The utility of a metagenomics approach for marine biomonitoring

Gregory A. C. Singer[1], Shahrokh Shekarriz[1], Avery McCarthy[1], Nicole Fahner[1], Mehrdad Hajibabaei[1,2,*]

[1] Centre for Environmental Genomics Applications, eDNAtec Inc, St. John's, NL, Canada

[2] Centre for Biodiversity Genomics & Department of Integrative Biology, University of Guelph, Guelph, ON, Canada

[*] Corresponding author. Email: hajibabaei@gmail.com

1 Abstract

The isolation and analysis of environmental DNA (eDNA) for ecosystem assessment and monitoring has become increasingly popular. A majority of studies have taken a metabarcoding approach—that is, amplifying and sequencing one or more gene targets of interest. Shotgun sequencing of eDNA—also called metagenomics—while popular in microbial community analysis has not seen much adoption for the analysis of other groups of organisms. Especially in light of the existence of extremely high-capacity DNA sequencers, we decided to test the performance of a shotgun approach side-by-side with a metabarcoding approach on marine water samples obtained from offshore Newfoundland. We found that metabarcoding remains the most efficient technique, but that metagenomics also has significant power to reveal biodiversity patterns, and in fact can be treated as an independent confirmation of ecological gradients. Moreover, we show that metagenomics can also be used to infer factors related to ecosystem health and function.

2 Introduction

The Earth's oceans are under increasing environmental stress, and there have been calls for increased biomonitoring to understand these changes and (Canonico et al. 2019). Unfortunately, conventional observation-based techniques of biomonitoring are infeasible at such scales. However, the combination of DNA barcoding with next-generation DNA sequencing has led to the emergence of "metabarcoding" (Hajibabaei et al. 2011; Taberlet et al. 2012), and hundreds of studies over the past years have firmly established its utility as a means of rapid and accurate bioassessment.

Despite its capabilities, metabarcoding has some drawbacks. Although there exist a number of "universal" primers that purport to amplify common barcode regions over a very broad range of taxa, it is known that these primers don't bind equally well to different templates, leading to biased amplification (Alberdi et al. 2018; Günther et al. 2018; Kelly, Shelton, and Gallego 2019). This is the primary reason why sequence read counts are only weakly correlated to the original starting DNA concentrations (Evans et al. 2016; Fonseca 2018; Piñol, Senar, and Symondson n.d.), and why abundance measurements from the technique are therefore unreliable. Moreover, these PCR biases can even lead to entire taxonomic groups missing (Tedersoo et al. 2015). Different techniques have been employed to ameliorate this bias, but it is most common to employ multiple gene regions (Porter and Hajibabaei 2018) and multiple primer pairs per gene target (Hajibabaei et al. 2019).

Alternatively, rather than amplifying and sequencing standardized DNA barcode regions from samples, it is possible to sequence random segments of DNA directly from the sample—essentially the whole genome shotgun sequencing of environmental samples, which is a technique commonly referred to as "metagenomics" (Handelsman et al. 1998; Chen and Pachter 2005; Tringe and Rubin 2005). Because there is no amplification of the DNA, the read counts should be directly proportional to the proportion of starting material, which itself is correlated to biomass (Yates, Fraser, and Derry 2019; Salter et al. 2019; Bista et al. n.d.). Curiously, although this was the first genomics approach used to characterize ocean biodiversity (Venter et al. 2004), it has since fallen out of favour except in the microbial sphere (Gregory et al. 2019) or in small systems such as the analysis of gut contents (Srivathsan et al. 2015) or small mock communities (Bista et al. n.d.). This is probably due to the relative inefficiency of the procedure: as shown in Table 1, while eukaryotic DNA is expected to be in low copy number within environmental samples (Azam and Malfatti 2007), the genome sizes are thousands of times larger than those of prokaryotes. Moreover, most of this extra DNA is noncoding, meaning it will not find a match in a reference database unless it happens to be closely related to an organism that has undergone whole genome sequencing. As has been observed by Stat et al (Stat et al. 2017), this leads to two outcomes: (1) the fraction of identifiable reads will be much lower than those from a metabarcoding experiment; and (2) while the reads may be an unbiased sample of the original eDNA, the *identifiable* reads will be highly biased towards genomes that have been completely sequenced—i.e., those from prokaryotes and eukaryotic model organisms. Indeed, the eukaryotic taxonomic recovery in previous studies is below 0.5% (Tedersoo et al. 2015; Stat et al. 2017).

With the recent introduction of the Illumina NovaSeq 6000—an ultra-high capacity DNA sequencing instrument that has a very low cost-per-read—we decided to revisit the utility of metagenomics for environmental characterization. We obtained samples from the near offshore of Newfoundland, Canada, and subjected those samples to both metabarcoding and metagenomics protocols.

Table 1: The likelihood of making a taxonomic classification on an environmental shotgun sequence read is an interplay between many factors, some of which are summarized here

eDNA source	Abundance	Genome size	Reference database
Bacteria	↑	↓	↑
Eukaryotic nuclear	↓	↑	↓
Eukaryotic organellar	−	↓	↑

