed weaker or inconsistent copper correlations, suggesting that not all eukaryotic taxa exhibit pronounced metal responses. Karst et al. (2018) found that arthropod rRNA reads dominated their marine-sediment dataset (~45 %). In the present study no single marker reproduced that proportion, but G18S recovered the largest share of arthropod sequences (~20 % of metazoan reads), NEM retrieved the highest relative abundance of nematodes (~40 %) and DM568 yielded the greatest proportion of platyhelminth reads (~10 %). Percentages drawn from WoRMS (e.g. Arthropoda 28 %, Mollusca 25 %) provide a global richness baseline, highlighting that read-based values are primer-dependent. In addition, variation in ribosomal-RNA gene-copy number can inflate read counts in fast-growing microbes (bacteria and protists), so apparent relative abundances should be interpreted with caution (Kembel et al., 2012). These differences reiterated that a multi-marker strategy is essential for balanced coverage of metazoan phyla. This domain-wide perspective showed that severe contamination can reshape entire eukaryotic communities, yet single-primer or smaller-scale studies might overlook important taxa (Borja and Dauer, 2008, Dewi et al., 2024).

7.6 Comparison of Archaeal, Bacterial and Eukaryotic Responses

Analysing domain responses across both the 12-site and 34-site datasets underscores how environmental heterogeneity affects threshold detection. In the smaller, more uniform Fal and Hayle subset:

- **Archaea** showed a sharper drop in diversity once copper surpassed 20 µg/L, suggesting they may serve as early-warning indicators in less confounded environments (Sandaa et al., 1999, Grant et al., 2023).
- **Bacteria** exhibited moderate declines but only underwent large compositional turnover at higher Cu loads.
- **Eukaryotes** (especially nematodes and alveolates) displayed clear shifts where copper levels were extreme, aligning with historical morphological evidence in southwestern estuaries (Somerfield et al., 1994a, Millward and Grant, 1995).

When considering the 34-site dataset, salinity, sediment texture and other variables blurred direct copper effects, indicating that archaea and eukaryotes can be reliable metal indicators only if site-level heterogeneity was accounted for (Nijland, 2020, Chen et al., 2022).

7.7 Significance of Multi-Marker Metabarcoding

This thesis demonstrates that no single primer set adequately captures the full complexity of benthic microbial and eukaryotic communities. Instead, employing multiple primers; 16S for bacteria; dedicated archaeal 16S sets and varied eukaryotic markers (18S, ITS, 28S, COI), optimises taxonomic coverage (Bik et al., 2012, Tytgat et al., 2019, Fonseca et al., 2022). While multi-marker metabarcoding significantly broadens our view, primer biases and incomplete databases remain limiting factors, especially for certain archaea (Nanoarchaeota) and eukaryotes (molluscs, foraminiferans). Nonetheless, this approach surpasses traditional methods by detecting cryptic or rare taxa that morphological surveys might miss, yielding a more refined picture of pollution-induced shifts (Coull, 1992, Borja and Dauer, 2008). Multi-primer data showed minimal replicate variation within each site compared to differences among sites, indicating that local conditions strongly shape benthic communities. However, despite extensive sequencing, metabarcoding did not always reflect physiological tolerance

assays reported for nematodes (Bik et al., 2012), highlighting that compositional data alone may overlook sub-lethal stress detected by bioassays. This discrepancy arises from insufficiently matched control locations, rather than shortcomings of the method. Expanding the number of sampled sites typically introduces further environmental differences that can mask contamination impacts. Consequently, finding "clean" sites with physical and chemical attributes mirroring those of polluted areas was more critical than simply adding loci or refining databases. Although this can be difficult in practice, it underscores the importance of precise site selection to isolate metal contamination as the primary driver of observed community shifts.

7.8 Management Implications

The findings confirm porewater copper as an important predictor of community change, especially when concentrations exceed 20 µg/L. While total sediment copper also matters, particularly for deposit feeders ingesting particles (Miller et al., 1984, Watling, 1998), porewater measurements often reflect bioavailable fractions more directly (Somerfield et al., 1994c, Ogilvie and Grant, 2008). Where copper loads exceed 100 µg/L, major compositional turnover was likely signalling a shift from contamination to ecological harm consistent with GESAMP's definition of pollution (Bryan and Hummerstone, 1971, Grant, 2010). Consequently, multi-domain monitoring (bacteria, archaea and eukaryotes) can improve early detection of pollution thresholds, informing targeted remediation. Researchers should, however, consider local confounding variables such as salinity, organic content and grain size that may mask moderate metal impacts or amplify them under certain redox conditions (Eggleton and Thomas, 2004, Chen et al., 2022).

7.9 Limitations and Caveats

Despite the advantages of multi-locus metabarcoding, a few limitations remain. Reference libraries for certain archaeal phyla and marine eukaryotes were still sparse (Bik et al., 2012, Krzmarzick et al., 2018), raising the risk of unassigned reads or underestimation of specific taxa. DNA-based approaches also conflate active, dormant and dead organisms, potentially overstating ecological presence if turnover rates were high (Taberlet et al., 2012). In large-scale

datasets spanning multiple estuaries and overlapping gradients can obscure moderate pollution signals (Zou et al., 2020, Pawlowski et al., 2024). Moreover, morphological or functional endpoints (e.g., feeding rates, growth assays) still play a valuable role in confirming that compositional shifts reflect genuine ecological harm rather than mere presence of metals (Bryan and Gibbs, 1983, Shipp and Grant, 2006). Finally, the cost and logistical burden of multi-marker sequencing may challenge routine regulatory monitoring, although these constraints continue to diminish as sequencing technologies evolve.

7.10 Directions for Future Research

Further methodological refinements could enhance pollution detection. Primer sets optimised for lesser-studied taxa, such as certain nematodes or archaeal sub-clades, would reduce coverage gaps (Bahram et al., 2019, Kounosu et al., 2019). Shotgun metagenomics and metatranscriptomics can profile the full complement of metal-resistance genes and their transcriptional activity in situ (Quince et al., 2017), while emerging sediment metaproteomics is beginning to uncover which detoxification proteins are actually expressed under high-Cu stress (Wilmes and Bond, 2006, Gracioso et al., 2014). Complementary, low-cost functional screens could be achieved with targeted qPCR assays for well-characterised copper-efflux genes (Navarro et al., 2009). Functional analyses (metatranscriptomics, metagenomics) offer a direct window into metabolic pathways under metal stress, clarifying whether tolerant lineages truly thrive or merely persist (Turner et al., 2013). Further microcosm and mesocosm experiments could isolate copper's role from confounding variables, illuminating threshold responses in controlled environments (Ogilvie and Grant, 2008). Integrating morphological data, particularly for nematodes and other meiobenthos would validate how well DNA-based approaches capture metal-induced changes, especially for extremely localised or patchily distributed taxa (Somerfield et al., 1994c, Warwick, 2001). Finally, bridging advanced geochemical models (e.g., AVS, equilibrium partitioning) with multi-domain metabarcoding can pinpoint the fraction of metals most responsible for ecological disruptions, offering more precise guidelines for remediation and management (Di Toro et al., 1991, Hall Jr and Anderson, 2022).

7.11 Final Remarks

In southwestern England, the legacy of copper contamination is long-standing and will continue to shape benthic communities for decades. This research has shown that multi-target metabarcoding is a powerful tool for detecting subtle yet meaningful shifts across bacteria, archaea and eukaryotes, revealing both domain-specific sensitivities and shared threshold responses. Key findings indicate that while moderate copper loads may remain partially masked by environmental heterogeneity, severe contamination (>100 µg/L porewater Cu) consistently alters community composition, underscoring a pollution threshold in line with GESAMP's criteria for ecological harm (Bryan and Hummerstone, 1971, Grant, 2010). By refining methodologies, expanding reference databases and incorporating morphological or functional validation, future studies can deepen our understanding of how metals, salinity and other stressors intersect to shape benthic ecology. A fuller understanding will also require linking taxonomic shifts to functional roles, including processes such as nutrient cycling or metal detoxification, through metagenomic or transcriptomic approaches. Ultimately, the insights gained here on threshold detection, domain-specific responses and the utility of multi-locus metabarcoding advance both scientific insight and practical strategies for coastal pollution monitoring, guiding more informed and effective environmental management.

References

ABED, R. M., AL-KINDI, S. & AL-KHARUSI, S. 2015. Diversity of bacterial communities along a petroleum contamination gradient in desert soils. *Microbial ecology,* 69**,** 95-105.

AGNIHOTRI, R., GUJRE, N., MITRA, S. & SHARMA, M. P. 2023. Decoding the PLFA profiling of microbial community structure in soils contaminated with municipal solid wastes. *Environmental Research,* 219**,** 114993.

ALLOWAY, B. J. 2012. *Heavy metals in soils: trace metals and metalloids in soils and their bioavailability*, Springer Science & Business Media.

AMARAL-ZETTLER, L. A., MCCLIMENT, E. A., DUCKLOW, H. W. & HUSE, S. M. 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PloS one,* 4**,** e6372.

ANTIL, S., ABRAHAM, J. S., SRIPOORNA, S., MAURYA, S., DAGAR, J., MAKHIJA, S., BHAGAT, P., GUPTA, R., SOOD, U., LAL, R. & TOTEJA, R. 2023. DNA barcoding, an effective tool for species identification: a review. *Molecular Biology Reports,* 50**,** 761-775.

ARDILA, P. A. R., ALONSO, R. Á., VALSERO, J. J. D., GARCÍA, R. M., CABRERA, F. Á., COSÍO, E. L. & LAFORET, S. D. 2023. Assessment of heavy metal pollution in marine sediments from southwest of Mallorca island, Spain. *Environmental Science and Pollution Research,* 30**,** 16852-16866.

ARRHENIUS, O. 1921. Species and area. *Journal of Ecology,* 9**,** 95-99.

ARUNACHALAM, K., SASIDHARAN, S. P., ARUNACHALAM, K. & SASIDHARAN, S. P. 2021. Gel electrophoresis and PCR amplification. *Bioassays in Experimental and Preclinical Pharmacology***,** 241-258.

AUSTEN, M. & SOMERFIELD, P. 1997. A community level sediment bioassay applied to an estuarine heavy metal gradient. *Marine Environmental Research,* 43**,** 315-328.

AYLAGAS, E., BORJA, Á. & RODRÍGUEZ-EZPELETA, N. 2014. Environmental status assessment using DNA metabarcoding: towards a genetics based marine biotic index (gAMBI). *PloS one,* 9**,** e90529.

BÅÅTH, E. 1989. Effects of heavy metals in soil on microbial processes and populations (a review). *Water, Air, and Soil Pollution,* 47**,** 335-379.

BADOTTI, F., DE OLIVEIRA, F. S., GARCIA, C. F., VAZ, A. B. M., FONSECA, P. L. C., NAHUM, L. A., OLIVEIRA, G. & GÓES-NETO, A. 2017. Effectiveness of ITS and

sub-regions as DNA barcode markers for the identification of Basidiomycota (Fungi). *BMC microbiology,* 17**,** 1-12.

BAETSCHER, D. S., LOCATELLI, N. S., WON, E., FITZGERALD, T., MCINTYRE, P. B. & THERKILDSEN, N. O. 2023. Optimizing a metabarcoding marker portfolio for species detection from complex mixtures of globally diverse fishes. *Environmental DNA,* 5**,** 1589-1607.

BAHRAM, M., ANSLAN, S., HILDEBRAND, F., BORK, P. & TEDERSOO, L. 2019. Newly designed 16S rRNA metabarcoding primers amplify diverse and novel archaeal taxa from the environment. *Environmental Microbiology Reports,* 11**,** 487-494.

BAILLY, J., FRAISSINET-TACHET, L., VERNER, M.-C., DEBAUD, J.-C., LEMAIRE, M., WÉSOLOWSKI-LOUVEL, M. & MARMEISSE, R. 2007. Soil eukaryotic functional diversity, a metatranscriptomic approach. *The ISME journal,* 1**,** 632-642.

BALSAMO, M., ARTOIS, T., SMITH III, J. P., TODARO, M. A., GUIDI, L., LEANDER, B. S. & VAN STEENKISTE, N. W. 2020. The curious and neglected soft-bodied meiofauna: Rouphozoa (Gastrotricha and Platyhelminthes). *Hydrobiologia,* 847**,** 2613-2644.

BARNES, M. A. & TURNER, C. R. 2016. The ecology of environmental DNA and implications for conservation genetics. *Conservation genetics,* 17**,** 1-17.

BARRENECHEA ANGELES, I., NGUYEN, N.-L., GRECO, M., TAN, K. S. & PAWLOWSKI, J. 2024. Assigning the unassigned: A signature-based classification of rDNA metabarcodes reveals new deep-sea diversity. *Plos one,* 19**,** e0298440.

BARTON, D. B. 1961. A history of copper mining in Cornwall and Devon. *(No Title).*

BEALE, D., CROSSWELL, J., KARPE, A., METCALFE, S., MORRISON, P., STALEY, C., AHMED, W., SADOWSKY, M., PALOMBO, E. & STEVEN, A. 2018. Seasonal metabolic analysis of marine sediments collected from Moreton Bay in South East Queensland, Australia, using a multi-omics-based approach. *Science of the Total Environment,* 631**,** 1328-1341.

BÉGUINOT, J. 2016. On general mathematical constraints applying to the kinetics of species discovery during progressive sampling: consequences on the theoretical expression of the Species Accumulation Curve. *Advances in Research,* 8**,** 1-17.

BERRY, D., MAHFOUDH, K. B., WAGNER, M. & LOY, A. 2011. Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and environmental microbiology,* 77**,** 7846-7849.

