Contribution ID: **154**   Type: **not specified**

Using AnnoTree to Get More Assignments, Faster, in DIAMOND+MEGAN Microbiome Analysis

Monday, November 28, 2022 1:30 PM (30 minutes)

In microbiome analysis, one main approach is to align metagenomic sequencing reads against a protein reference database, such as NCBI-nr, and then to perform taxonomic and functional binning based on the alignments. This approach is embodied, for example, in the standard DIAMOND+MEGAN analysis pipeline, which first aligns reads against NCBI-nr using DIAMOND and then performs taxonomic and functional binning using MEGAN. Here, we propose the use of the AnnoTree protein database, rather than NCBI-nr, in such alignment-based analyses to determine the prokaryotic content of metagenomic samples. We demonstrate a 2-fold speedup over the usage of the prokaryotic part of NCBI-nr and increased assignment rates, in particular assigning twice as many reads to KEGG. In addition to binning to the NCBI taxonomy, MEGAN now also bins to the GTDB taxonomy.

IMPORTANCE The NCBI-nr database is not explicitly designed for the purpose of microbiome analysis, and its increasing size makes its unwieldy and computationally expensive for this purpose. The AnnoTree protein database is only one-quarter the size of the full NCBI-nr database and is explicitly designed for metagenomic analysis, so it should be supported by alignment-based pipelines.

Presenter: GAUTAM, Anupam

Session Classification: Poster session