3   Methods

3.1   Sample collection

Triplicate 250mL water samples were taken from surface water simultaneously. Samples were taken from eight locations along two transects in Conception Bay, Newfoundland and Labrador, Canada, on October 13-14, 2017 (Figure 1). These samples cover a range from nearshore to approximately 10km offshore (with a sea bottom depth ranging from a few metres nearshore to approximately 200 metres in
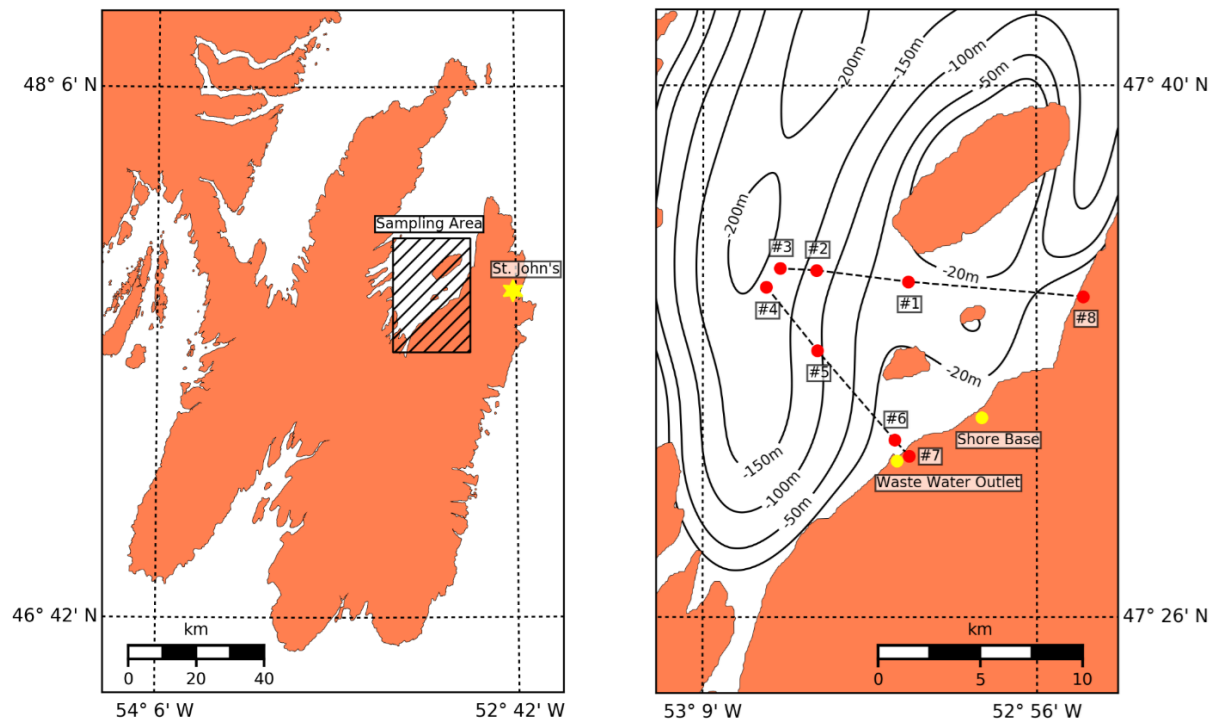
the middle of the bay).

Figure 1: The location of eight sampling sites from Conception Bay, Newfoundland and Labrador, Canada.

3.2   Laboratory

Sample filtration, DNA extraction, and amplicon library prep were detailed in Singer et al. (2018). Briefly, water samples were filtered with 0.22 µm PVDF Sterivex filters (MilliporeSigma) and extracted with the DNeasy PowerWater Kit (Qiagen). Amplicons were then generated from the 232 bp Fish_miniE region of COI with triplicate PCR reactions (Shokralla et al. 2015). These amplicons were indexed with Nextera unique dual indexes (IDT). Amplicons were quantified with the Quant-iT PicoGreen dsDNA assay for normalization and library pooling. Libraries were cleaned with AMPure XP right and left side selections. Negative controls were generated during filtration, extraction, and PCR to screen for contamination and cross-contamination.

Metagenomic libraries were prepared with the Nextera Flex library prep kit (Illumina). For the majority of samples, 5 µL of DNA extract (1.0-3.0 ng/µL) was added to each tagmentation reaction. Three shore samples with higher DNA content were diluted so that 10 ng was added to each reaction (CBNL-052, CBNL-053, and CBNL-054, were diluted 32X, 4X, and 5X respectively). Ten PCR cycles were used to amplify and index tagmented DNA. Unique dual Nextera indexes were used to mitigate index misassignment (IDT; 8-bp index codes). Samples were quantified with Quant-iT PicoGreen dsDNA assay with a Synergy HTX plate fluorometer (BioTek) and pooled to normalize DNA concentration. Libraries were cleaned with three successive AMPure XP cleanups: Left side selection with bead:DNA ratios of 0.75X, then 0.8X, and a right-side selection with 0.5X.

Libraries were sequenced with a 300-cycle S4 kit on the NovaSeq 6000 following the NovaSeq XP workflow. Libraries were pooled with amplicon libraries from other projects for sequencing.

3.3   Bioinformatics

3.3.1   Metabarcoding and metagenomics

All the raw amplicon reads as well as shotgun metagenomics reads, were trimmed for low-quality and Illumina sequencing adapters using Trimmomatic (Bolger, Lohse, and Usadel 2014). After filtering, there were a total of 132.8M total reads across the metabarcoding samples, and 217.2M reads across the metagenomics samples. A local k-mer based taxonomy database contains all of NCBI's nt database constructed using Kraken (Wood and Salzberg 2014), and the taxonomic composition of both amplicon and shotgun reads predicted using Kraken. A threshold cut-off of 0.05 was used to report high confidence taxonomic assignments for each read. The taxonomic assignment tables were merged into one biom file format (McDonald et al. 2012) using an in-house python script and the data imported into the Phyloseq package in R (McMurdie and Holmes 2013) for downstream analysis. The relative abundance as well as read number tables constructed for each dataset. The Shannon metric values and Bray-Curtis distances were calculated for each sample using the Phyloseq package.

3.3.2   Functional inference

Low-quality reads and Illumina universal adapters removed from the raw metagenomic reads using Trimmomatic (as above) and the quality of trimmed reads was assessed using fastqc (Andrews et al. 2012). The Comprehensive Antibiotic Resistance Database (CARD; Alcock et al. 2019) installed locally, and the shotgun reads per sample aligned against CARD's protein homolog models using the Resistance Gene Identifier tool (Alcock et al. 2019). More specifically, FASTQ sequences aligned to the canonical CARD reference sequences that contain available sequences in GenBank with clear evidence for increased minimum inhibitory concentration (MIC) in a peer-reviewed journal available in PubMed.

4   Results and Discussion

4.1   Fewer metagenomics reads can be classified than metabarcoding reads

As noted above, the metabarcoding approach specifically targets genes that have large reference databases, whereas shotgun sequencing will produce random segments from nuclear and organellar genomes that may or may not have close matches in GenBank, and therefore it's expected that a lower proportion of shotgun reads will be identifiable relative to metabarcoding reads. Indeed, this is exactly the case as illustrated in Figure 2. Regardless of the confidence threshold set for taxonomic classification, the result is the same: the majority of reads from metabarcoding studies can at least be identified to the level of kingdom, whereas the majority of reads from shotgun sequencing cannot even be identified at this very course level. The discrepancy is even more stark as attempts to map finer-level taxonomic information onto the reads are made. For example, while only 11.3% of metabarcoding reads were unclassified at the phylum level with 95% confidence, 62.7% of shotgun reads went unclassified.