BESAURY, L., GHIGLIONE, J. F. & QUILLET, L. 2014. Abundance, activity, and diversity of archaeal and bacterial communities in both uncontaminated and highly copper-contaminated marine sediments. *Mar Biotechnol (NY),* 16**,** 230-42.

BESZTERI, B., TEMPERTON, B., FRICKENHAUS, S. & GIOVANNONI, S. J. 2010. Average genome size: a potential source of bias in comparative metagenomics. *The ISME journal,* 4**,** 1075-1077.

BIK, H. M., PORAZINSKA, D. L., CREER, S., CAPORASO, J. G., KNIGHT, R. & THOMAS, W. K. 2012. Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution,* 27**,** 233-243.

BIRRER, S. C., WEMHEUER, F., DAFFORN, K. A., GRIBBEN, P. E., STEINBERG, P. D., SIMPSON, S. L., POTTS, J., SCANES, P., DOBLIN, M. A. & JOHNSTON, E. L. 2021. Legacy Metal Contaminants and Excess Nutrients in Low Flow Estuarine Embayments Alter Composition and Function of Benthic Bacterial Communities. *Frontiers in Microbiology,* 12.

BLANCK, H. 2002. A Critical Review of Procedures and Approaches Used for Assessing Pollution-Induced Community Tolerance (PICT) in Biotic Communities. *Human and Ecological Risk Assessment: An International Journal,* 8**,** 1003-1034.

BLANCK, H., WÄNGBERG, S.-Å. & MOLANDER, S. 1988. Pollution-induced community tolerance—a new ecotoxicological tool. *Functional testing of aquatic biota for estimating hazards of chemicals.* ASTM International.

BLAXTER, M. L., DE LEY, P., GAREY, J. R., LIU, L. X., SCHELDEMAN, P., VIERSTRAETE, A., VANFLETEREN, J. R., MACKEY, L. Y., DORRIS, M., FRISSE, L. M., VIDA, J. T. & THOMAS, W. K. 1998. A molecular evolutionary framework for the phylum Nematoda. *Nature,* 392**,** 71-75.

BORJA, A. & DAUER, D. M. 2008. Assessing the environmental quality status in estuarine and coastal systems: comparing methodologies and indices. *Ecological indicators,* 8**,** 331-337.

BRANNOCK, P. M. & HALANYCH, K. M. 2015. Meiofaunal community analysis by high-throughput sequencing: Comparison of extraction, quality filtering, and clustering methods. *Marine Genomics,* 23**,** 67-75.

BRANNOCK, P. M., ORTMANN, A. C., MOSS, A. G. & HALANYCH, K. M. 2016. Metabarcoding reveals environmental factors influencing spatio-temporal variation in pelagic micro-eukaryotes. *Molecular Ecology,* 25**,** 3593-3604.

BRAY, J. R. & CURTIS, J. T. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecological monographs,* 27**,** 326-349.

BRYAN, G. 1971. The effects of heavy metals (other than mercury) on marine and estuarine organisms. *Proceedings of the Royal Society of London. Series B. Biological Sciences,* 177**,** 389-410.

BRYAN, G., GIBBS, P., HUMMERSTONE, L. & BURT, G. 1987. Copper, zinc, and organotin as long-term factors governing the distribution of organisms in the Fal Estuary in southwest England. *Estuaries,* 10**,** 208-219.

BRYAN, G. & GIBBS, P. E. 1983. Heavy metals in the Fal Estuary, Cornwall: a study of long-term contamination by mining waste and its effects on estuarine organisms.

BRYAN, G. & HUMMERSTONE, L. 1971. Adaptation of the polychaete Nereis diversicolor to estuarine sediments containing high concentrations of heavy metals. I. General

observations and adaptation to copper. *Journal of the Marine Biological Association of the United Kingdom,* 51**,** 845-863.

BRYAN, G. & LANGSTON, W. 1992. Bioavailability, accumulation and effects of heavy metals in sediments with special reference to United Kingdom estuaries: a review. *Environmental pollution,* 76**,** 89-131.

BUCKLIN, A., LINDEQUE, P. K., RODRIGUEZ-EZPELETA, N., ALBAINA, A. & LEHTINIEMI, M. 2016. Metabarcoding of marine zooplankton: prospects, progress and pitfalls. *Journal of Plankton Research,* 38**,** 393-400.

BURKI, F., ROGER, A. J., BROWN, M. W. & SIMPSON, A. G. 2020. The new tree of eukaryotes. *Trends in ecology & evolution,* 35**,** 43-55.

BURKI, F., SANDIN, M. M. & JAMY, M. 2021. Diversity and ecology of protists revealed by metabarcoding. *Current Biology,* 31**,** R1267-R1280.

BURTON, G. A. 2018. *Sediment toxicity assessment,* CRC Press.

CAHILL, A. E., PEARMAN, J. K., BORJA, A., CARUGATI, L., CARVALHO, S., DANOVARO, R., DASHFIELD, S., DAVID, R., FÉRAL, J. P. & OLENIN, S. 2018. A comparative analysis of metabarcoding and morphology-based identification of benthic communities across different regional seas. *Ecology and Evolution,* 8**,** 8908-8920.

CAI, M., RICHTER-HEITMANN, T., YIN, X., HUANG, W.-C., YANG, Y., ZHANG, C., DUAN, C., PAN, J., LIU, Y. & LIU, Y. 2021. Ecological features and global distribution of Asgard archaea. *Science of The Total Environment,* 758**,** 143581.

CALLAHAN, B. J., MCMURDIE, P. J. & HOLMES, S. P. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal,* 11**,** 2639-2643.

CAO, Z., LI, P. & LI, Z.-H. 2021. A latest review on the application of microcosm model in environmental research. *Environmental Science and Pollution Research,* 28**,** 60438-60447.

CAPORASO, J. G., LAUBER, C. L., WALTERS, W. A., BERG-LYONS, D., HUNTLEY, J., FIERER, N., OWENS, S. M., BETLEY, J., FRASER, L. & BAUER, M. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME journal,* 6**,** 1621-1624.

CAPORASO, J. G., LAUBER, C. L., WALTERS, W. A., BERG-LYONS, D., LOZUPONE, C. A., TURNBAUGH, P. J., FIERER, N. & KNIGHT, R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the national academy of sciences,* 108**,** 4516-4522.

CAREW, M. E., KELLAR, C. R., PETTIGROVE, V. J. & HOFFMANN, A. A. 2018. Can high-throughput sequencing detect macroinvertebrate diversity for routine monitoring of an urban river? *Ecological indicators,* 85**,** 440-450.

CARLIER, J. D., ETTAMIMI, S., COX, C. J., HAMMANI, K., GHAZAL, H. & COSTA, M. C. 2020. Prokaryotic diversity in stream sediments affected by acid mine drainage. *Extremophiles,* 24**,** 809-819.

CHAPMAN, P. M. & WANG, F. 2001. Assessing sediment contamination in estuaries. *Environmental Toxicology and Chemistry: An International Journal,* 20**,** 3-22.

CHAPMAN, P. M., WANG, F., JANSSEN, C., PERSOONE, G. & ALLEN, H. E. 1998. Ecotoxicology of metals in aquatic sediments: binding and release, bioavailability, risk assessment, and remediation. *Canadian Journal of Fisheries and Aquatic Sciences,* 55**,** 2221-2243.

CHAUDHARY, A. & SINGH, H. S. 2013. Secondary structure and phylogenetic utility of the ribosomal large subunit (28S) in monogeneans of the genus Thaparocleidus and Bifurcohaptor (Monogenea: Dactylogyridae). *Journal of parasitic diseases,* 37**,** 74-83.

CHEN, H., JI, C., HU, H., HU, S., YUE, S. & ZHAO, M. 2022. Bacterial community response to chronic heavy metal contamination in marine sediments of the East China Sea. *Environmental Pollution,* 307**,** 119280.

CHEN, J., MCILROY, S. E., ARCHANA, A., BAKER, D. M. & PANAGIOTOU, G. 2019. A pollution gradient contributes to the taxonomic, functional, and resistome diversity of microbial communities in marine sediments. *Microbiome,* 7**,** 104.

CHEN, M., LI, W., TENG, H., HU, W., DONG, Z., ZHANG, D., LIU, T. & ZHENG, Q. 2024. Impact of Combined Pollution of Ciprofloxacin and Copper on the Diversity of Archaeal Communities and Antibiotic-Resistance Genes. *Antibiotics,* 13**,** 734.

CHITTAVICHAI, T., SATHITNAITHAM, S., UTTHIYA, S., PROMPICHAI, W., PROMMARIT, K., VUTTIPONGCHAIKIJ, S. & WONNAPINIJ, P. 2025. Limitations of 18S rDNA Sequence in Species-Level Classification of Dictyostelids. *Microorganisms,* 13**,** 275.

CHOI, J. & PARK, J. S. 2020. Comparative analyses of the V4 and V9 regions of 18S rDNA for the extant eukaryotic community using the Illumina platform. *Scientific Reports,* 10**,** 6519.

CHRISMAS, N., ALLEN, R., ALLEN, M. J., BIRD, K. & CUNLIFFE, M. 2023. A 17-year time-series of fungal environmental DNA from a coastal marine ecosystem reveals long-term seasonal-scale and inter-annual diversity patterns. *Proceedings of the Royal Society B: Biological Sciences,* 290**,** 20222129.

CLARKE, K. & AINSWORTH, M. 1993. A method of linking multivariate community structure to environmental variables. *Marine Ecology-Progress Series,* 92**,** 205-205.

CLARKE, K. R. 1990a. Comparisons of dominance curves. *Journal of experimental marine biology and ecology,* 138**,** 143-157.

CLARKE, K. R. 1990b. Comparisons of dominance curves. *Journal of Experimental Marine Biology and Ecology,* 138**,** 143-157.

CLARKE, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. *Australian journal of ecology,* 18**,** 117-143.

CLARKE, K. R., GORLEY, R. N., SOMERFIELD, P. & WARWICK, R. M. 2014. Change in marine communities: an approach to statistical analysis and interpretation.

COLE, J. R., CHAI, B., FARRIS, R. J., WANG, Q., KULAM, S., MCGARRELL, D. M., GARRITY, G. M. & TIEDJE, J. M. 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic acids research,* 33**,** D294-D296.

COPPO, G., PAIS, F. S., FERREIRA, T. O., HALANYCH, K. M., DONNELLY, K., MAZZUCO, A. C. & BERNARDINO, A. F. 2023. Transition of an estuarine benthic meiofauna assemblage 1.7 and 2.8 years after a mining disaster. *PeerJ,* 11**,** e14992.

CORCOLL, N., YANG, J., BACKHAUS, T., ZHANG, X. & ERIKSSON, K. M. 2019. Copper Affects Composition and Functioning of Microbial Communities in Marine Biofilms at Environmentally Relevant Concentrations. *Frontiers in Microbiology,* 9.

COULL, B. 1992. Pollution and meiofauna: field, laboratory, and mesocosm studies. *Oceanogr. mar. biol. ann. Rev.,* 30**,** 191-271.

COULL, B. C. 1999. Role of meiofauna in estuarine soft-bottom habitats. *Australian Journal of Ecology,* 24**,** 327-343.

COX, M. J., COOKSON, W. O. C. M. & MOFFATT, M. F. 2013. Sequencing the human microbiome in health and disease. *Human Molecular Genetics,* 22**,** R88-R94.

CREER, S., DEINER, K., FREY, S., PORAZINSKA, D., TABERLET, P., THOMAS, W. K., POTTER, C. & BIK, H. M. 2016. The ecologist's field guide to sequence-based identification of biodiversity. *Methods in ecology and evolution,* 7**,** 1008-1018.

CRIBDON, B., WARE, R., SMITH, O., GAFFNEY, V. & ALLABY, R. G. 2020. PIA: more accurate taxonomic assignment of metagenomic data demonstrated on sedaDNA from the North Sea. *Frontiers in Ecology and Evolution,* 8**,** 84.

DAUVIN, J.-C. 2007. Paradox of estuarine quality: benthic indicators and indices, consensus or debate for the future. *Marine Pollution Bulletin,* 55**,** 271-281.

DE LUCA, D., PIREDDA, R., SARNO, D. & KOOISTRA, W. H. 2021. Resolving cryptic species complexes in marine protists: phylogenetic haplotype networks meet global DNA metabarcoding datasets. *The ISME journal,* 15**,** 1931-1942.

DEAGLE, B. E., KIRKWOOD, R. & JARMAN, S. N. 2009. Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular ecology,* 18**,** 2022-2038.

DEAGLE, B. E., THOMAS, A. C., MCINNES, J. C., CLARKE, L. J., VESTERINEN, E. J., CLARE, E. L., KARTZINEL, T. R. & EVESON, J. P. 2019. Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular ecology,* 28**,** 391-406.

DEINER, K., WALSER, J.-C., MÄCHLER, E. & ALTERMATT, F. 2015. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological conservation,* 183**,** 53-63.

DELAGE, E., HENRY, N., SALAZAR, G., RUSCHEWEYH, H.-J., POULAIN, J., SUNAGAWA, S., DE VARGAS, C. & CHAFFRON, S. 2023. Tara Oceans (2009-2013) rDNA 18S V4 ASV table (dada2).

DELMONT, T. O., ROBE, P., CLARK, I., SIMONET, P. & VOGEL, T. M. 2011. Metagenomic comparison of direct and indirect soil DNA extraction approaches. *Journal of microbiological methods,* 86**,** 397-400.

