How Accurate is Species Detection with Metabarcoding

Milena Nunes, Paula Pappalardo, Stephanie Bush, Karen Osborn

OCE-1560088

Introduction:

Metabarcoding has the potential to allow rapid determination of the species present in an entire community, which in turn allows monitoring and greatly increased power to compare communities in space or time. However, taxonomic identification of metabarcodes requires the use of a reference library of known sequences and building libraries with sufficient taxonomic coverage takes a great deal of time and taxonomic expertise. Here we test 3 different reference libraries to determine the importance of reference library choice.

We used two groups of planktonic taxa to determine *how accurate metabarcode species assignment is with different reference libraries.*

The inclusion of both barcodes and metabarcodes in a phylogenetic tree allowed the assessment of metabarcode taxonomic assignment.
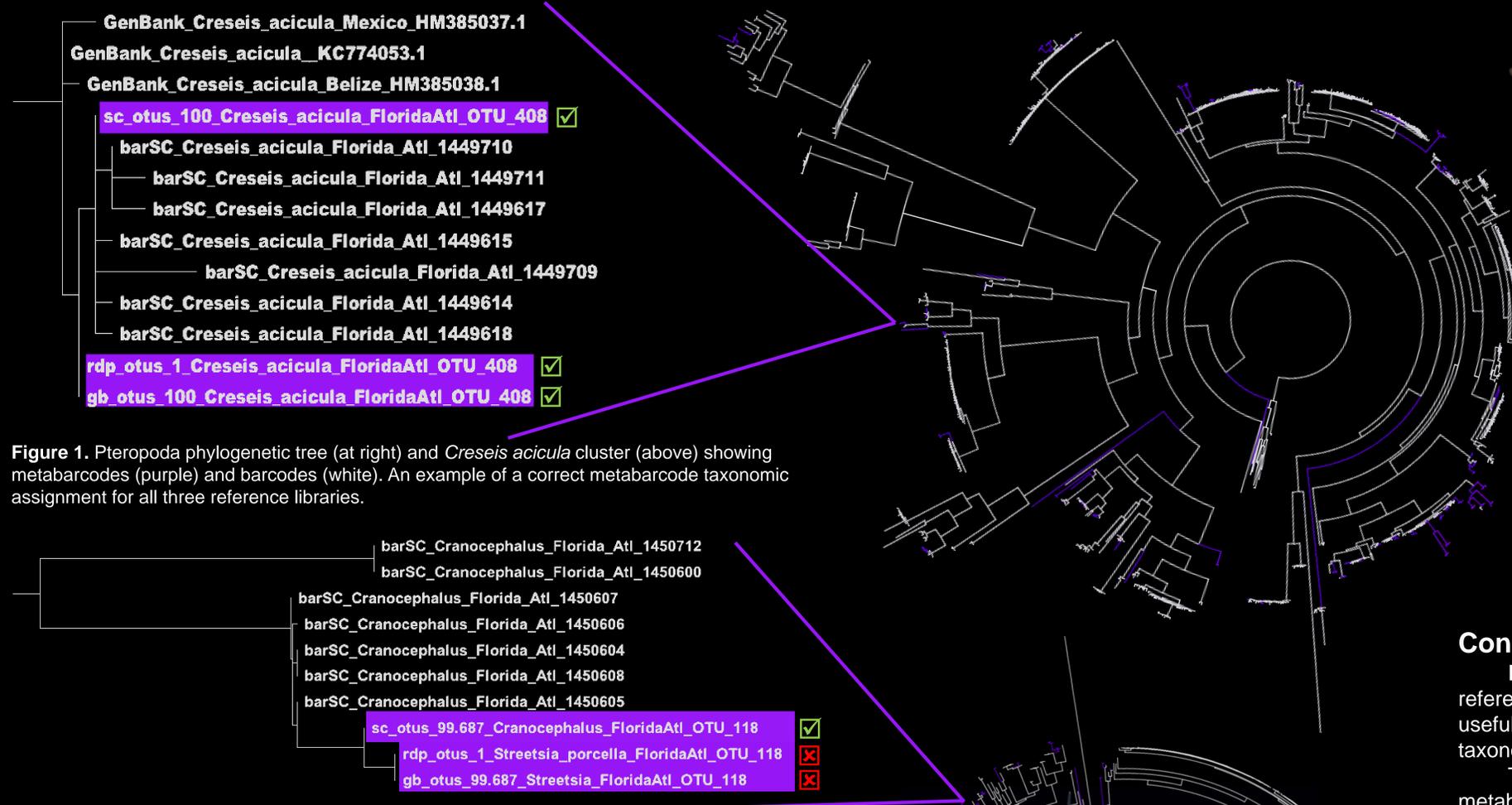
Methods:

- Pteropoda (Mollusca, Gastropoda) and Hyperiidea (Arthropoda, Amphipoda) were chosen as test cases because we have many unpublished barcodes available.
- Collected all available (many unpublished) zooplankton barcodes and metabarcodes assigned to species using 1. RDP Classifier- using the Midori curated reference library, 2. BLASTn-GenBank using all GenBank sequences, and 3. BLASTn-StreamCode using our own curated library.
- Sequence alignment using MAFFTT in Geneious v2021.1.1.
- Tree building using RAXML, with 1000 replicates, GTRGAMMA model, with partitions, run on the SI/HPC.
- Analyzed metabarcodes species assignment in each clade and categorized them as "Yes", "No Clear Evidence", and "No" based on their position on the tree.

Figure 1. Pteropoda phylogenetic tree (at right) and *Creseis acicula* cluster (above) showing metabarcodes (purple) and barcodes (white). An example of a correct metabarcode taxonomic assignment for all three reference libraries.
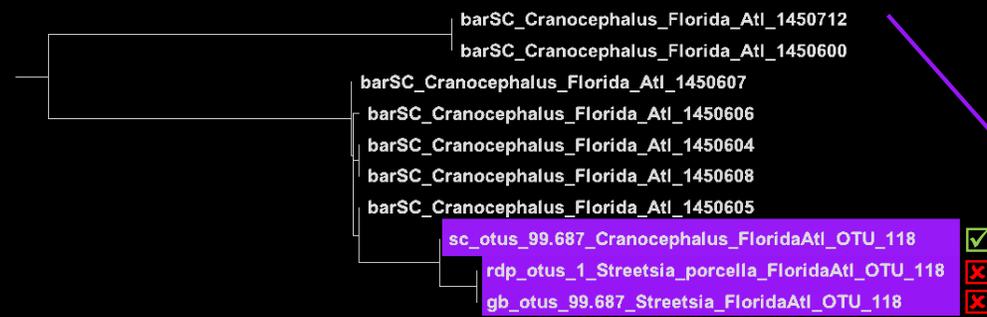

Figure 2. Hyperiidea phylogenetic tree (at right) and *Cranocephalus* cluster (above) showing metabarcodes (purple) and barcodes (white). Here is an example of two incorrect and one correct metabarcode taxonomic assignments.
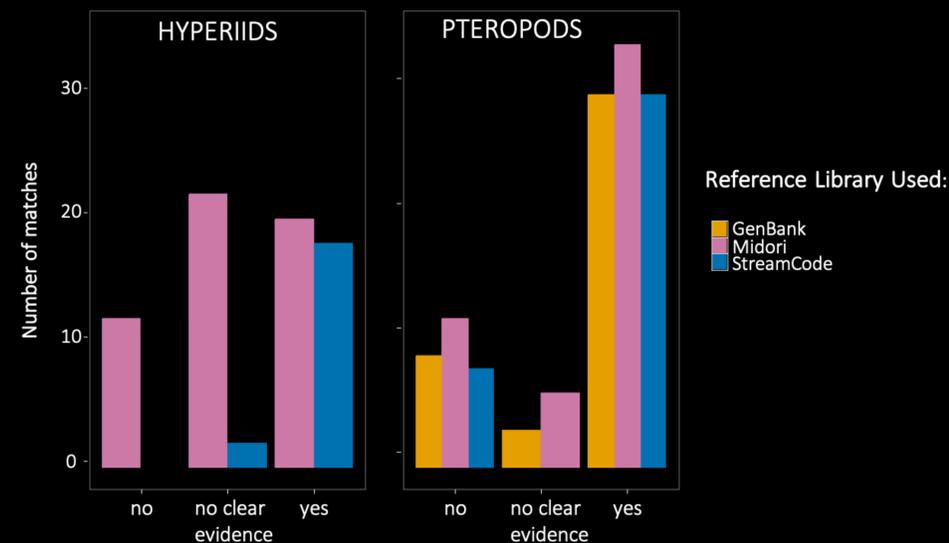