Figure 2 also upholds the second prediction made above: that taxonomic assignment to reads is also biased. Interestingly, we see opposite trends between the two methodologies here. In metabarcoding experiments, Eukaryota dominate the results—no doubt thanks to the enormous international efforts to build up these reference libraries (Adamowicz 2015). Conversely, the shotgun sequencing experiment shows a much larger proportion of Bacteria than Eukaryota—again likely the reflection of the relative completeness of the whole genome reference libraries to match against.
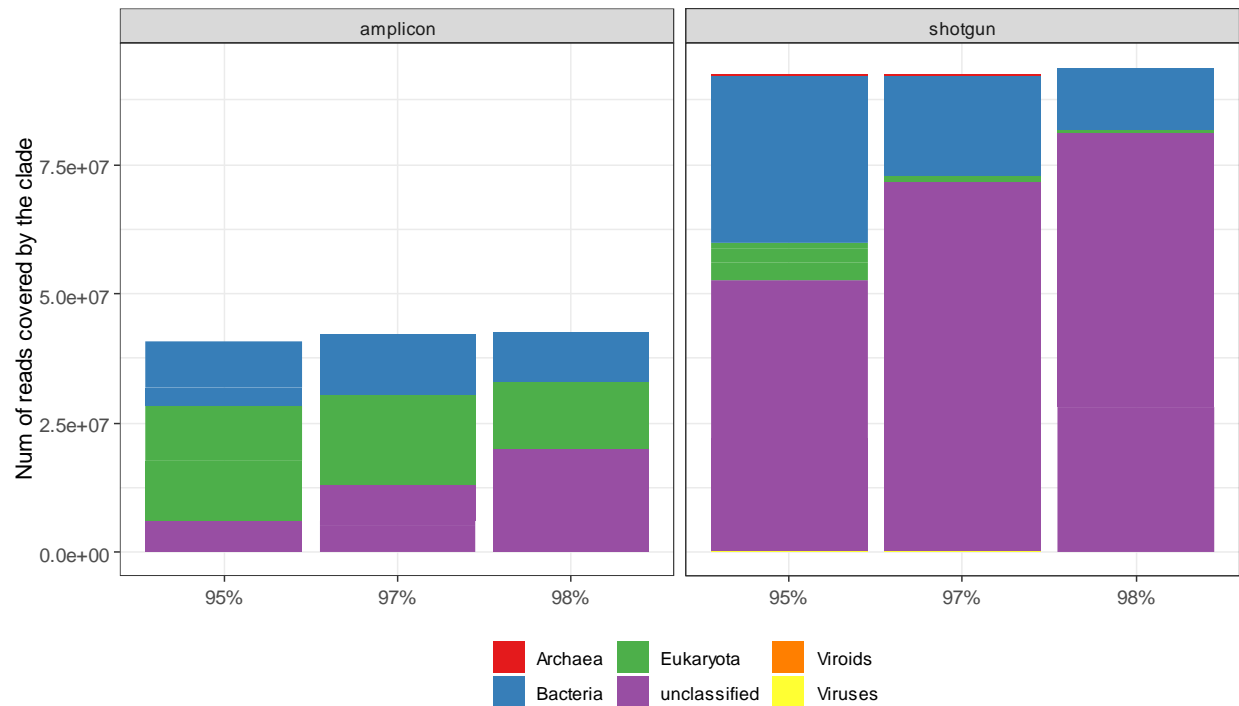
Figure 2: The proportion of unclassified reads (purple) is greater in the metagenomics experiment (right) than in the metabarcoding experiment (left). Data are plotted for a single sample (001). Note that the relative proportion of bacteria to eukaryotes among identified reads is much greater in the metagenomics experiment than in the metabarcoding experiment.

4.2 The pattern of diversity revealed by metagenomics is starkly different than that revealed through metabarcoding

If we restrict our analysis to just the Metazoan phyla, the two methodologies produce very different pictures of the biodiversity present in the environment (Figure 3). The metabarcoding experiments show a very high proportion of arthropods and few chordates, while this situation is reversed in the metagenomics experiment. This difference is even more stark when raw read counts are plotted (Figure 4). Note that despite the two-fold greater sequencing depth in the metagenomics experiments, the proportion of reads mapping to metazoan phyla is approximately half that of the metabarcoding experiments, again underscoring the relative inefficiency of the metagenomics approach. Also, the pattern of biodiversity is again starkly different—so much so that it is almost difficult to believe they came from the exact same DNA extracts. In this way, metagenomics data can be seen as complementary to metabarcoding data—both providing useful information that does not perfectly overlap. This is similar to the notion that metabarcoding data itself is complementary to traditional observational methodologies (Emilson et al. 2017)