DEN BULCKE LAURE, V., ANNELIES, D. B., HANS, H., SARA, M., STEPHIE, S., WILLEM, W., JAN, W., KRIS, H. & SOFIE, D. 2024. Comparative study of traditional and DNA-based methods for environmental impact assessment: A case study of marine aggregate extraction in the North Sea. *Science of The Total Environment,* 946**,** 174106.

DENG, S., KE, T., LI, L., CAI, S., ZHOU, Y., LIU, Y., GUO, L., CHEN, L. & ZHANG, D. 2018. Impacts of environmental factors on the whole microbial communities in the rhizosphere of a metal-tolerant plant: Elsholtzia haichowensis Sun. *Environmental Pollution,* 237**,** 1088-1097.

DERYCKE, S., BACKELJAU, T., VLAEMINCK, C., VIERSTRAETE, A., VANFLETEREN, J., VINCX, M. & MOENS, T. 2007. Spatiotemporal analysis of population genetic structure in Geomonhystera disjuncta (Nematoda, Monhysteridae) reveals high levels of molecular diversity. *Marine Biology,* 151**,** 1799-1812.

DERYCKE, S., DE LEY, P., TANDINGAN DE LEY, I., HOLOVACHOV, O., RIGAUX, A. & MOENS, T. 2010. Linking DNA sequences to morphology: cryptic diversity and population genetic structure in the marine nematode Thoracostoma trachygaster (Nematoda, Leptosomatidae). *Zoologica Scripta,* 39**,** 276-289.

DERYCKE, S., REMERIE, T., VIERSTRAETE, A., BACKELJAU, T., VANFLETEREN, J., VINCX, M. & MOENS, T. 2005. Mitochondrial DNA variation and cryptic speciation within the free-living marine nematode Pellioditis marina. *Marine Ecology Progress Series,* 300**,** 91-103.

DEWI, F. R. P., KIM, H.-W., KIM, A. R., LEE, S. R., THEN, A. Y.-H., ZAMRONI, M., PALIMIRMO, F. S., WAHYUNI, H. T. & AMIN, M. H. F. A. 2024. Exploring fish communities in the mud volcano polluted river using environmental DNA metabarcoding. *Environmental Advances,* 16**,** 100534.

DI CESARE, A., PJEVAC, P., ECKERT, E., CURKOV, N., MIKO ŠPARICA, M., CORNO, G. & ORLIĆ, S. 2020. The role of metal contamination in shaping microbial communities in heavily polluted marine sediments. *Environmental Pollution,* 265**,** 114823.

DI TORO, D. M., ZARBA, C. S., HANSEN, D. J., BERRY, W. J., SWARTZ, R. C., COWAN, C. E., PAVLOU, S. P., ALLEN, H. E., THOMAS, N. A. & PAQUIN, P. R. 1991. Technical basis for establishing sediment quality criteria for nonionic organic

chemicals using equilibrium partitioning. *Environmental Toxicology and Chemistry: An International Journal,* 10**,** 1541-1583.

DU, M., ZHENG, M., LIU, A., WANG, L., PAN, X., LIU, J. & RAN, X. 2022. Effects of emerging contaminants and heavy metals on variation in bacterial communities in estuarine sediments. *Science of The Total Environment,* 832**,** 155118.

EATON, J. W. & MOSS, B. 1966. The estimation of numbers and pigment content in epipelic algal populations. *Limnology and Oceanography,* 11**,** 584-595.

EDGAR, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics,* 26**,** 2460-2461.

EDGAR, R. C. 2017. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *bioRxiv***,** 192211.

EDGAR, R. C. 2018. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics,* 34**,** 2371-2375.

EGGLETON, J. & THOMAS, K. V. 2004. A review of factors affecting the release and bioavailability of contaminants during sediment disturbance events. *Environment international,* 30**,** 973-980.

EL KHAWAND, M., CROMBIE, A. T., JOHNSTON, A., VAVLLINE, D. V., MCAULIFFE, J. C., LATONE, J. A., PRIMAK, Y. A., LEE, S. K., WHITED, G. M. & MCGENITY, T. J. 2016. Isolation of isoprene degrading bacteria from soils, development of isoA gene probes and identification of the active isoprene-degrading soil community using DNA-stable isotope probing. *Environmental microbiology,* 18**,** 2743-2753.

ELBADRI, G. A. A., DE LEY, P., WAEYENBERGE, L., VIERSTRAETE, A., MOENS, M. & VANFLETEREN, J. 2002. Intraspecific variation in Radopholus similis isolates assessed with restriction fragment length polymorphism and DNA sequencing of the internal transcribed spacer region of the ribosomal RNA cistron. *International Journal for Parasitology,* 32**,** 199-205.

ELBRECHT, V. & LEESE, F. 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PloS one,* 10**,** e0130324.

ELBRECHT, V., PEINERT, B. & LEESE, F. 2017. Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecology and evolution,* 7**,** 6918-6926.

ELOE-FADROSH, E. A., IVANOVA, N. N., WOYKE, T. & KYRPIDES, N. C. 2016. Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nature Microbiology,* 1**,** 1-4.

EMILSON, C. E., THOMPSON, D. G., VENIER, L. A., PORTER, T. M., SWYSTUN, T., CHARTRAND, D., CAPELL, S. & HAJIBABAEI, M. 2017. DNA metabarcoding and morphological macroinvertebrate metrics reveal the same changes in boreal watersheds across an environmental gradient. *Scientific reports,* 7**,** 1-11.

EMMONS, S. C., COMPSON, Z. G., MALISH, M. C., BUSCH, M. H., SAENZ, V., HIGGINS, K. T. & ALLEN, D. C. 2023. DNA metabarcoding captures different macroinvertebrate biodiversity than morphological identification approaches across a continental scale. *Environmental DNA*.

ENDLER, R. 2009. Sediment physical properties of the DYNAS study area. *Journal of Marine Systems,* 75**,** 317-329.

EULER, S., JEFFREY, L. C., MAHER, D. T., MACKENZIE, D. & TAIT, D. R. 2020. Shifts in methanogenic archaea communities and methane dynamics along a subtropical estuarine land use gradient. *PLoS One,* 15**,** e0242339.

FADROSH, D. W., MA, B., GAJER, P., SENGAMALAY, N., OTT, S., BROTMAN, R. M. & RAVEL, J. 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome,* 2**,** 1-7.

FAIS, M., BELLISARIO, B., DUARTE, S., VIEIRA, P. E., SOUSA, R., CANCHAYA, C. & COSTA, F. O. 2020. Meiofauna metabarcoding in Lima estuary (Portugal) suggests high taxon replacement within a background of network stability. *Regional Studies in Marine Science,* 38**,** 101341.

FANG, X., PENG, B., WANG, X., SONG, Z., ZHOU, D., WANG, Q., QIN, Z. & TAN, C. 2019. Distribution, contamination and source identification of heavy metals in bed sediments from the lower reaches of the Xiangjiang River in Hunan province, China. *Science of The Total Environment,* 689**,** 557-570.

FÉLIX, M.-A., BRAENDLE, C. & CUTTER, A. D. 2014. A Streamlined System for Species Diagnosis in Caenorhabditis (Nematoda: Rhabditidae) with Name Designations for 15 Distinct Biological Species. *PLOS ONE,* 9**,** e94723.

FICETOLA, G. F., BOYER, F., VALENTINI, A., BONIN, A., MEYER, A., DEJEAN, T., GABORIAUD, C., USSEGLIO-POLATERA, P. & TABERLET, P. 2021. Comparison of markers for the monitoring of freshwater benthic biodiversity through DNA metabarcoding. *Molecular Ecology,* 30**,** 3189-3202.

FOLK, R. L. & WARD, W. C. 1957. Brazos River bar [Texas]; a study in the significance of grain size parameters. *Journal of sedimentary research,* 27**,** 3-26.

FONSECA, V. G. 2018. Pitfalls in relative abundance estimation using eDNA metabarcoding. Wiley Online Library.

FONSECA, V. G., KIRSE, A., GIEBNER, H., VAUSE, B. J., DRAGO, T., POWER, D. M., PECK, L. S. & CLARK, M. S. 2022. Metabarcoding the Antarctic Peninsula biodiversity using a multi-gene approach. *ISME Communications,* 2.

FOX, J. & WEISBERG, S. 2019. Package 'car'. Companion to Applied Regression, version, 3.1-3. Available September 27, 2024 at https://cran.r-project.org/web/packages/car/car.pdf. *SAGE*.

GAONKAR, C. C. & CAMPBELL, L. 2024. A full-length 18S ribosomal DNA metabarcoding approach for determining protist community diversity using Nanopore sequencing. *Ecology and Evolution,* 14**,** e11232.

GATTONI, K., GENDRON, E. M. S., SANDOVAL-RUIZ, R., BORGEMEIER, A., MCQUEEN, J. P., R, M. S., SLOS, D., T, O. P. & D, L. P. 2023. 18S-NemaBase: Curated 18S rRNA Database of Nematode Sequences. *J Nematol,* 55**,** 20230006.

GELLER, J., MEYER, C., PARKER, M. & HAWK, H. 2013. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular ecology resources,* 13**,** 851-861.

GERALDI, N. R., KRAUSE-JENSEN, D., ØRBERG, S. B., FRÜHE, L., SEJR, M. K., HANSEN, J. L., LUND-HANSEN, L. & DUARTE, C. M. 2024. Environmental drivers of Arctic communities based on metabarcoding of marine sediment eDNA. *Proceedings of the Royal Society B,* 291**,** 20231614.

GIELINGS, R., FAIS, M., FONTANETO, D., CREER, S., COSTA, F. O., RENEMA, W. & MACHER, J.-N. 2021. DNA Metabarcoding Methods for the Study of Marine Benthic Meiofauna: A Review. *Frontiers in Marine Science,* 8.

GIERE, O. 2009. The microscopic motile fauna of aquatic sediments. Meiobenthology. Springer-Berlin/Heidelberg.

GILBERT, M. T. P., HANSEN, A. J., WILLERSLEV, E., TURNER-WALKER, G. & COLLINS, M. 2006. Insights into the processes behind the contamination of degraded human teeth and bone samples with exogenous sources of DNA. *International Journal of Osteoarchaeology,* 16**,** 156-164.

GILLAN, D. C., DANIS, B., PERNET, P., JOLY, G. & DUBOIS, P. 2005. Structure of sediment-associated microbial communities along a heavy-metal contamination gradient in the marine environment. *Appl Environ Microbiol,* 71**,** 679-90.

GOLDBERG, C. S., TURNER, C. R., DEINER, K., KLYMUS, K. E., THOMSEN, P. F., MURPHY, M. A., SPEAR, S. F., MCKEE, A., OYLER-MCCANCE, S. J. & CORNMAN, R. S. 2016. Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in ecology and evolution,* 7**,** 1299-1307.

GOŁĘBIEWSKI, M., DEJA-SIKORA, E., CICHOSZ, M., TRETYN, A. & WRÓBEL, B. 2014. 16S rDNA pyrosequencing analysis of bacterial community in heavy metals polluted soils. *Microbial ecology,* 67**,** 635-647.

GOSWAMI, A., ADKINS-JABLONSKY, S. J., FILHO, M. M. B., SHILLING, M. D., DAWSON, A., HEISER, S., O'CONNOR, A., WALKER, M., ROBERTS, Q. & MORRIS, J. J. 2023. Heavy Metal Pollution Impacts Soil Bacterial Community Structure and Antimicrobial Resistance at the Birmingham 35th Avenue Superfund Site. *Microbiology Spectrum,* 11**,** e02426-22.

GOTELLI, N. J. & CHAO, A. 2013. Measuring and estimating species richness, species diversity, and biotic similarity from sampling data.

GOTELLI, N. J. & COLWELL, R. K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters,* 4**,** 379-391.

GOUGH, H. L. & STAHL, D. A. 2011. Microbial community structures in anoxic freshwater lake sediment along a metal contamination gradient. *The ISME journal,* 5**,** 543-558.

GOWER, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika,* 53**,** 325-338.

GRACIOSO, L. H., AVANZI, I. R., GALLUZZI BALTAZAR, M. D. P., VEIGA, M. N., GIMENES, L. J., DO NASCIMENTO, C. A. O. & PERPETUO, E. A. 2014. Proteomic profiles comparison of three isolated bacteria strains from a copper processing area. *BMC Proceedings,* 8**,** P197.

GRANT, A. 2002. Pollution-tolerant species and communities: intriguing toys or invaluable monitoring tools? *Human and Ecological Risk Assessment,* 8**,** 955-970.

GRANT, A. 2010. Detecting ecological effects of pollutants in the aquatic environment. *Ecology of industrial pollution. 1st ed. Cambridge University Press, British Ecological Society***,** 147-61.

GRANT, A., ALEIDAN, A., DAVIES, C. S., UDOCHI, S. C., FRITSCHER, J., BAHRAM, M. & HILDEBRAND, F. 2023. Improved taxonomic annotation of Archaea communities using LotuS2, the Genome Taxonomy Database and RNAseq data. *bioRxiv***,** 2023.08.21.554127.

GRANT, A., HATELEY, J. G. & JONES, N. V. 1989. Mapping the ecological impact of heavy metals on the estuarine polychaete Nereis diversicolor using inherited metal tolerance. *Marine Pollution Bulletin,* 20**,** 235-238.

GRAY, J. S., CLARKE, K., WARWICK, R. & HOBBS, G. 1990. Detection of initial effects of pollution on marine benthos: an example from the Ekofisk and Eldfisk oilfields, North Sea. *Marine ecology progress series***,** 285-299.