Figure 3. The number of metabarcode taxonomic assignments that were correctly assigned (Yes), incorrectly assigned (No), or have too little data to verify assignment (No Clear Evidence). For Hyperiids, StreamCode had fewer errors than Midori, but Midori had slightly more correct assignments than StreamCode. GenBank was not tested for hyperiids for the species level. For pteropods, Midori had the most errors in comparison to the two other reference libraries, but overall, there was no significant difference between the amount of correctly assigned metabarcodes.

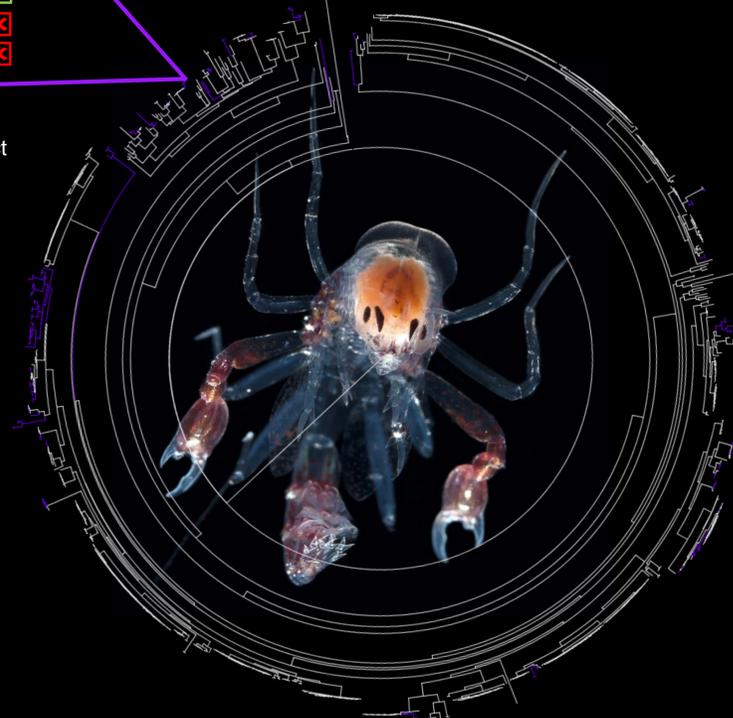**Figure 4.** The pteropod *Clione*

Figure 5. The hyperiid *Phronima*

Conclusion:

Including metabarcodes and all reference barcodes in a tree is a useful way to assess metabarcode taxonomic assignments.

Taxonomic assignment of metabarcode sequences has been shown to be highly dependent on the representation of each taxonomic group within a reference library[1]. Likely the difference in reference library impact was because pteropods are well represented in all of these databases, while hyperiids are not well represented only in the StreamCode reference library.

Based on this, a complete reference library is the most important thing for metabarcoding to be an effective tool.

I recommend focusing effort on improving the reference libraries by increasing the amount of good quality barcodes.

References:
1. Pappalardo P et al, 2021. The role of taxonomic expertise in interpretation of metabarcoding studies. ICES Journal of Marine Science