One observation of note is that there is surprisingly little stochastic effect from one replicate to the next in the metagenomics experiments. Although this was a deep sequencing experiment, we'd still expect a very small fraction of the total eDNA to have been sequenced, and yet the taxonomic composition from one sample to the next is consistent.
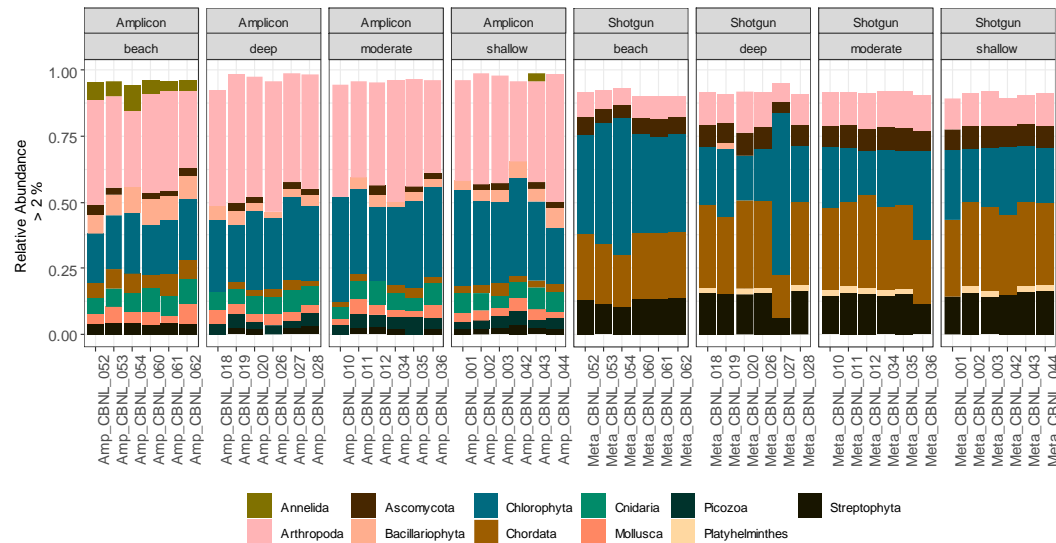
Figure 3: The relative proportions of identified taxa are very different between the metabarcoding experiments (left) vs the metagenomics experiment (right), although both have excellent internal consistency. Only phyla exceeding 2% are shown which is why the bars aren't a uniform height.
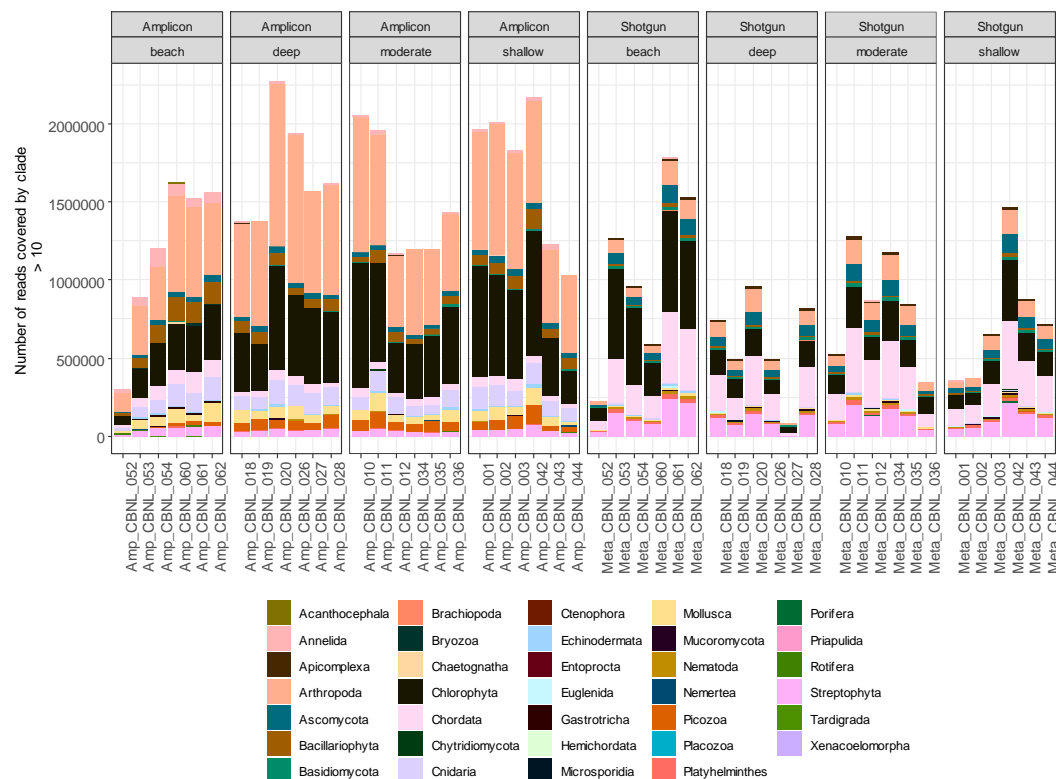

Figure 4: Absolute read counts for each phylum are now shown, illustrating a combination of the greater efficiency of obtaining identified taxa in metabarcoding (taller bars to the left than to the right), as well as the wide disparity in the composition of biodiversity between the two experiments.

4.3 Despite lower resolution, metagenomics data can still uncover meaningful biological patterns

Based on beta diversity, a very clear pattern emerged that showed a distinction between the water samples obtained close to shore ("beach") and those from deeper waters (Figure 5A). Interestingly, despite the lower taxonomic resolution of the metagenomic data, the same pattern emerged in that study as well, albeit not as starkly (Figure 5B). When both metabarcoding and metagenomic data are combined, not surprisingly the primary differentiator is the methodology used, and then both methods show a