GREENWOOD, N. 2001. *Metal speciation and toxicity in sediments from the Fal and Hayle estuaries, Cornwall, UK.* University of East Anglia.

GUBELIT, Y., POLYAK, Y., DEMBSKA, G., PAZIKOWSKA-SAPOTA, G., ZEGAROWSKI, L., KOCHURA, D., KRIVOROTOV, D., PODGORNAYA, E., BUROVA, O. & MAAZOUZI, C. 2016. Nutrient and metal pollution of the eastern Gulf of Finland coastline: sediments, macroalgae, microbiota. *Science of the Total Environment,* 550**,** 806-819.

GUPTA, A., SAHA, A. & SAR, P. 2021. Thermoplasmata and Nitrososphaeria as dominant archaeal members in acid mine drainage sediment of Malanjkhand Copper Project, India. *Arch Microbiol,* 203**,** 1833-1841.

HADZIAVDIC, K., LEKANG, K., LANZEN, A., JONASSEN, I., THOMPSON, E. M. & TROEDSSON, C. 2014. Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PloS one,* 9**,** e87624.

HAJIBABAEI, M., SHOKRALLA, S., ZHOU, X., SINGER, G. A. C. & BAIRD, D. J. 2011. Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos. *PLOS ONE,* 6**,** e17497.

HAJIHASSANI, A., SINGH, P. R. & GITONGA, D. 2023. The root-knot nematode Meloidogyne incognita infects passion fruit in the USA. *Australasian Plant Disease Notes,* 18.

HALL JR, L. W. & ANDERSON, R. D. 2022. Historical Global Review of Acid-Volatile Sulfide Sediment Monitoring Data. *Soil Systems,* 6**,** 71.

HAMA AZIZ, K. H., MUSTAFA, F. S., OMER, K. M., HAMA, S., HAMARAWF, R. F. & RAHMAN, K. O. 2023. Heavy metal pollution in the aquatic environment: efficient and low-cost removal approaches to eliminate their toxicity: a review. *RSC Adv,* 13**,** 17595-17610.

HAMADY, M., WALKER, J. J., HARRIS, J. K., GOLD, N. J. & KNIGHT, R. 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature methods,* 5**,** 235-237.

HARNELLY, E., KUSUMA, H. I., THOMY, Z. & SAMINGAN, S. 2022. Internal Transcribed Spacer (ITS) gene as an accurate DNA barcode for identification of macroscopic fungus in Aceh. *Biodiversitas Journal of Biological Diversity,* 23.

HE, Y., HE, L., WANG, Z., LIANG, T., SUN, S. & LIU, X. 2022. Salinity Shapes the Microbial Communities in Surface Sediments of Salt Lakes on the Tibetan Plateau, China. *Water,* 14**,** 4043.

HEIP, C., VINCX, M. & VRANKEN, G. 1985. The ecology of marine nematodes.

HELBLING, D. E., ACKERMANN, M., FENNER, K., KOHLER, H.-P. E. & JOHNSON, D. R. 2012. The activity level of a microbial community function can be predicted from its metatranscriptome. *The ISME journal,* 6**,** 902-904.

HERMANS, S. M., BUCKLEY, H. L. & LEAR, G. 2018. Optimal extraction methods for the simultaneous analysis of DNA from diverse organisms and sample types. *Molecular ecology resources,* 18**,** 557-569.

HOLLAWAY, S. L., FAW, G. M. & SIZEMORE, R. K. 1980. The bacterial community composition of an active oil field in the northwestern Gulf of Mexico. *Marine Pollution Bulletin,* 11**,** 153-156.

HOTELLING, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology,* 24**,** 417.

HUANG, L.-N., KUANG, J.-L. & SHU, W.-S. 2016. Microbial Ecology and Evolution in the Acid Mine Drainage Model System. *Trends in Microbiology,* 24**,** 581-593.

HUANG, W.-C., LIU, Y., ZHANG, X., ZHANG, C.-J., ZOU, D., ZHENG, S., XU, W., LUO, Z., LIU, F. & LI, M. 2021. Comparative genomic analysis reveals metabolic flexibility of Woesearchaeota. *Nature Communications,* 12**,** 5281.

HUGALL, A., STANTON, J. & MORITZ, C. 1999. Reticulate evolution and the origins of ribosomal internal transcribed spacer diversity in apomictic Meloidogyne. *Molecular Biology and Evolution,* 16**,** 157-164.

HUGERTH, L. W., MULLER, E. E. L., HU, Y. O. O., LEBRUN, L. A. M., ROUME, H., LUNDIN, D., WILMES, P. & ANDERSSON, A. F. 2014. Systematic Design of 18S rRNA Gene Primers for Determining Eukaryotic Diversity in Microbial Consortia. *PLOS ONE,* 9**,** e95567.

HUMMON, W. D. 1981. Extraction by sieving: a biased procedure in studies of stream meiobenthos. *Transactions of the American Microscopical Society***,** 278-284.

HURLBERT, S. H. 1971. The nonconcept of species diversity: a critique and alternative parameters. *Ecology,* 52**,** 577-586.

JACCARD, P. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat,* 37**,** 547-579.

JACKSON, T. A., VLAAR, S., NGUYEN, N. C., LEPPARD, G. G. & FINAN, T. M. 2015. Effects of Bioavailable Heavy Metal Species, Arsenic, and Acid Drainage from Mine Tailings on a Microbial Community Sampled Along a Pollution Gradient in a Freshwater Ecosystem. *Geomicrobiology Journal,* 32**,** 724 - 750.

JÄRUP, L. 2003. Hazards of heavy metal contamination. *British medical bulletin,* 68**,** 167-182.

JAY, Z. J., BEAM, J. P., KOZUBAL, M. A., JENNINGS, R. D., RUSCH, D. B. & INSKEEP, W. P. 2016. The distribution, diversity and function of predominant Thermoproteales in high-temperature environments of Yellowstone National Park. *Environmental microbiology,* 18**,** 4755-4769.

JIANG, J., ALDERISIO, K. A., SINGH, A. & XIAO, L. 2005. Development of procedures for direct extraction of Cryptosporidium DNA from water concentrates and for relief of PCR inhibitors. *Applied and environmental microbiology,* 71**,** 1135-1141.

JOHNSON, C. A. 1986. The regulation of trace element concentrations in river and estuarine waters contaminated with acid mine drainage: The adsorption of Cu and Zn on amorphous Fe oxyhydroxides. *Geochimica et Cosmochimica Acta,* 50**,** 2433-2438.

KALU, E., REYES-PRIETO, A. & BARBEAU, M. 2023. Community dynamics of microbial eukaryotes in intertidal mudflats in the hypertidal Bay of Fundy. ISME COMMUN, 3 (21).

KARST, S. M., DUEHOLM, M. S., MCILROY, S. J., KIRKEGAARD, R. H., NIELSEN, P. H. & ALBERTSEN, M. 2018. Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. *Nature biotechnology,* 36**,** 190-195.

KECK, F., BLACKMAN, R. C., BOSSART, R., BRANTSCHEN, J., COUTON, M., HÜRLEMANN, S., KIRSCHNER, D., LOCHER, N., ZHANG, H. & ALTERMATT, F. 2022. Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for biological community assessment. *Molecular Ecology,* 31**,** 1820-1835.

KEMBEL, S. W., WU, M., EISEN, J. A. & GREEN, J. L. 2012. Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance. *PLOS Computational Biology,* 8**,** e1002743.

KENDALL, M. A. & WIDDICOMBE, S. 1999. Small scale patterns in the structure of macrofaunal assemblages of shallow soft sediments. *Journal of Experimental Marine Biology and Ecology,* 237**,** 127-140.

KENNISH, M. J. 2002. Environmental threats and environmental future of estuaries. *Environmental conservation,* 29**,** 78-107.

KENNY, A. J. & SOTHERAN, I. 2013. Characterising the physical properties of seabed habitats. *Methods for the study of marine benthos***,** 47-95.

KIM, D. Y., COUNTWAY, P. D., YAMASHITA, W. & CARON, D. A. 2012. A combined sequence-based and fragment-based characterization of microbial eukaryote assemblages provides taxonomic context for the Terminal Restriction Fragment Length Polymorphism (T-RFLP) method. *Journal of microbiological methods,* 91**,** 527-536.

KIM, H., KIM, H., HWANG, H. S. & KIM, W. 2017. Metagenomic analysis of the marine coastal invertebrates of South Korea as assessed by Ilumina MiSeq. *Animal Cells and Systems,* 21**,** 37-44.

KIM, S.-Y., CHEON, S., PARK, C. & SOH, H. Y. 2025. Integrating DNA metabarcoding and morphological analysis improves marine zooplankton biodiversity assessment. *Scientific Reports,* 15**,** 7283.

KLUNDER, L., DUINEVELD, G. C. A., LAVALEYE, M. S. S., VAN DER VEER, H. W., PALSBØLL, P. J. & VAN BLEIJSWIJK, J. D. L. 2019. Diversity of Wadden Sea macrofauna and meiofauna communities highest in DNA from extractions preceded by cell lysis. *Journal of Sea Research,* 152**,** 101764.

KOEPPEL, A. F., WERTHEIM, J. O., BARONE, L., GENTILE, N., KRIZANC, D. & COHAN, F. M. 2013. Speedy speciation in a bacterial microcosm: new species can arise as frequently as adaptations within a species. *The ISME Journal,* 7**,** 1080-1091.

KORZHENKOV, A. A., TOSHCHAKOV, S. V., BARGIELA, R., GIBBARD, H., FERRER, M., TEPLYUK, A. V., JONES, D. L., KUBLANOV, I. V., GOLYSHIN, P. N. & GOLYSHINA, O. V. 2019. Archaea dominate the microbial community in an ecosystem with low-to-moderate temperature and extreme acidity. *Microbiome,* 7**,** 11.

KOUNOSU, A., MURASE, K., YOSHIDA, A., MARUYAMA, H. & KIKUCHI, T. 2019. Improved 18S and 28S rDNA primer sets for NGS-based parasite detection. *Scientific Reports,* 9**,** 15789.

KOZICH, J. J., WESTCOTT, S. L., BAXTER, N. T., HIGHLANDER, S. K. & SCHLOSS, P. D. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and environmental microbiology,* 79**,** 5112-5120.

KREHENWINKEL, H., WOLF, M., LIM, J. Y., ROMINGER, A. J., SIMISON, W. B. & GILLESPIE, R. G. 2017. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific reports,* 7**,** 17668.

KRUSKAL, J. B. 1964. Nonmetric multidimensional scaling: a numerical method. *Psychometrika,* 29**,** 115-129.

KRYCH, Ł., CASTRO-MEJÍA, J. L., FORERO-JUNCO, L. M., MOESBY, D. N., MIKKELSEN, M. B., RASMUSSEN, M. A., SYKULSKI, M. & NIELSEN, D. S. 2019. DNA enrichment and tagmentation method for species-level identification and strain-level differentiation using ON-rep-seq. *Communications biology,* 2**,** 1-10.

KRZMARZICK, M. J., TAYLOR, D. K., FU, X. & MCCUTCHAN, A. L. 2018. Diversity and niche of archaea in bioremediation. *Archaea,* 2018**,** 3194108.

LAFFERTY, K. 2024. Metabarcoding is (usually) more cost effective than seining or qPCR for detecting tidewater gobies and other estuarine fishes. *PeerJ,* 12**,** e16847.

LALLIAS, D., HIDDINK, J. G., FONSECA, V. G., GASPAR, J. M., SUNG, W., NEILL, S. P., BARNES, N., FERRERO, T., HALL, N., LAMBSHEAD, P. J. D., PACKER, M., THOMAS, W. K. & CREER, S. 2015. Environmental metabarcoding reveals heterogeneous drivers of microbial eukaryote diversity in contrasting estuarine ecosystems. *The ISME Journal,* 9**,** 1208-1221.

LAMB, P. D., HUNTER, E., PINNEGAR, J. K., CREER, S., DAVIES, R. G. & TAYLOR, M. I. 2019. How quantitative is metabarcoding: A meta-analytical approach. *Mol Ecol,* 28**,** 420-430.

LANE, D. 1991. 16S/23S sequencing. In: Stackebrant E, Goodfellow M (eds) Nucleic acid technologies in bacterial systematic. *Wiley, New York,* pp. 115–175.

LATZ, M. A. C., GRUJCIC, V., BRUGEL, S., LYCKEN, J., JOHN, U., KARLSON, B., ANDERSSON, A. & ANDERSSON, A. F. 2022. Short- and long-read metabarcoding of the eukaryotic rRNA operon: Evaluation of primers and comparison to shotgun metagenomics sequencing. *Molecular Ecology Resources,* 22**,** 2304-2318.

LAWES, J. C., NEILAN, B. A., BROWN, M. V., CLARK, G. F. & JOHNSTON, E. L. 2016. Elevated nutrients change bacterial community composition and connectivity: high throughput sequencing of young marine biofilms. *Biofouling,* 32**,** 57-69.

LEARY, D. H., LI, R. W., HAMDAN, L. J., HERVEY IV, W. J., LEBEDEV, N., WANG, Z., DESCHAMPS, J. R., KUSTERBECK, A. W. & VORA, G. J. 2014. Integrated metagenomic and metaproteomic analyses of marine biofilm communities. *Biofouling,* 30**,** 1211-1223.

LEE, P. Y., COSTUMBRADO, J., HSU, C.-Y. & KIM, Y. H. 2012. Agarose gel electrophoresis for the separation of DNA fragments. *JoVE (Journal of Visualized Experiments)***,** e3923.

