**Ian Bishop**

**BIO-594\_2019**

**4/25/19**

**Work count: 818**

This week we read two articles that each in their own way sought to characterize "hidden" diversity in two different environments using NGS. Boussarie et al. (BO) used NGS in order to assess the insufficiency of traditional methods to determine species richness in mobile and elusive sharks.  Sunagawa et al. (SU) described the massive picoplankton diversity in epi- and mesopelagic waters across the global ocean. First I'll summarize each paper then contrast them a bit.

Boussarie et al. hypothesized that estimates of dark diversity, that is the difference between expected and observed richness for a given locality, is often overestimated because traditional survey methods are not suitable for rare, mobile animals with aversive behavior like sharks. They compared three methods that detect diversity in marine coastal ecosystems: underwater visual census (UVC; traditional), baited remote underwater video station surveys (BRUVS; traditional), and environmental DNA surveys (eDNA; next generation). They found that dark diversity is reduced when eDNA was analyzed (via metabarcoding of COI), where many more species were detected with orders of magnitude fewer samples collected and each sample's cost and effort was lower as well. They found that each eDNA sample detected significantly more diversity than traditional methods, and that only rarefaction curves for eDNA could accurately measure shark diversity in the region. While eDNA is not a silver bullet (it cannot provide information on behavior, health, age, sex, size or abundance), it can accurately measure diversity and should be implemented to facilitate conservation and restoration efforts in at-risk animal groups like sharks.

In contrast, Sunagawa et al. used shotgun metagenomics in addition to metabarcoding (16S, not COI) to examine phylogenetic and functional diversity in marine plankton. They found ~35,000 OTUs across the globe, found that temperature, not geographic distance, was the dominant driver of community differentiation in their samples, and that there is considerable overlap in the core functional gene sets of gut and marine plankton. Most of the metagenomic sequences they collected were new to science and that trove was compiled into the Ocean Microbial Reference Gene Catalog (OMR-GC). Importantly, the combination of amplicon sequencing for phylogenetic marker genes (16S) and shotgun metagenomics allowed the Tara consortium to pair who is present in a sample with data on gene function, allowing them to describe functional redundancy across the globe, which they hypothesize is a common emergent property of microbial ecosystems, given that such redundancy is also found in the human gut microbiome.

One difference between these two papers is scale. While BO collected DNA at the local scale, SU collected DNA at stations in every ocean basic except for the Arctic (which was visited during a subsequent Tara cruise). The amplicon sequencing in both instances successfully characterized the hidden diversity of each ecosystem. BO did however note that scaling eDNA up to broader spatial scales might be an issue in terms of DNA advection with currents. While this wasn't a problem in their case, as the reef of focus was isolated, it could cause a problem in more connected and expansive habitat, such as the Great Barrier Reef.

Another point stressed by BO that is applicable to both studies is the idea that minimal sample numbers can uncover considerable diversity. Both groups demonstrated their sampling efficiency by plotting rarefaction curves. BO estimated that by 200 samples one would nearly reach the rarefaction asymptote, and noted the relative ease and cost-efficiency of such minimal and non-invasive sampling. SU required only 243 samples to find tens of million of functional genes, which again is not very many samples. The sampling in this case is costly first and foremost because of sampling location and ship time, not by the collection and analysis of these specific DNA samples.

Lastly, in their methods sections, two important considerations differed between the papers as they approached bioinformatic processing of eDNA: singleton retention and cluster similarity. BO opted to remove singletons (where only a single read is found for a given OTU), whether or not it was tied to prior knowledge in the BOLD database. In looking at other recent survey work, it is clear that singletons can represent a large fraction of the sequencing data one collects in a metabarcoding effort (e.g. Leray & Knowlton 2015), and therefore contamination or sequencing error could have a big false positive effect on overall or sample-specific richness. Singleton inclusion or exclusion is not mentioned in SU. And second, each paper used a different threshold for cluster similarity, 97% vs 99%. I would expect this has to do with taxonomic group and barcode specific differences, but in general it illustrates that there is no single cutoff one can apply without considering context. Newer oligotyping methods might get around this altogether by examining differences in Shannon entropy and not relying on a priori cluster cutoffs (Eren et al. 2013).

***References:***

Boussarie, G., Bakker, J., Wangensteen, O. S., Mariani, S., Bonnin, L., Juhel, J.-B., … Mouillot, D. (2018). Environmental DNA illuminates the dark diversity of sharks. *Science Advances*, *4*(5), eaap9661. <https://doi.org/10.1126/sciadv.aap9661>

Eren, A. M., Maignien, L., Sul, W. J., Murphy, L. G., Grim, S. L., Morrison, H. G., & Sogin, M. L. (2013). Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods in Ecology and Evolution*, *4*(12), 1111–1119. <https://doi.org/10.1111/2041-210X.12114>

Leray, M., & Knowlton, N. (n.d.). *DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity*. 6.

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., … Velayoudon, D. (2015). Structure and function of the global ocean microbiome. *Science*, *348*(6237), 1261359–1261359. <https://doi.org/10.1126/science.1261359>