difference between beach and other water samples along the secondary axis (Figure 5C). Curiously, the directionality of the difference between beach and other samples differs between the two methods, however, with beach samples higher on the y-axis than the other water samples, yet lower on the y-axis than the other water samples in the metagenomics experiments. This demonstrates that while both techniques are able to distinguish the sampling sites from each other, the exact taxa that are used to do so are completely different. In this way, a metagenomics assay can be seen as an independent confirmation of biological trends observed within metabarcoding experiments, since different taxonomic groups are likely to respond to similar ecological gradients (Ritter et al. 2019).
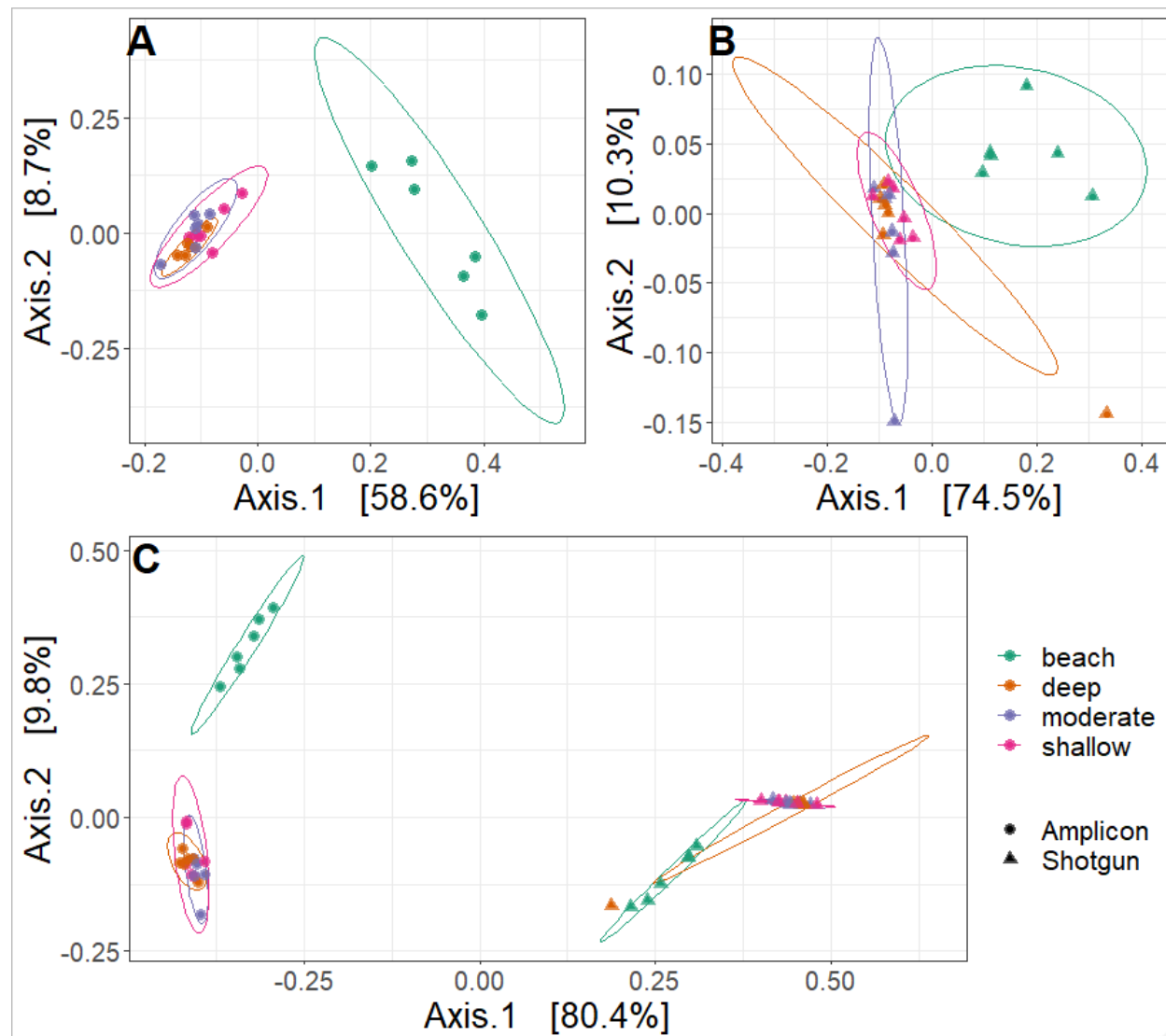
Figure 5: The beach samples cluster separately from the other samples in the metabarcoding data (A), and this is also true of the metabarcoding data (B), although with greater variation primarily due to two odd datapoints from deep- and moderate-depth water. When the data from both experiments are compared (C), the primary axis of variation (accounting for >80% of the variation in the data) separates the methodology (metabarcoding on the left and metagenomics on the right), while the second axis separates the beach samples from the other water samples. Interestingly, the polarity is inverted between the two experiments: the beach samples have higher values on the y-axis than the other water samples among the metabarcoding data, but this relationship is inverse among the metagenomics samples.