LERAY, M., HO, S.-L., LIN, I.-J. & MACHIDA, R. J. 2018. MIDORI server: a webserver for taxonomic assignment of unknown metazoan mitochondrial-encoded sequences using a curated database. *Bioinformatics,* 34**,** 3753-3754.

LERAY, M. & KNOWLTON, N. 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences,* 112**,** 2076-2081.

LERAY, M. & KNOWLTON, N. 2017. Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. *PeerJ,* 5**,** e3006.

LERAY, M., YANG, J. Y., MEYER, C. P., MILLS, S. C., AGUDELO, N., RANWEZ, V., BOEHM, J. T. & MACHIDA, R. J. 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology,* 10**,** 34.

LI, S., LI, L., GAO, Q., DONG, S. & SHI, S. 2022. Deep-sea cage culture altered microbial community composition in the sediments of the Yellow Sea Cold Water Mass. *Marine Pollution Bulletin,* 183**,** 114081.

LI, Y., MA, J., SHEN, X., LI, X., ZHANG, R., NIU, Y. & CUI, B. 2024. Quantitative effects of anthropogenic and natural factors on the spatial distribution of heavy metals and bacterial communities in sediments of the Pearl River Delta. *Environmental Technology & Innovation,* 35**,** 103648.

LIN, Y., YE, Y., HU, Y. & SHI, H. 2019. The variation in microbial community structure under different heavy metal contamination levels in paddy soils. *Ecotoxicology and environmental safety,* 180**,** 557-564.

LIU, C., QI, R.-J., JIANG, J.-Z., ZHANG, M.-Q. & WANG, J.-Y. 2019. Development of a Blocking Primer to Inhibit the PCR Amplification of the 18S rDNA Sequences of Litopenaeus vannamei and Its Efficacy in Crassostrea hongkongensis. *Frontiers in Microbiology,* 10.

LIU, J.-J., WU, W.-X., DING, Y., SHI, D.-Z. & CHEN, Y.-X. 2010. Ammonia-oxidizing archaea and their important roles in nitrogen biogeochemical cycling: A review. *Yingyong Shengtai Xuebao,* 21.

LIU, J. & ZHANG, H. 2021. Combining multiple markers in environmental DNA metabarcoding to assess deep-sea benthic biodiversity. *Frontiers in Marine Science,* 8**,** 684955.

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

LIU, X., LI, M., CASTELLE, C. J., PROBST, A. J., ZHOU, Z., PAN, J., LIU, Y., BANFIELD, J. F. & GU, J. D. 2018. Insights into the ecology, evolution, and metabolism of the widespread Woesearchaeotal lineages. *Microbiome,* 6**,** 102.

LIU, X., WANG, Y. & GU, J.-D. 2021. Ecological distribution and potential roles of Woesearchaeota in anaerobic biogeochemical cycling unveiled by genomic analysis. *Computational and Structural Biotechnology Journal,* 19**,** 794-800.

LOSADA, P. M., TÜMMLER, B., WIEHLMANN, L. & CHOUVARINE, P. 2014. Whole metagenome shotgun sequencing analysis of microbiome of cystic fibrosis-and COPD patients. *European Respiratory Journal,* 44.

LOZUPONE, C. & KNIGHT, R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology,* 71**,** 8228-8235.

LUDDINGTON, I. A., KACZMARSKA, I. & LOVEJOY, C. 2012. Distance and character-based evaluation of the V4 region of the 18S rRNA gene for the identification of diatoms (Bacillariophyceae).

LUO, M., JI, Y., WARTON, D. & YU, D. W. 2023. Extracting abundance information from DNA-based data. *Molecular Ecology Resources,* 23**,** 174-189.

LUOMA, S. N. & RAINBOW, P. S. 2008. *Metal contamination in aquatic environments: science and lateral management*, Cambridge university press Cambridge.

MACHERIOTOU, L., GUILINI, K., BEZERRA, T. N., TYTGAT, B., NGUYEN, D. T., PHUONG NGUYEN, T. X., NOPPE, F., ARMENTEROS, M., BOUFAHJA, F., RIGAUX, A., VANREUSEL, A. & DERYCKE, S. 2019. Metabarcoding free-living marine nematodes using curated 18S and CO1 reference sequence databases for species-level taxonomic assignments. *Ecology and Evolution,* 9**,** 1211-1226.

MACHIDA, R. J. & KNOWLTON, N. 2012. PCR Primers for Metazoan Nuclear 18S and 28S Ribosomal DNA Sequences. *PLOS ONE,* 7**,** e46180.

MACKAY-ROBERTS, N., BOCK, C., LANNIG, G., LUCASSEN, M., PAUL, N., SCHAUM, E., POGODA, B. & GERDTS, G. 2024. A benthic mesocosm system for long-term multi-factorial experiments applying predicted warming and realistic microplastic pollution scenarios. *Limnology and Oceanography: Methods,* 22**,** 910-929.

MACKIEWICZ, P. & GAGAT, P. 2014. Monophyly of Archaeplastida supergroup and relationships among its lineages in the light of phylogenetic and phylogenomic studies. Are we close to a consensus? *Acta Societatis Botanicorum Poloniae,* 83.

MANI, K., TAIB, N., HUGONI, M., BRONNER, G., BRAGANÇA, J. M. & DEBROAS, D. 2020. Transient dynamics of archaea and bacteria in sediments and brine across a salinity gradient in a solar saltern of Goa, India. *Frontiers in microbiology,* 11**,** 1891.

MAOSA, J. O., WANG, S., LIU, S., LI, H., QING, X. & BERT, W. 2024. Exploring the use of metabarcoding to reveal eukaryotic associations with mononchids nematodes. *Journal of Nematology,* 56.

MARINCHEL, N., CASABIANCA, S., MARCHESINI, A., VERNESI, C., SCARDI, M. & PENNA, A. 2024. Structural variability of protist assemblages in surface sediments across Italian Mediterranean marine subregions. *Frontiers in Marine Science,* 11**,** 1427357.

MARTIN-POZAS, T., CUEZVA, S., FERNANDEZ-CORTES, A., BENAVENTE, D., SAIZ-JIMENEZ, C. & SANCHEZ-MORAL, S. 2023. Prokaryotic communities inhabiting a high-radon subterranean ecosystem (Castañar Cave, Spain): Environmental and substrate-driven controls. *Microbiological Research,* 277**,** 127511.

MARTIN, K. R., WAITS, L. P. & PARENT, C. E. 2021. Teaching an old shell new tricks: extracting DNA from current, historical, and ancient Mollusk shells. *Bioscience,* 71**,** 235-248.

MAZURKIEWICZ, M., PAWŁOWSKA, J., ANGELES, I. B., GRZELAK, K., DEJA, K., ZABORSKA, A., PAWŁOWSKI, J. & WŁODARSKA-KOWALCZUK, M. 2024. Sediment DNA metabarcoding and morphology provide complementary insight into macrofauna and meiobenthos response to environmental gradients in an Arctic glacial fjord. *Marine Environmental Research,* 198**,** 106552.

MCCLAIN, C. R., WEBB, T. J., HEIM, N. A., KNOPE, M. L., MONARREZ, P. M. & PAYNE, J. L. 2025. Size bias in the documentation of marine biodiversity. *Oikos,* 2025**,** e10828.

MCMURDIE, P. J. & HOLMES, S. 2013. Package 'phyloseq'. *gan,* 2.

MEDLIN, L., ELWOOD, H. J., STICKEL, S. & SOGIN, M. L. 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene,* 71**,** 491-499.

MILLER, D. C., JUMARS, P. A. & NOWELL, A. R. M. 1984. Effects of sediment transport on deposit feeding: Scaling arguments. *Limnology and Oceanography,* 29**,** 1202-1217.

MILLWARD, R. N. 1995. *The effects of chronic and acute metal-enrichment on the nematode community structure, composition and function in Restronguet Creek, S W England.* PhD thesis, University of East Anglia.

MILLWARD, R. N. & GRANT, A. 1995. Assessing the impact of copper on nematode communities from a chronically metal-enriched estuary using pollution-induced community tolerance. *Marine Pollution Bulletin,* 30**,** 701-706.

MILLWARD, R. N. & GRANT, A. 2000. Pollution-induced tolerance to copper of nematode communities in the severely contaminated restronguet creek and adjacent estuaries, Cornwall, United Kingdom. *Environmental Toxicology and Chemistry: An International Journal,* 19**,** 454-461.

MITCHELL, A., ROMANO, G. H., GROISMAN, B., YONA, A., DEKEL, E., KUPIEC, M., DAHAN, O. & PILPEL, Y. 2009. Adaptive prediction of environmental changes by microorganisms. *Nature,* 460**,** 220-224.

MITSI, K., ARROYO, A. S. & RUIZ-TRILLO, I. 2019. A global metabarcoding analysis expands molecular diversity of Platyhelminthes and reveals novel early-branching clades. *Biology Letters,* 15**,** 20190182.

MOORE, C. G. & BETT, B. J. 1989. The use of meiofauna in marine pollution impact assessment. *Zoological Journal of the Linnean Society,* 96**,** 263-280.

MOSELEY, R. & MANOYLOV, K. 2012. Diatom community composition from low human impact areas in Southeast U.S. *Georg. J. Sci,* 70**,** 123-143.

MÚRRIA, C., WANGENSTEEN, O. S., SOMMA, S., VÄISÄNEN, L., FORTUÑO, P., ARNEDO, M. A. & PRAT, N. 2024. Taxonomic accuracy and complementarity between bulk and eDNA metabarcoding provides an alternative to morphology for biological assessment of freshwater macroinvertebrates. *Science of The Total Environment,* 935**,** 173243.

MURTAGH, F. & LEGENDRE, P. 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of classification,* 31**,** 274-295.

NAVARRO, C. A., ORELLANA, L. H., MAURIACA, C. & JEREZ, C. A. 2009. Transcriptional and functional studies of Acidithiobacillus ferrooxidans genes related to survival in the presence of copper. *Applied and Environmental Microbiology,* 75**,** 6102-6109.

NAVAS, M., PÉREZ-ESTEBAN, J., TORRES, M. A., HONTORIA, C. & MOLINER, A. 2021. Taxonomic and functional analysis of soil microbial communities in a mining site across a metal (loid) contamination gradient. *European Journal of Soil Science,* 72**,** 1190-1205.

NEEDHAM, D., FICHOT, E., PARADA, A. & YEH, Y. 2019. Fuhrman Lab 515F-926R 16S and 18S rRNA Gene Sequencing Protocol V. 2. protocolsio. *Fuhrman Lab 515F-926R 16S and 18S rRNA Gene Sequencing Protocol V. 2. protocolsio.*

NGUYEN, H. T., TRINH, Q. P., COUVREUR, M., NGUYEN, T. D. & BERT, W. 2021. Description of Hemicycliophora cardamomi sp. n. (Nematoda: Hemicycliophoridae) associated with Amomum longiligulare T.L. Wu and a web-based key for the identification of Hemicycliophora spp. *Journal of Helminthology,* 95**,** e2.

NICHOLS, R. V., VOLLMERS, C., NEWSOM, L. A., WANG, Y., HEINTZMAN, P. D., LEIGHTON, M., GREEN, R. E. & SHAPIRO, B. 2018. Minimizing polymerase biases in metabarcoding. *Molecular ecology resources,* 18**,** 927-939.

NIJLAND, R. 2020. Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. *Molecular Ecology.*

NIKODEMOVA, M., HOLZHAUSEN, E. A., DEBLOIS, C. L., BARNET, J. H., PEPPARD, P. E., SUEN, G. & MALECKI, K. M. 2023. The effect of low-abundance OTU filtering methods on the reliability and variability of microbial composition assessed by 16S rRNA amplicon sequencing. *Front Cell Infect Microbiol,* 13**,** 1165295.

NIU, J., CHEN, H., CAI, L., HE, M., ZHANG, R. & WANG, L. 2022. Grinding Beads Influence Microbial DNA Extraction from Organic-Rich Sub-Seafloor Sediment. *Microorganisms,* 10**,** 2505.

NUGENT, N. & RHINARD, M. 2015. *The european commission*, Bloomsbury Publishing.

NYATI, S., BHATTACHARYA, D., WERTH, S. & HONEGGER, R. 2013. Phylogenetic analysis of LSU and SSU r DNA group I introns of lichen photobionts associated with the genera X anthoria and X anthomendoza (T eloschistaceae, lichenized A scomycetes). *Journal of phycology,* 49**,** 1154-1166.

O'DONNELL, J. L., KELLY, R. P., LOWELL, N. C. & PORT, J. A. 2016. Indexed PCR primers induce template-specific bias in large-scale DNA sequencing studies. *PloS one,* 11**,** e0148698.

OGILVIE, L. A. & GRANT, A. 2008. Linking pollution induced community tolerance (PICT) and microbial community structure in chronically metal polluted estuarine sediments. *Marine Environmental Research,* 65**,** 187-198.

OKSANEN, J., SIMPSON, G., BLANCHET, F., KINDT, R., LEGENDRE, P., MINCHIN, P., O'HARA, R., SOLYMOS, P., STEVENS, M. & SZOECS, E. 2022. vegan: Community Ecology Package. R package version 2.6-4 [online]. Retrieved in 2023, October 10.

OLSGARD, F. & GRAY, J. S. 1995. A comprehensive analysis of the effects of offshore oil and gas exploration and production on the benthic communities of the Norwegian continental shelf. *Marine Ecology Progress Series,* 122**,** 277-306.

