BBSPLIT SUPPLEMENTAL METHODS

BBSplit maps sequences to reference databases of sequences provided by the user. The output of BBSplit is a file of sequences that have mapped to each reference database and a file for all sequences that do not map to any reference databases.

We decided to build our BBSplit reference databases based on sequences we already knew to be a part of our dataset (i.e. the output of a previous denoising pipeline). Because dada2 provides more ASVs than unoise3 (Nearing et al. 2018), we chose to use the dada2 ASV list to create our reference databases. We split the output of the MEGAN taxonomic alignment into sequences that mapped to the predator (*Heteropoda venatoria*) and those that mapped to prey (Supplement R code). We know that our predator *H. venatoria* is the only member of its genus and family on Palmyra Atoll, and so sequences that matched these higher classifications in MEGAN (Genus: *Heteropoda* or Family: Sparassidae; Handler et al.) were also split into the predator reference file. We then split our dada2 ASV list based on whether sequences mapped to these predator or prey ASV lists from MEGAN. Our final result was a set of two reference files, one including all predator ASVs, and one containing all prey ASVs.

After we had built these reference databases of predator and prey ASVs, we ran the BBSplit program using both the predator and prey reference files to map our trimmed sequences. We kept defaults for most settings of the BBSplit command, including that reads that ambiguously mapped to both databases should go in the best fit database (the default “ambiguous” and “ambiguous2” parameters equal to “best”). We kept this default because we expected that most sequences would be predator sequences. However, because raw sequences or sequences with high error rates had not been denoised yet, would not fit perfectly to the ASV list for predators, but would fit more closely to this ASV list than to the ASVs in the prey reference file (Code Supplement/Data). The output of BBSplit was one set of sequences that mapped to the prey reference ASVs, one that mapped to predator ASVs, and one that did not map to either of these (unmapped).

Although we were most interested in running dada2 and unoise3 again on the prey ASVs split with BBSplit, we also wanted to ensure that the splitting process did an accurate job of removing predator ASVS (i.e. predator sequence file should all map to predator after dada2 and unoise3), and that we weren’t missing any prey in the prey sequence file (by looking at the ASV list of the unmapped sequence file after dada2 and unoise3). Therefore, we ran dada2 and unoise3 against each mapped set of sequences: the prey-mapped sequences, the predator-mapped sequences, and the unmapped sequences. As a result, we had a total of six more ASV lists and ASV tables matched to each sample (3 from Dada2 and 3 from unoise3 in USEARCH). We then used BLAST and a database of all nucleotide sequences on GenBank (downloaded on November 20, 2019) and the BOLD IDEngine (accessed February 5-16, 2020) to match taxonomies to each of these ASV files. Again, we selected the subtree in MEGAN with likely prey items (Kingdom:Animalia, Clade: Bilateria) and exported the same files. For BOLD, we again used the Species Level Barcode Records database.