4.4   Metagenomics give us additional useful biological insights

One of the primary benefits of a metagenomics approach is that other biological information beyond simple organism identification can be obtained. Most notably, gene fragments can be identified and provide insight into the functional genomics of the environment. Here, we are talking about which genes are *present* in the environment and not which genes are *active*, which would require an RNA-seq assay (Wang, Gerstein, and Snyder 2009). Nevertheless, particularly within the microbial world, simply knowing which genes are present can give important insights into the functional diversity of the environment (Dinsdale et al. 2008). In the present analysis, we mapped shotgun sequencing reads against the CARD antimicrobial resistance gene database (Alcock et al. 2019). A principal coordinates analysis of the relative abundances of these genes reveals that the beach samples are highly distinct from the samples taken from further offshore (Figure 6A), and hierarchical clustering supports this conclusion (Figure 6B).

To determine which gene sets were driving this difference, we plotted the relative proportions of different classes of antimicrobial resistance genes and found that the beach samples were relatively rich in antibiotic inactivation genes (Figure 7). Further analysis of these genes revealed that 24 gene families contributed to this division, 13 of which were beta-lactamases that provide resistance to broad spectrums of antimicrobials including penicillins, cephalosporins, cephamycins, and carbapenems. These genes are common within the bacterial family Enterobacteriaceae, which includes many human pathogens including *E. coli*, *Salmonella*, and *Shigella*. We turned back to the metabarcoding data to see if this family occurred in higher relative proportion in the beach samples compared to samples from further ashore, and indeed this was the case (Figure 8).
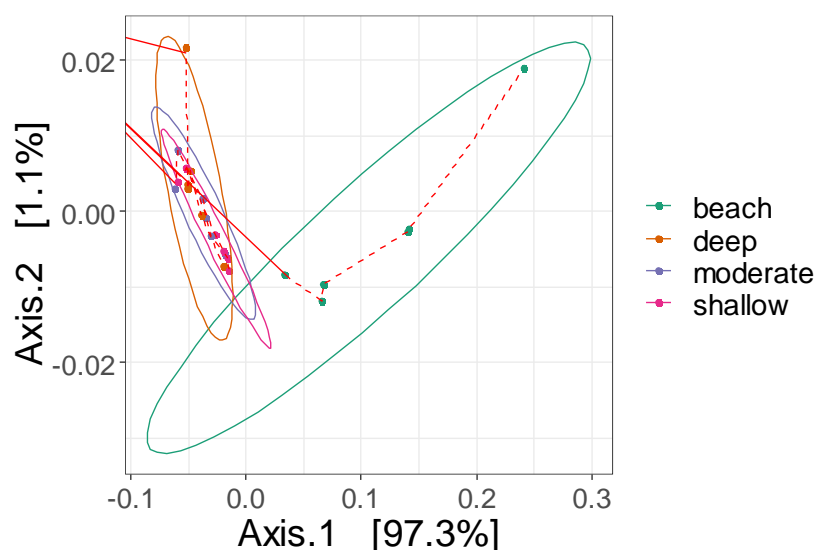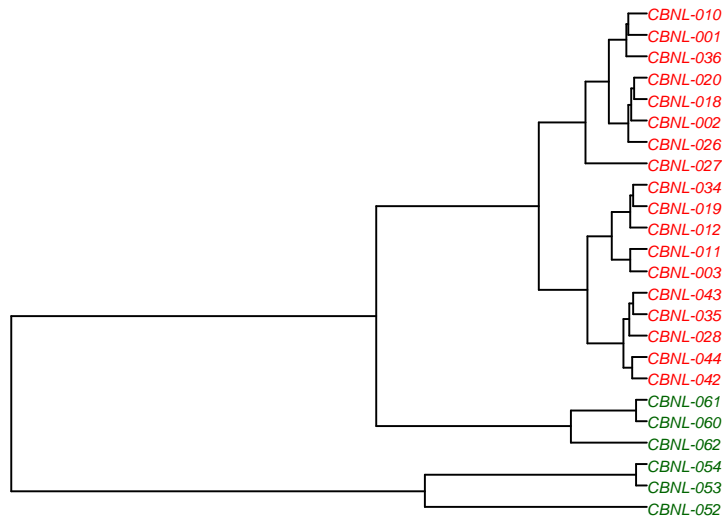
Figure 6: (top) PCoA based on the Bray-Curtis distances of antibiotic resistance genes. The coastal samples are clearly delineated from the non-coastal. (bottom) Hierarchical clustering of the data reveals the same pattern, with coastal (green) and non-coastal (red) samples clustering into their own clades. Interestingly, the coastal samples closest to the storm drain outlet (052, 053, and 054) are an outgroup to the rest of the samples.
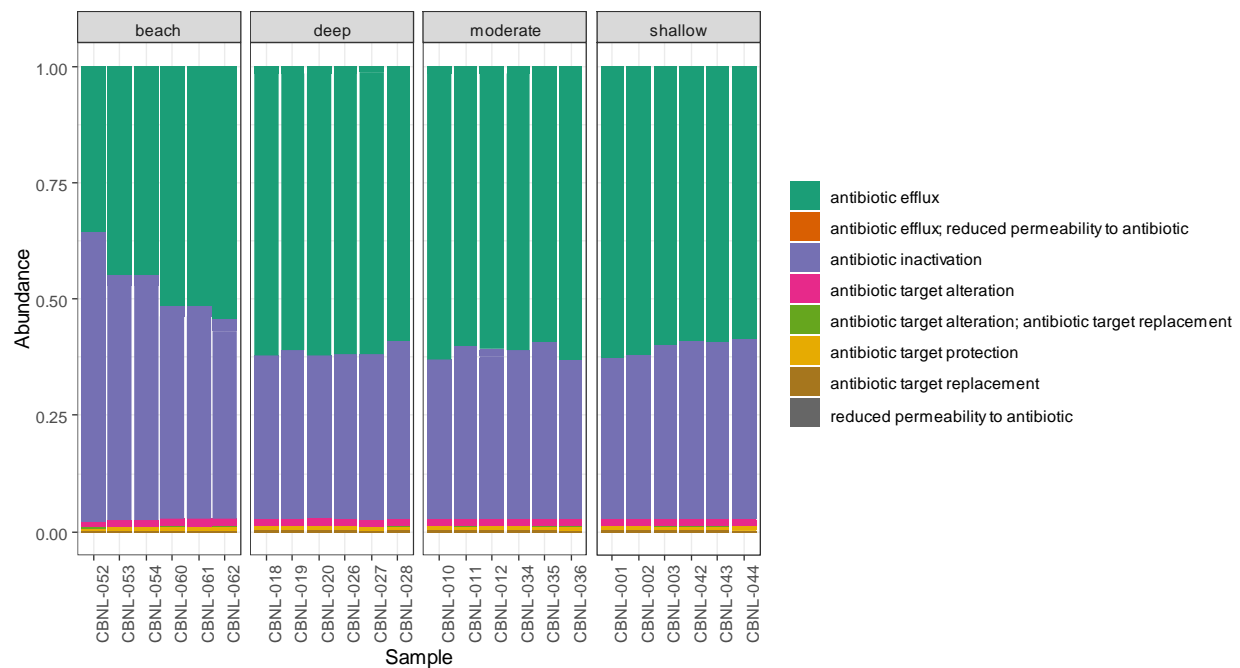

Figure 7: Water samples obtained close to shore have higher proportions of "antibiotic inactivation" genes present than samples taken from further offshore
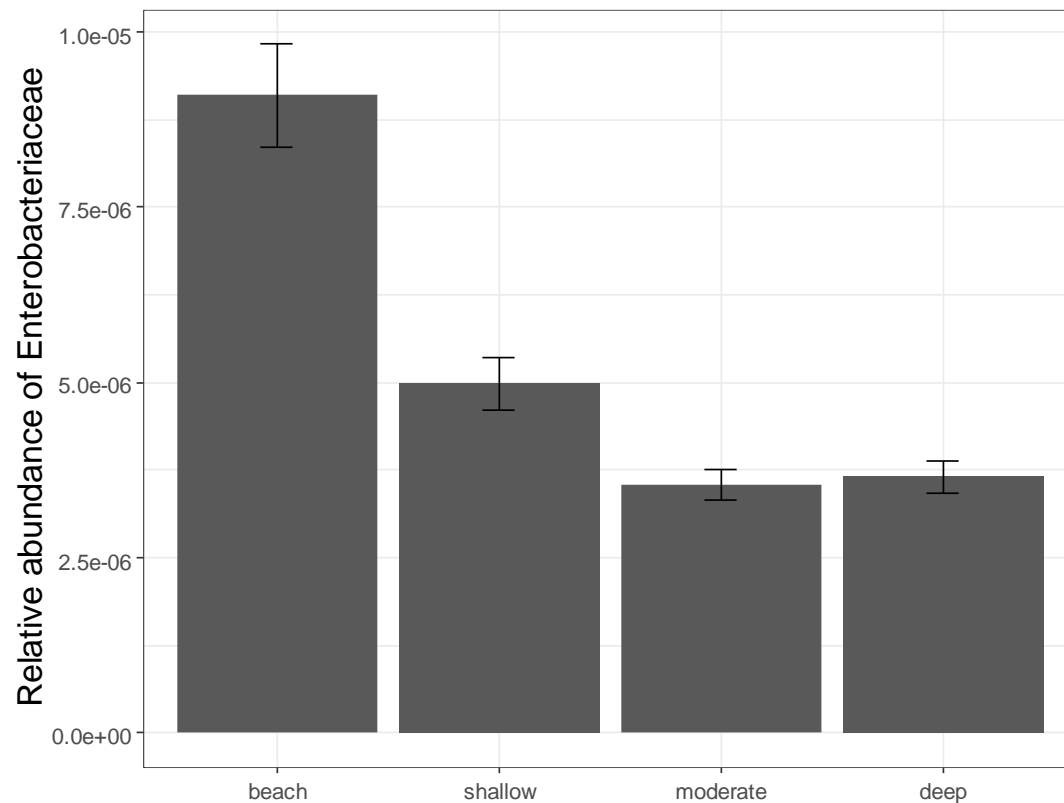
Figure 8: Beta-lactamase genes are known to be prevalent among the Enterobaceriacae. Here we turn back to the metabarcoding data and show that the relative proportion of this family of bacteria is indeed higher among the beach samples than in samples from deeper waters. Error bars indicate standard errors from the mean.