OWENS, L. A., FRIANT, S., MARTORELLI DI GENOVA, B., KNOLL, L. J., CONTRERAS, M., NOYA-ALARCON, O., DOMINGUEZ-BELLO, M. G. & GOLDBERG, T. L. 2024. VESPA: an optimized protocol for accurate metabarcoding-based characterization of vertebrate eukaryotic endosymbiont and parasite assemblages. *Nature Communications,* 15**,** 402.

ÖZKURT, E., FRITSCHER, J., SORANZO, N., NG, D. Y., DAVEY, R. P., BAHRAM, M. & HILDEBRAND, F. 2022. LotuS2: an ultrafast and highly accurate tool for amplicon sequencing analysis. *Microbiome,* 10**,** 176.

PAN, Y., XIE, J., YAN, W., ZHANG, T. C. & CHEN, C. 2022. Response of microbial community to different land-use types, nutrients and heavy metals in urban river sediment. *J Environ Manage,* 321**,** 115855.

PANG, S. E., SLIK, J. F., ZURELL, D. & WEBB, E. L. 2023. The clustering of spatially associated species unravels patterns in tropical tree species distributions. *Ecosphere,* 14**,** e4589.

PANOV, A., ESIKOVA, T., SOKOLOV, S., KOSHELEVA, I. & BORONIN, A. 2013. Influence of soil pollution on the composition of a microbial community. *Microbiology,* 82**,** 241-248.

PARADA, A. E., NEEDHAM, D. M. & FUHRMAN, J. A. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental microbiology,* 18**,** 1403-1414.

PAWLOWSKI, J., CERMAKOVA, K., CORDIER, T., FRONTALINI, F., APOTHÉLOZ-PERRET-GENTIL, L. & MERZI, T. 2024. Assessing the potential of nematode metabarcoding for benthic monitoring of offshore oil platforms. *Science of The Total Environment,* 933**,** 173092.

PAWLOWSKI, J., ESLING, P., LEJZEROWICZ, F., CEDHAGEN, T. & WILDING, T. A. 2014. Environmental monitoring through protist next-generation sequencing metabarcoding: Assessing the impact of fish farming on benthic foraminifera communities. *Molecular ecology resources,* 14**,** 1129-1140.

PAWLOWSKI, J., KELLY-QUINN, M., ALTERMATT, F., APOTHÉLOZ-PERRET-GENTIL, L., BEJA, P., BOGGERO, A., BORJA, A., BOUCHEZ, A., CORDIER, T. & DOMAIZON, I. 2018. The future of biotic indices in the ecogenomic era: Integrating (e) DNA metabarcoding in biological assessment of aquatic ecosystems. *Science of the Total Environment,* 637**,** 1295-1310.

PEARMAN, J. K., KEELEY, N. B., WOOD, S. A., LAROCHE, O., ZAIKO, A., THOMSON-LAING, G., BIESSY, L., ATALAH, J. & POCHON, X. 2020. Comparing sediment DNA extraction methods for assessing organic enrichment associated with marine aquaculture. *PeerJ,* 8**,** e10231.

PEARSON, K. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science,* 2**,** 559-572.

PEARSON, T. A. 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. *Oceanography and marine biology: an annual review,* 16**,** 229-311.

PENG, B., WANG, M., WU, Y., HUANG, S., ZHANG, Y., HUANG, J., WANG, Y. & CHEN, C. 2024. Anthropogenic activities affect the diverse autotrophic communities of coastal sediments. *Environmental Pollution,* 361**,** 124817.

PENNANEN, T., FROSTEGARD, A., FRITZE, H. & BAATH, E. 1996. Phospholipid fatty acid composition and heavy metal tolerance of soil microbial communities along two heavy metal-polluted gradients in coniferous forests. *Applied and Environmental Microbiology,* 62**,** 420-428.

PEOPLES, L. M., GRAMMATOPOULOU, E., POMBROL, M., XU, X., OSUNTOKUN, O., BLANTON, J., ALLEN, E. E., NUNNALLY, C. C., DRAZEN, J. C., MAYOR, D. J. & BARTLETT, D. H. 2019. Microbial Community Diversity Within Sediments from Two Geographically Separated Hadal Trenches. *Front Microbiol,* 10**,** 347.

PERCZE, K., HARKAI, Á. & MÉSZÁROS, T. 2024. A Cost-Effective Approach for Single-Stranded DNA Amplification Using Primer-Blocked Asymmetric PCR. *Current Protocols,* 4**,** e1125.

PEREIRA, T. J., FONSECA, G., MUNDO-OCAMPO, M., GUILHERME, B. C. & ROCHA-OLIVARES, A. 2010. Diversity of free-living marine nematodes (Enoplida) from Baja California assessed by integrative taxonomy. *Marine Biology,* 157**,** 1665-1678.

PERRYMAN, S. A. M. 1996. The effect of heavy metal contamination on estuarine benthic fauna at varying levels of biological organisation.

PETRO, C., STARNAWSKI, P., SCHRAMM, A. & KJELDSEN, K. U. 2017. Microbial community assembly in marine sediments. *Aquatic Microbial Ecology,* 79**,** 177-195.

PIOLA, R. F. & JOHNSTON, E. L. 2008. Pollution reduces native diversity and increases invader dominance in marine hard-substrate communities. *Diversity and Distributions,* 14**,** 329-342.

POCHON, X., ZAIKO, A., HOPKINS, G. A., BANKS, J. C. & WOOD, S. A. 2015. Early detection of eukaryotic communities from marine biofilm using high-throughput sequencing: an assessment of different sampling devices. *Biofouling,* 31**,** 241-251.

POLINSKI, J. M., BUCCI, J. P., GASSER, M. & BODNAR, A. G. 2019. Metabarcoding assessment of prokaryotic and eukaryotic taxa in sediments from Stellwagen Bank National Marine Sanctuary. *Sci Rep,* 9**,** 14820.

POMPANON, F., DEAGLE, B. E., SYMONDSON, W. O., BROWN, D. S., JARMAN, S. N. & TABERLET, P. 2012. Who is eating what: diet assessment using next generation sequencing. *Molecular ecology,* 21**,** 1931-1950.

PORAZINSKA, D. L., GIBLIN-DAVIS, R. M., FALLER, L., FARMERIE, W., KANZAKI, N., MORRIS, K., POWERS, T. O., TUCKER, A. E., SUNG, W. & THOMAS, W. K. 2009. Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources,* 9**,** 1439-1450.

PRUESSE, E., QUAST, C., KNITTEL, K., FUCHS, B. M., LUDWIG, W., PEPLIES, J. & GLÖCKNER, F. O. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research,* 35**,** 7188-7196.

QUAIL, M. A., SWERDLOW, H. & TURNER, D. J. 2009. Improved protocols for the illumina genome analyzer sequencing system. *Current protocols in human genetics,* 62**,** 18.2. 1-18.2. 27.

QUINCE, C., LANZEN, A., DAVENPORT, R. J. & TURNBAUGH, P. J. 2011. Removing noise from pyrosequenced amplicons. *BMC bioinformatics,* 12**,** 1-18.

QUINCE, C., WALKER, A. W., SIMPSON, J. T., LOMAN, N. J. & SEGATA, N. 2017. Shotgun metagenomics, from sampling to analysis. *Nature biotechnology,* 35**,** 833-844.

RAINBOW, P. S. 2020. Mining-contaminated estuaries of Cornwall–field research laboratories for trace metal ecotoxicology. *Journal of the Marine Biological Association of the United Kingdom,* 100**,** 195-210.

RAJKOWSKA, K., OTLEWSKA, A., BRONCEL, N. & KUNICKA-STYCZYŃSKA, A. 2023. Microbial diversity and bioactive compounds in dried Lycium barbarum fruits (goji): A comparative study. *Molecules,* 28**,** 4058.

RAMETTE, A. 2007. Multivariate analyses in microbial ecology. *FEMS microbiology ecology,* 62**,** 142-160.

RANJARD, L., DEQUIEDT, S., PRÉVOST-BOURÉ, N. C., THIOULOUSE, J., SABY, N., LELIEVRE, M., MARON, P., MORIN, F., BISPO, A. & JOLIVET, C. 2013. Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nature communications,* 4**,** 1-10.

RAUSCH, P., BASIC, M., BATRA, A., BISCHOFF, S. C., BLAUT, M., CLAVEL, T., GLÄSNER, J., GOPALAKRISHNAN, S., GRASSL, G. A. & GÜNTHER, C. 2016. Analysis of factors contributing to variation in the C57BL/6J fecal microbiota across German animal facilities. *International Journal of Medical Microbiology,* 306**,** 343-355.

REDDY, B. & DUBEY, S. K. 2019. River Ganges water as reservoir of microbes with antibiotic and metal ion resistance genes: high throughput metagenomic approach. *Environmental Pollution,* 246**,** 443-451.

REGUEIRA-IGLESIAS, A., BALSA-CASTRO, C., BLANCO-PINTOS, T. & TOMÁS, I. 2023. Critical review of 16S rRNA gene sequencing workflow in microbiome studies: From primer selection to advanced data analysis. *Molecular Oral Microbiology,* 38**,** 347-399.

REYNOLDS, N. K., JUSINO, M. A., STAJICH, J. E. & SMITH, M. E. 2022. Understudied, underrepresented, and unknown: methodological biases that limit detection of early diverging fungi from environmental samples. *Molecular ecology resources,* 22**,** 1065-1085.

RICOTTA, C. & PODANI, J. 2017. On some properties of the Bray-Curtis dissimilarity and their ecological meaning. *Ecological Complexity,* 31**,** 201-205.

RIESENFELD, C. S., SCHLOSS, P. D. & HANDELSMAN, J. 2004. Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.,* 38**,** 525-552.

RINKE, C., CHUVOCHINA, M., MUSSIG, A. J., CHAUMEIL, P.-A., DAVÍN, A. A., WAITE, D. W., WHITMAN, W. B., PARKS, D. H. & HUGENHOLTZ, P. 2021. A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nature microbiology,* 6**,** 946-959.

RIVERA-GARCIA, L. G., HILL-SPANIK, K. M., BERTHRONG, S. T. & PLANTE, C. J. 2018. Tidal stage changes in structure and diversity of intertidal benthic diatom assemblages: a case study from two contrasting Charleston harbor flats. *Estuaries and coasts,* 41**,** 772-783.

ROLLINSON, G., PIRRIE, D., POWER, M. & CUNDY, A. 2007. Geochemical and mineralogical record of historical mining, Hayle Estuary, Cornwall, UK. *Geoscience in south-west England,* 11**,** 326-337.

ROSENBERG, G. 2014. A new critical estimate of named species-level diversity of the recent Mollusca. *American Malacological Bulletin,* 32**,** 308-322.

RUPPERT, K. M., KLINE, R. J. & RAHMAN, M. S. 2019. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation,* 17**,** e00547.

SALGAONKAR, B. B., DAS, D. & BRAGANÇA, J. M. 2016. Resistance of extremely halophilic archaea to zinc and zinc oxide nanoparticles. *Applied Nanoscience,* 6**,** 251-258.

SALMASO, N., VASSELON, V., RIMET, F., VAUTIER, M., ELERSEK, T., BOSCAINI, A., DONATI, C., MORETTO, M., PINDO, M., RICCIONI, G., STEFANI, E., CAPELLI, C., LEPORI, F., KURMAYER, R., MISCHKE, U., KLEMENČIČ, A. K., NOVAK, K., GRECO, C., FRANZINI, G., FUSATO, G., GIACOMAZZI, F., LEA, A., MENEGON, S., ZAMPIERI, C., MACOR, A., VIRGILIO, D., ZANUT, E., ZORZA, R., BUZZI, F. & DOMAIZON, I. 2022. DNA sequence and taxonomic gap analyses to quantify the coverage of aquatic cyanobacteria and eukaryotic microalgae in reference databases: Results of a survey in the Alpine region. *Science of The Total Environment,* 834**,** 155175.

SAMBROOK, J. 1989. Molecular cloning: a laboratory manual. *Cold Spring Habor Laboratory*.

SANDAA, R.-A., ENGER, Ø. & TORSVIK, V. 1999. Abundance and Diversity of Archaea in Heavy-Metal-Contaminated Soils. *Applied and environmental microbiology,* 65**,** 3293-7.

SANDERS, H. L. 1968. Marine benthic diversity: a comparative study. *The American Naturalist,* 102**,** 243-282.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences,* 74**,** 5463-5467.

SAPKOTA, R. & NICOLAISEN, M. 2015. High-throughput sequencing of nematode communities from total soil DNA extractions. *BMC Ecology,* 15**,** 3.

SAZYKIN, I., KHMELEVTSOVA, L., AZHOGINA, T. & SAZYKINA, M. 2023. Heavy Metals Influence on the Bacterial Community of Soils: A Review. *Agriculture,* 13**,** 653.

SCHENK, J., HÖSS, S., KLEINBÖLTING, N. & TRAUNSPURGER, W. 2022. Suitability of molecular taxonomy for assessing polluted sediments using the NemaSPEAR [%] index. *Ecological Indicators,* 137**,** 108761.

SCHLOSS, P. D., JENIOR, M. L., KOUMPOURAS, C. C., WESTCOTT, S. L. & HIGHLANDER, S. K. 2016. Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ,* 4**,** e1869.

SCHMIDT, P.-A., BÁLINT, M., GRESHAKE, B., BANDOW, C., RÖMBKE, J. & SCHMITT, I. 2013. Illumina metabarcoding of a soil fungal community. *Soil Biology and Biochemistry,* 65**,** 128-132.