4.5   Metabarcoding experiments can be augmented with metagenomics data for "free"

It is interesting to note that on Illumina DNA sequencing platforms—the most popular instruments for metabarcoding experiments at the moment (Singer et al. 2019)—adding a metagenomics library to the flow cell can potentially be performed without adding any additional sequencing costs. This is because Illumina recommends spiking in a control library composed of PhiX genome to increase base complexity on the flow cell, with a recommended proportion of PhiX ranging from 5% to 50% (Illumina, 2017). We suggest that a shotgun environmental eDNA library could be used in place of (or, better, in combination with) PhiX to accomplish this same goal, without "wasting" space on the flow cell. Therefore, there is little downside to obtaining these useful, complementary data.

5   References

Adamowicz, Sarah J. 2015. 'International Barcode of Life: Evolution of a Global Research Community'. *Genome* 58 (5): 151–62. https://doi.org/10.1139/gen-2015-0094.

Alberdi, Antton, Ostaizka Aizpurua, M. Thomas P. Gilbert, and Kristine Bohmann. 2018. 'Scrutinizing Key Steps for Reliable Metabarcoding of Environmental Samples'. *Methods in Ecology and Evolution* 9 (1): 134–47. https://doi.org/10.1111/2041-210X.12849.

Alcock, Brian P., Amogelang R. Raphenya, Tammy T. Y. Lau, Kara K. Tsang, Mégane Bouchard, Arman Edalatmand, William Huynh, et al. 2019. 'CARD 2020: Antibiotic Resistome Surveillance with the Comprehensive Antibiotic Resistance Database'. *Nucleic Acids Research*, October. https://doi.org/10.1093/nar/gkz935.

Andrews, Simon, Felix Krueger, Anne Segonds-Pichon, Laura Biggins, Christel Krueger, and Steven Wingett. 2012. *FastQC*. Babraham, UK.

Azam, Farooq, and Francesca Malfatti. 2007. 'Microbial Structuring of Marine Ecosystems'. *Nature Reviews Microbiology* 5 (10): 782–91. https://doi.org/10.1038/nrmicro1747.

Bista, Iliana, Gary R. Carvalho, Min Tang, Kerry Walsh, Xin Zhou, Mehrdad Hajibabaei, Shadi Shokralla, et al. n.d. 'Performance of Amplicon and Shotgun Sequencing for Accurate Biomass Estimation in Invertebrate Community Samples'. *Molecular Ecology Resources* 0 (0). Accessed 29 June 2018. https://doi.org/10.1111/1755-0998.12888.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data'. *Bioinformatics* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Canonico, Gabrielle, Pier Luigi Buttigieg, Enrique Montes, Frank E. Muller-Karger, Carol Stepien, Dawn Wright, Abigail Benson, et al. 2019. 'Global Observational Needs and Resources for Marine Biodiversity'. *Frontiers in Marine Science* 6. https://doi.org/10.3389/fmars.2019.00367.

Chen, Kevin, and Lior Pachter. 2005. 'Bioinformatics for Whole-Genome Shotgun Sequencing of Microbial Communities'. *PLOS Computational Biology* 1 (2): e24. https://doi.org/10.1371/journal.pcbi.0010024.

Dinsdale, Elizabeth A., Robert A. Edwards, Dana Hall, Florent Angly, Mya Breitbart, Jennifer M. Brulc, Mike Furlan, et al. 2008. 'Functional Metagenomic Profiling of Nine Biomes'. *Nature* 452 (7187): 629–32. https://doi.org/10.1038/nature06810.

Emilson, Caroline E., Dean G. Thompson, Lisa A. Venier, Teresita M. Porter, Tom Swystun, Derek Chartrand, Scott Capell, and Mehrdad Hajibabaei. 2017. 'DNA Metabarcoding and Morphological Macroinvertebrate Metrics Reveal the Same Changes in Boreal Watersheds across an Environmental Gradient'. *Scientific Reports* 7 (1): 12777. https://doi.org/10.1038/s41598-017-13157-x.

Evans, Nathan T., Brett P. Olds, Mark A. Renshaw, Cameron R. Turner, Yiyuan Li, Christopher L. Jerde, Andrew R. Mahon, Michael E. Pfrender, Gary A. Lamberti, and David M. Lodge. 2016. 'Quantification of Mesocosm Fish and Amphibian Species Diversity via Environmental DNA Metabarcoding'. *Molecular Ecology Resources* 16 (1): 29–41. https://doi.org/10.1111/1755-0998.12433.

Fonseca, Vera G. 2018. '"Pitfalls in Relative Abundance Estimation Using EDNA Metabarcoding"'. *Molecular Ecology Resources* 18 (5): 923–26. https://doi.org/10.1111/1755-0998.12902.

Gregory, Ann C., Ahmed A. Zayed, Nádia Conceição-Neto, Ben Temperton, Ben Bolduc, Adriana Alberti, Mathieu Ardyna, et al. 2019. 'Marine DNA Viral Macro- and Microdiversity from Pole to Pole'. *Cell* 177 (5): 1109-1123.e14. https://doi.org/10.1016/j.cell.2019.03.040.

Günther, Babett, Thomas Knebelsberger, Hermann Neumann, Silke Laakmann, and Pedro Martínez Arbizu. 2018. 'Metabarcoding of Marine Environmental DNA Based on Mitochondrial and Nuclear Genes'. *Scientific Reports* 8 (1): 1–13. https://doi.org/10.1038/s41598-018-32917-x.

Hajibabaei, Mehrdad, Teresita M. Porter, Michael Wright, and Josip Rudar. 2019. 'COI Metabarcoding Primer Choice Affects Richness and Recovery of Indicator Taxa in Freshwater Systems'. *BioRxiv*, March, 572628. https://doi.org/10.1101/572628.

Hajibabaei, Mehrdad, Shadi Shokralla, Xin Zhou, Gregory A. C. Singer, and Donald J. Baird. 2011. 'Environmental Barcoding: A next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos'. *PloS One* 6 (4): e17497. https://doi.org/10.1371/journal.pone.0017497.

Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman. 1998. 'Molecular Biological Access to the Chemistry of Unknown Soil Microbes: A New Frontier for Natural Products'. *Chemistry & Biology* 5 (10): R245-249. https://doi.org/10.1016/s1074-5521(98)90108-9.