SCHOENLE, A., HOHLFELD, M., HERMANNS, K., MAHÉ, F., DE VARGAS, C., NITSCHE, F. & ARNDT, H. 2021. High and specific diversity of protists in the deep-sea basins dominated by diplonemids, kinetoplastids, ciliates and foraminiferans. *Communications Biology,* 4**,** 501.

SCHRATZBERGER, M., HOLTERMAN, M., VAN OEVELEN, D. & HELDER, J. 2019. A Worm's World: Ecological Flexibility Pays Off for Free-Living Nematodes in Sediments and Soils. *BioScience,* 69**,** 867-876.

SCHRATZBERGER, M., WARR, K. & ROGERS, S. 2006. Patterns of nematode populations in the southwestern North Sea and their link to other components of the benthic fauna. *Journal of Sea Research,* 55**,** 113-127.

SCHUIJT, L. M., VAN SMEDEN, J., VAN DRIMMELEN, C. K. E., BUIJSE, L. L., WU, D., BOERWINKEL, M.-C., BELGERS, D. J. M., MATSER, A. M., ROESSINK, I., HEIKAMP-DE JONG, I., BEENTJES, K. K., TRIMBOS, K. B., SMIDT, H. & VAN DEN BRINK, P. J. 2024. Effects of antidepressant exposure on aquatic communities assessed by a combination of morphological identification, functional measurements, environmental DNA metabarcoding and bioassays. *Chemosphere,* 349**,** 140706.

SCHUSTER, S. C. 2008. Next-generation sequencing transforms today's biology. *Nature methods,* 5**,** 16-18.

SEMMOURI, I., DE SCHAMPHELAERE, K. A., MEES, J., JANSSEN, C. R. & ASSELMAN, J. 2020. Evaluating the potential of direct RNA nanopore sequencing: Metatranscriptomics highlights possible seasonal differences in a marine pelagic crustacean zooplankton community. *Marine Environmental Research,* 153**,** 104836.

SEMPRUCCI, F., SBROCCA, C., ROCCHI, M. & BALSAMO, M. 2015. Temporal changes of the meiofaunal assemblage as a tool for the assessment of the ecological quality status. *Journal of the Marine Biological Association of the United Kingdom,* 95**,** 247-254.

SHAKYA, M., LO, C.-C. & CHAIN, P. S. 2019. Advances and challenges in metatranscriptomic analysis. *Frontiers in genetics,* 10**,** 904.

SHEARER, C. A. & LANE, L. C. 1983. Comparison of Three Techniques for the Study of Aquatic Hyphomycete Communities. *Mycologia,* 75**,** 498-508.

SHI, Y., TYSON, G. W., EPPLEY, J. M. & DELONG, E. F. 2011. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *The ISME Journal,* 5**,** 999-1013.

SHI, Z., KONG, Q., LI, X., XU, W., MAO, C., WANG, Y., SONG, W. & HUANG, J. 2022. The effects of DNA extraction kits and primers on prokaryotic and eukaryotic microbial community in freshwater sediments. *Microorganisms,* 10**,** 1213.

SHIPP, E. & GRANT, A. 2006. Hydrobia ulvae feeding rates: a novel way to assess sediment toxicity. *Environmental Toxicology and Chemistry: An International Journal,* 25**,** 3246-3252.

SHYLLA, J. A., GHATANI, S. & TANDON, V. 2013. Utility of divergent domains of 28S ribosomal RNA in species discrimination of paramphistomes (Trematoda: Digenea: Paramphistomoidea). *Parasitology research,* 112**,** 4239-4253.

SIKDER, M. M., VESTERGÅRD, M., KYNDT, T., KUDJORDJIE, E. N. & NICOLAISEN, M. 2021. Phytohormones selectively affect plant parasitic nematodes associated with Arabidopsis roots. *New Phytologist,* 232**,** 1272-1285.

SIKDER, M. M., VESTERGÅRD, M., SAPKOTA, R., KYNDT, T. & NICOLAISEN, M. 2020. Evaluation of metabarcoding primers for analysis of soil nematode communities. *Diversity,* 12**,** 388.

SIMPSON, S. L., ROSNER, J. & ELLIS, J. 2000. Competitive displacement reactions of cadmium, copper, and zinc added to a polluted, sulfidic estuarine sediment. *Environmental Toxicology and Chemistry: An International Journal,* 19**,** 1992-1999.

SINGH, S., SINGH, A., SINGH, R. & SHARMA, N. R. 2023. Heavy Metal Pollution in the Aquatic System and Their Toxicological Effects on Clarias Batrachus. *International Journal of Membrane Science and Technology*.

SMITH, D. P., PEAY, K. G., ACKERMANN, G., APPRILL, A., BAUER, M., BERG-LYONS, D., BETLEY, J., BRUNS, T., CAPORASO, J. G. & FIERER, N. 2018. EMP ITS Illumina amplicon protocol.

SOKAL, R. R. & ROHLF, F. J. 1962. The comparison of dendrograms by objective methods. *Taxon***,** 33-40.

SOMERFIELD, P., GEE, J. & WARWICK, R. 1994a. Soft sediment meiofaunal community structure in relation to a long-term heavy metal gradient in the Fal estuary system. *Marine Ecology Progress Series,* 105**,** 79-88.

SOMERFIELD, P. J., , GEE, J. M., , WARWICK, R. M. & 1994b. Soft sediment meiofaunal community structure in relation to a long-term heavy metal gradient in the Fa1 estuary system.

SOMERFIELD, P. J., MICHAEL GEE, J. & WARWICK, R. M. 1994c. Benthic community structure in relation to an instantaneous discharge of waste water from a tin mine. *Marine Pollution Bulletin,* 28**,** 363-369.

SONG, W., QI, R., ZHAO, L., XUE, N., WANG, L. & YANG, Y. 2019. Bacterial community rather than metals shaping metal resistance genes in water, sediment and biofilm in lakes from arid northwestern China. *Environmental Pollution,* 254**,** 113041.

SONNENBERG, R., NOLTE, A. W. & TAUTZ, D. 2007. An evaluation of LSU rDNA D1-D2 sequences for their use in species identification. *Frontiers in Zoology,* 4**,** 6.

SRIVASTAVA, D., KOLASA, J., BENGTSSON, J., GONZALEZ, A., LAWLER, S., MILLER, T., MUNGUIA, P., ROMANUK, T., SCHNEIDER, D. & TRZCINSKI, M. 2004. Are natural microcosms useful model systems for ecology? *Trends in ecology & evolution,* 19 7**,** 379-384.

ST, L. & WOLD, S. 1989. Analysis of variance (ANOVA). *Chemometrics and intelligent laboratory systems,* 6**,** 259-272.

STOECK, T., BASS, D., NEBEL, M., CHRISTEN, R., JONES, M. D. M., BREINER, H.-W. & RICHARDS, T. A. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology,* 19**,** 21-31.

SUN, J., ZHANG, A., ZHANG, Z., LIU, Y., ZHOU, H., CHENG, H., CHEN, Z., LI, H., ZHANG, R. & WANG, Y. 2023. Distinct assembly processes and environmental adaptation of abundant and rare archaea in Arctic marine sediments. *Marine Environmental Research,* 190**,** 106082.

SZULC, J., OTLEWSKA, A., OKRASA, M., MAJCHRZYCKA, K., SULYOK, M. & GUTAROWSKA, B. 2017. Microbiological contamination at workplaces in a combined heat and power (CHP) station processing plant biomass. *International journal of environmental research and public health,* 14**,** 99.

TABERLET, P., COISSAC, E., HAJIBABAEI, M. & RIESEBERG, L. H. 2012. Environmental DNA. *Molecular ecology,* 21**,** 1789-1793.

TAGLIABUE, A., MATTERSON, K. O., PONTI, M., TURICCHIA, E., ABBIATI, M. & COSTANTINI, F. 2023. Sediment and bottom water eDNA metabarcoding to support coastal management. *Ocean & Coastal Management,* 244**,** 106785.

TEAM, R. C. 2024. *A language and environment for statistical computing* [Online]. R Core Team. Available: http://www.R-project.org/ [Accessed].

TEBBE, C. C. & VAHJEN, W. 1993. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Applied and environmental microbiology,* 59**,** 2657-2665.

TEDERSOO, L., BAHRAM, M., ZINGER, L., NILSSON, R. H., KENNEDY, P. G., YANG, T., ANSLAN, S. & MIKRYUKOV, V. 2022. Best practices in metabarcoding of fungi: from experimental design to results. *Molecular ecology,* 31**,** 2769-2795.

TEDERSOO, L., HOSSEYNI MOGHADDAM, M. S., MIKRYUKOV, V., HAKIMZADEH, A., BAHRAM, M., NILSSON, R. H., YATSIUK, I., GEISEN, S., SCHWELM, A., PIWOSZ, K., PROUS, M., SILDEVER, S., CHMOLOWSKA, D., RUECKERT, S., SKALOUD, P., LAAS, P., TINES, M., JUNG, J.-H., CHOI, J. H., ALKAHTANI, S. & ANSLAN, S. 2024. EUKARYOME: the rRNA gene reference database for identification of all eukaryotes. *Database,* 2024.

TEMPERTON, B., FIELD, D., OLIVER, A., TIWARI, B., MÜHLING, M., JOINT, I. & GILBERT, J. A. 2009. Bias in assessments of marine microbial biodiversity in fosmid libraries as evaluated by pyrosequencing. *The ISME Journal,* 3**,** 792-796.

TERLIZZI, A., SCUDERI, D., FRASCHETTI, S. & ANDERSON, M. 2005. Quantifying effects of pollution on biodiversity: a case study of highly diverse molluscan assemblages in the Mediterranean. *Marine Biology,* 148**,** 293-305.

THORNTON, D. H., BRANCH, L. C. & SUNQUIST, M. E. 2011. The influence of landscape, patch, and within-patch factors on species presence and abundance: a review of focal patch studies. *Landscape Ecology,* 26**,** 7-18.

TIAN, B.-Y., CAO, Y. & ZHANG, K.-Q. 2015. Metagenomic insights into communities, functions of endophytes and their associates with infection by root-knot nematode, Meloidogyne incognita, in tomato roots. *Scientific Reports,* 5**,** 17087.

TIMM, V. F., GONÇALVES, L. T., VALENTE, V. L. D. S. & DEPRA, M. 2022. The efficiency of the COI gene as a DNA barcode and an overview of Orthoptera (Caelifera and Ensifera) sequences in the BOLD System. *Canadian Journal of Zoology,* 100**,** 710-718.

TUREKIAN, K. K. & WEDEPOHL, K. H. 1961. Distribution of the elements in some major units of the earth's crust. *Geological society of America bulletin,* 72**,** 175-192.

TURNER, S., PRYER, K. M., MIAO, V. P. W. & PALMER, J. D. 1999. Investigating Deep Phylogenetic Relationships among Cyanobacteria and Plastids by Small Subunit rRNA Sequence Analysis1. *Journal of Eukaryotic Microbiology,* 46**,** 327-338.

TURNER, T. R., RAMAKRISHNAN, K., WALSHAW, J., HEAVENS, D., ALSTON, M., SWARBRECK, D., OSBOURN, A., GRANT, A. & POOLE, P. S. 2013. Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants. *The ISME Journal,* 7**,** 2248-2258.

TYTGAT, B., NGUYEN, D. T., NGUYEN, T. X. P., PHAM, T. M., LONG, P. K., VANREUSEL, A. & DERYCKE, S. 2019. Monitoring of marine nematode communities through 18S rRNA metabarcoding as a sensitive alternative to morphology. *Ecological Indicators,* 107**,** 105554.

UDOCHI, S. 2020. *Predicting the ecological effects of metals in marine sediments.* University of East Anglia.

URICH, T., LANZÉN, A., QI, J., HUSON, D. H., SCHLEPER, C. & SCHUSTER, S. C. 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PloS one,* 3**,** e2527.

VAULOT, D., GEISEN, S., MAHÉ, F. & BASS, D. 2022. pr2-primers: An 18S rRNA primer database for protists. *Molecular Ecology Resources,* 22**,** 168-179.

VOGT, P., MILJUTINA, M. & RAUPACH, M. J. 2014. The application of DNA sequence data for the identification of benthic nematodes from the North Sea. *Helgoland Marine Research,* 68**,** 549-558.

VOICA, D. M., BARTHA, L., BANCIU, H. L. & OREN, A. 2016. Heavy metal resistance in halophilic Bacteria and Archaea. *FEMS Microbiology Letters,* 363.

VOJTKOVÁ, H. & JANULKOVÁ, R. 2012. Phenotypic characterization of Pseudomonas bacteria isolated from polluted sites of Ostrava, Czech Republic. *GeoScience Engineering,* 58**,** 52.

VON AMMON, U., WOOD, S. A., LAROCHE, O., ZAIKO, A., TAIT, L., LAVERY, S., INGLIS, G. & POCHON, X. 2018. The impact of artificial surfaces on marine bacterial and eukaryotic biofouling assemblages: A high-throughput sequencing analysis. *Marine Environmental Research,* 133**,** 57-66.

VRAIN, T., WAKARCHUK, D., LEVESQUE, A. & HAMILTON, R. 1992. Intraspecific rDNA restriction fragment length polymorphism in the Xiphinema americanum group. *Fundamental and applied Nematology,* 15**,** 563-573.

WADE, W. 2002. Unculturable bacteria—the uncharacterized organisms that cause oral infections. *Journal of the Royal Society of Medicine,* 95**,** 81-83.

WAEYENBERGE, L., DE SUTTER, N., VIAENE, N. & HAEGEMAN, A. 2019. New Insights Into Nematode DNA-metabarcoding as Revealed by the Characterization of Artificial and Spiked Nematode Communities. *Diversity,* 11**,** 52.