'How Much PhiX Spike-in Is Recommended When Sequencing Low Diversity Libraries on Illumina Platforms?' n.d. Accessed 14 November 2018. https://support.illumina.com/bulletins/2017/02/how-much-phix-spike-in-is-recommended-when-sequencing-low-divers.html.

Kelly, Ryan P., Andrew Olaf Shelton, and Ramón Gallego. 2019. 'Understanding PCR Processes to Draw Meaningful Conclusions from Environmental DNA Studies'. *Scientific Reports* 9 (1): 1–14. https://doi.org/10.1038/s41598-019-48546-x.

McDonald, Daniel, Jose C. Clemente, Justin Kuczynski, Jai Ram Rideout, Jesse Stombaugh, Doug Wendel, Andreas Wilke, et al. 2012. 'The Biological Observation Matrix (BIOM) Format or: How I Learned to Stop Worrying and Love the Ome-Ome'. *GigaScience* 1 (1): 7. https://doi.org/10.1186/2047-217X-1-7.

McMurdie, Paul J., and Susan Holmes. 2013. 'Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data'. *PLOS ONE* 8 (4): e61217. https://doi.org/10.1371/journal.pone.0061217.

Piñol, Josep, Miquel A. Senar, and William O. C. Symondson. n.d. 'The Choice of Universal Primers and the Characteristics of the Species Mixture Determines When DNA Metabarcoding Can Be Quantitative'. *Molecular Ecology* 0 (ja). Accessed 3 July 2018. https://doi.org/10.1111/mec.14776.

Porter, Teresita M., and Mehrdad Hajibabaei. 2018. 'Scaling up: A Guide to High-Throughput Genomic Approaches for Biodiversity Analysis'. *Molecular Ecology* 27 (2): 313–38. https://doi.org/10.1111/mec.14478.

Ritter, Camila D., Søren Faurby, Dominic J. Bennett, Luciano N. Naka, Hans ter Steege, Alexander Zizka, Quiterie Haenel, R. Henrik Nilsson, and Alexandre Antonelli. 2019. 'The Pitfalls of Biodiversity Proxies: Differences in Richness Patterns of Birds, Trees and Understudied Diversity across Amazonia'. *Scientific Reports* 9 (1): 1–13. https://doi.org/10.1038/s41598-019-55490-3.

Salter, Ian, Mourits Joensen, Regin Kristiansen, Petur Steingrund, and Poul Vestergaard. 2019. 'Environmental DNA Concentrations Are Correlated with Regional Biomass of Atlantic Cod in Oceanic Waters'. *Communications Biology* 2 (1): 1–9. https://doi.org/10.1038/s42003-019-0696-8.

Shokralla, Shadi, Rosalee S. Hellberg, Sara M. Handy, Ian King, and Mehrdad Hajibabaei. 2015. 'A DNA Mini-Barcoding System for Authentication of Processed Fish Products'. *Scientific Reports* 5 (October): 15894. https://doi.org/10.1038/srep15894.

Singer, G. a. C., N. A. Fahner, J. G. Barnes, A. McCarthy, and M. Hajibabaei. 2019. 'Comprehensive Biodiversity Analysis via Ultra-Deep Patterned Flow Cell Technology: A Case Study of EDNA Metabarcoding Seawater'. *Scientific Reports* 9 (1): 1–12. https://doi.org/10.1038/s41598-019-42455-9.

Srivathsan, Amrita, John C. M. Sha, Alfried P. Vogler, and Rudolf Meier. 2015. 'Comparing the Effectiveness of Metagenomics and Metabarcoding for Diet Analysis of a Leaf-Feeding Monkey (Pygathrix Nemaeus)'. *Molecular Ecology Resources* 15 (2): 250–61. https://doi.org/10.1111/1755-0998.12302.

Stat, Michael, Megan J. Huggett, Rachele Bernasconi, Joseph D. DiBattista, Tina E. Berry, Stephen J. Newman, Euan S. Harvey, and Michael Bunce. 2017. 'Ecosystem Biomonitoring with EDNA: Metabarcoding across the Tree of Life in a Tropical Marine Environment'. *Scientific Reports* 7 (1): 12240. https://doi.org/10.1038/s41598-017-12501-5.

Taberlet, Pierre, Eric Coissac, François Pompanon, Christian Brochmann, and Eske Willerslev. 2012. 'Towards Next-Generation Biodiversity Assessment Using DNA Metabarcoding'. *Molecular Ecology* 21 (8): 2045–50. https://doi.org/10.1111/j.1365-294X.2012.05470.x.

Tedersoo, Leho, Sten Anslan, Mohammad Bahram, Sergei Põlme, Taavi Riit, Ingrid Liiv, Urmas Kõljalg, et al. 2015. 'Shotgun Metagenomes and Multiple Primer Pair-Barcode Combinations of Amplicons Reveal Biases in Metabarcoding Analyses of Fungi'. *MycoKeys* 10 (May): 1–43. https://doi.org/10.3897/mycokeys.10.4852.

Tringe, Susannah Green, and Edward M. Rubin. 2005. 'Metagenomics: DNA Sequencing of Environmental Samples'. *Nature Reviews Genetics* 6 (11): 805–14. https://doi.org/10.1038/nrg1709.

Venter, J. Craig, Karin Remington, John F. Heidelberg, Aaron L. Halpern, Doug Rusch, Jonathan A. Eisen, Dongying Wu, et al. 2004. 'Environmental Genome Shotgun Sequencing of the Sargasso Sea'. *Science* 304 (5667): 66–74. https://doi.org/10.1126/science.1093857.

Wang, Zhong, Mark Gerstein, and Michael Snyder. 2009. 'RNA-Seq: A Revolutionary Tool for Transcriptomics'. *Nature Reviews Genetics* 10 (1): 57–63. https://doi.org/10.1038/nrg2484.

Wood, Derrick E., and Steven L. Salzberg. 2014. 'Kraken: Ultrafast Metagenomic Sequence Classification Using Exact Alignments'. *Genome Biology* 15 (3): R46. https://doi.org/10.1186/gb-2014-15-3-r46.

Yates, Matthew C., Dylan J. Fraser, and Alison M. Derry. 2019. 'Meta-Analysis Supports Further Refinement of EDNA for Monitoring Aquatic Species-Specific Abundance in Nature'. *Environmental DNA* 1 (1): 5–13. https://doi.org/10.1002/edn3.7.