WALTERS, W., HYDE EMBRIETTE, R., BERG-LYONS, D., ACKERMANN, G., HUMPHREY, G., PARADA, A., GILBERT JACK, A., JANSSON JANET, K., CAPORASO, J. G., FUHRMAN JED, A., APPRILL, A. & KNIGHT, R. 2015. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems,* 1**,** 10.1128/msystems.00009-15.

WANG, H., BIER, R., ZGLESZEWSKI, L., PEIPOCH, M., OMONDI, E., MUKHERJEE, A., CHEN, F., ZHANG, C. & KAN, J. 2020. Distinct distribution of Archaea from soil to freshwater to estuary: implications of archaeal composition and function in different environments. *Frontiers in Microbiology,* 11**,** 576661.

WANG, W., LEI, J., LI, M., ZHANG, X., XIANG, X., WANG, H., LU, X., MA, L., LIU, X. & TUOVINEN, O. H. 2023a. Archaea are better adapted to antimony stress than their bacterial counterparts in Xikuangshan groundwater. *Science of The Total Environment,* 905**,** 166999.

WANG, Y., TIAN, R. M., GAO, Z. M., BOUGOUFFA, S. & QIAN, P.-Y. 2014. Optimal Eukaryotic 18S and Universal 16S/18S Ribosomal RNA Primers and Their Application in a Study of Symbiosis. *PLOS ONE,* 9**,** e90053.

WANG, Z., LIU, X., LIANG, D., WANG, Q., ZHANG, L. & ZHANG, P. 2023b. VertU: universal multilocus primer sets for eDNA metabarcoding of vertebrate diversity, evaluated by both artificial and natural cases. *Frontiers in Ecology and Evolution,* 11**,** 1164206.

WARWICK, R. 1986. A new method for detecting pollution effects on marine macrobenthic communities. *Marine biology,* 92**,** 557-562.

WARWICK, R., GOSS-CUSTARD, J., KIRBY, R., GEORGE, C., POPE, N. & ROWDEN, A. 1991. Static and dynamic environmental factors determining the community structure of estuarine macrobenthos in SW Britain: why is the Severn estuary different? *Journal of Applied Ecology***,** 329-345.

WARWICK, R. & UNCLES, R. 1980. Distribution of benthic macrofauna associations in the Bristol Channel in relation to tidal stress. *Marine Ecology Progress Series,* 3**,** 97-103.

WARWICK, R. M. 2001. Evidence for the Effects of Metal Contamination on the Intertidal Macrobenthic Assemblages of the Fal Estuary. *Marine Pollution Bulletin,* 42**,** 145-148.

WARWICK, R. M. 2018. The contrasting histories of marine and freshwater meiobenthic research – a result of differing life histories and adaptive strategies? *Journal of Experimental Marine Biology and Ecology,* 502**,** 4-11.

WARWICK, R. M. & CLARKE, K. R. 1991. A Comparison of some methods for analysing changes in benthic community structure. *Journal of the Marine Biological Association of the United Kingdom,* 71**,** 225-244.

WARWICK, R. M. & CLARKE, K. R. 1998. Taxonomic distinctness and environmental assessment. Blackwell Publishing Ltd.

WATANABE, T., HIRAI, J., SILDEVER, S., TADOKORO, K., HIDAKA, K., TANITA, I., NISHIUCHI, K., IGUCHI, N., KASAI, H. & NISHI, N. 2023. Improving taxonomic classification of marine zooplankton by molecular approach: registration of taxonomically verified 18S and 28S rRNA gene sequences. *PeerJ,* 11**,** e15427.

WATLING, L. 1998. Small-scale features of marine sediments and their importance to the study of deposit-feeding. *Marine Ecology-Progress Series,* 47**,** 135.

WEBSTER, G., NEWBERRY, C. J., FRY, J. C. & WEIGHTMAN, A. J. 2003. Assessment of bacterial community structure in the deep sub-seafloor biosphere by 16S rDNA-based techniques: a cautionary tale. *Journal of Microbiological Methods,* 55**,** 155-164.

WEBSTER, G., O'SULLIVAN, L. A., MENG, Y., WILLIAMS, A. S., SASS, A. M., WATKINS, A. J., PARKES, R. J. & WEIGHTMAN, A. J. 2015. Archaeal community diversity and abundance changes along a natural salinity gradient in estuarine sediments. *FEMS microbiology ecology,* 91**,** 1-18.

WEISS, S., XU, Z. Z., PEDDADA, S., AMIR, A., BITTINGER, K., GONZALEZ, A., LOZUPONE, C., ZANEVELD, J. R., VÁZQUEZ-BAEZA, Y., BIRMINGHAM, A., HYDE, E. R. & KNIGHT, R. 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome,* 5**,** 27.

WHITE, T. J., BRUNS, T., LEE, S. & TAYLOR, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications,* 18**,** 315-322.

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

WILSON, J.-J., BRANDON-MONG, G.-J., GAN, H.-M. & SING, K.-W. 2019. High-throughput terrestrial biodiversity assessments: mitochondrial metabarcoding, metagenomics or metatranscriptomics? *Mitochondrial DNA Part A,* 30**,** 60-67.

WURCH, L., GIANNONE, R. J., BELISLE, B. S., SWIFT, C., UTTURKAR, S., HETTICH, R. L., REYSENBACH, A.-L. & PODAR, M. 2016. Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. *Nature communications,* 7**,** 12115.

XUE, Z., LU, H. & LIU, W.-T. 2014. Membrane biofouling characterization: effects of sample preparation procedures on biofilm structure and the microbial community. *Biofouling,* 30**,** 813-821.

YAN, C., WANG, F., GENG, H., LIU, H., PU, S., TIAN, Z., CHEN, H., ZHOU, B., YUAN, R. & YAO, J. 2020. Integrating high-throughput sequencing and metagenome analysis to reveal the characteristic and resistance mechanism of microbial community in metal contaminated sediments. *Science of The Total Environment,* 707**,** 136116.

YAP, T. W.-C., GAN, H.-M., LEE, Y.-P., LEOW, A. H.-R., AZMI, A. N., FRANCOIS, F., PEREZ-PEREZ, G. I., LOKE, M.-F., GOH, K.-L. & VADIVELU, J. 2016. Helicobacter pylori eradication causes perturbation of the human gut microbiome in young adults. *PloS one,* 11**,** e0151893.

YARZA, P., YILMAZ, P., PRUESSE, E., GLÖCKNER, F. O., LUDWIG, W., SCHLEIFER, K.-H., WHITMAN, W. B., EUZÉBY, J., AMANN, R. & ROSSELLÓ-MÓRA, R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology,* 12**,** 635-645.

YAZAKI, E., YABUKI, A., IMAIZUMI, A., KUME, K., HASHIMOTO, T. & INAGAKI, Y. 2022. The closest lineage of Archaeplastida is revealed by phylogenomics analyses that include Microheliella maris. *Open Biology,* 12**,** 210376.

YENNI, G., ADLER, P. B. & ERNEST, S. M. 2017. Do persistent rare species experience stronger negative frequency dependence than common species? *Global Ecology and Biogeography,* 26**,** 513-523.

YI, J., LO, L. S. H., LIU, H., QIAN, P.-Y. & CHENG, J. 2021. Study of Heavy Metals and Microbial Communities in Contaminated Sediments Along an Urban Estuary. *Frontiers in Marine Science,* 8.

YIN, H., NIU, J., REN, Y., CONG, J., ZHANG, X., FAN, F., XIAO, Y., ZHANG, X., DENG, J. & XIE, M. 2015. An integrated insight into the response of sedimentary microbial communities to heavy metal contamination. *Scientific reports,* 5**,** 14266.

YIN, T., ZHANG, X., LONG, Y., JIANG, J., ZHOU, S., CHEN, Z., HU, J. & MA, S. 2024. Impact of soil physicochemical factors and heavy metals on co-occurrence pattern of

bacterial in rural simple garbage dumping site. *Ecotoxicology and Environmental Safety,* 280**,** 116476.

YIN, X., WANG, W., WANG, A., HE, M., LIN, C., OUYANG, W. & LIU, X. 2022. Microbial community structure and metabolic potential in the coastal sediments around the Yellow River Estuary. *Science of The Total Environment,* 816**,** 151582.

YODER, M., DE LEY, I. T., KING, I. W., MUNDO-OCAMPO, M., MANN, J., BLAXTER, M., POIRAS, L. & DE LEY, P. 2006. DESS: a versatile solution for preserving morphology and extractable DNA of nematodes. *Nematology,* 8**,** 367-376.

YU, C., MENG, K., ZHU, Z., LIU, S., ZHOU, Z., ZHANG, H. & XU, M. 2024. Impacts of cadmium accumulation on the diversity, assembly processes, and co-occurrence patterns of archaeal communities in marine sediments. *Science of The Total Environment,* 926**,** 171936.

YU, G., SMITH, D. K., ZHU, H., GUAN, Y. & LAM, T. T. Y. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution,* 8**,** 28-36.

ZAIKO, A., SCHIMANSKI, K., POCHON, X., HOPKINS, G. A., GOLDSTIEN, S., FLOERL, O. & WOOD, S. A. 2016. Metabarcoding improves detection of eukaryotes from early biofouling communities: implications for pest monitoring and pathway management. *Biofouling,* 32**,** 671-684.

ZENG, X., LI, Y., WANG, S., JIA, Y. & ZHU, L. 2023. Diversity, functional potential, and assembly of bacterial communities in metal (loid)-contaminated sediments from a coastal industrial basin. *Science of The Total Environment,* 886**,** 163831.

ZHANG, H., WEI, T., ZHANG, J., LI, Q., FU, L., HE, L. & WANG, Y. 2024a. Microbial mediation of cryptic methane cycle in semiclosed marine water column. *The Innovation Geoscience,* 2**,** 100082-1-100082-11.

ZHANG, Q., ZHANG, H., SU, Z., LI, X. & ZHANG, C. 2007. Effects of long-term petroleum and heavy metals pollution on the diversity and community structure of Pesudomonas populations in agricultural soils. *Ying Yong Sheng tai xue bao= The Journal of Applied Ecology,* 18**,** 1327-1332.

ZHANG, S., ZHAO, J. & YAO, M. 2020. A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish. *Methods in Ecology and Evolution,* 11**,** 1609-1625.

ZHANG, X.-Y., TANG, G.-L., XU, X.-Y., NONG, X.-H. & QI, S.-H. 2014. Insights into deep-sea sediment fungal communities from the East Indian Ocean using targeted environmental sequencing combined with traditional cultivation. *PLoS One,* 9**,** e109118.

ZHANG, X., YANG, H. & CUI, Z. 2017. Assessment on cadmium and lead in soil based on a rhizosphere microbial community. *Toxicology research,* 6**,** 671-677.

ZHANG, Z., XU, D., HUANG, T., ZHANG, Q., LI, Y., ZHOU, J., ZOU, R., LI, X. & CHEN, J. 2024b. High levels of cadmium altered soil archaeal activity, assembly, and co-occurrence network in volcanic areas. *Science of The Total Environment,* 924**,** 171529.

ZHAO, X., HUANG, J., ZHU, X., CHAI, J. & JI, X. 2020. Ecological effects of heavy metal pollution on soil microbial community structure and diversity on both sides of a river around a mining area. *International journal of environmental research and public health,* 17**,** 5680.

ZHAO, Z., LI, H., SUN, Y., ZHAN, A., LAN, W., WOO, S. P., SHAU-HWAI, A. T. & FAN, J. 2022. Bacteria versus fungi for predicting anthropogenic pollution in subtropical coastal sediments: Assembly process and environmental response. *Ecological Indicators,* 134**,** 108484.

ZHENG, P.-F., WEI, Z., ZHOU, Y., LI, Q., QI, Z., DIAO, X. & WANG, Y. 2022. Genomic evidence for the recycling of complex organic carbon by novel Thermoplasmatota clades in deep-sea sediments. *Msystems,* 7**,** e00077-22.

ZIMMERMANN, H. H., HARÐARDÓTTIR, S. & RIBEIRO, S. 2024. Assessing the performance of short 18S rDNA markers for environmental DNA metabarcoding of marine protists. *Environmental DNA,* 6**,** e580.

ZIMMERMANN, J., JAHN, R. & GEMEINHOLZER, B. 2011. Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity & Evolution,* 11**,** 173.

ZOU, D., LIU, H. & LI, M. 2020. Community, Distribution, and Ecological Roles of Estuarine Archaea. *Frontiers in Microbiology,* 11.

Appendix

Table A.1 Primer Version Codes and Total Versions for Forward and Reverse Primers Used in the Project.

Primer	Versioned primers codes				Total combinations
	Forward		Reverse		
515F	515F	A1 to A20	926R	B1 to B22	440
ARF	SSU1ArF	E41 to E52	SSU520R	F41 to F50	120
ITS	ITS1f12	E1 to E22	ITS2	F1 to F21	462
VRAIN	VRAIN2F	L31 and L32	VRAIN2R	M31 and M32	4
E1391	1391f	C1	EukBr	D1	1
D512	D512F	T1 and T2	D978R	U1 and U2	4
TAR	TAReuk454FD1	C41 to C52	TAReukREV3	D41 to D50	120
G18S	G18S4	V1 to V12	22R	W1 to W10	120
NEM	NEM	Q31 to Q38	18Sr2b	R31 to R38	64
JB3	JB3 adjusted	N31 to N38	JB5	P31 to P38	64
DM568	DM568F	X1 to X12	RM3R	Y1 to Y10	120