Challenges in capturing the mycobiome from shotgun metagenome data: lack of software and databases

Authors: Ekaterina Avershina[1], Arfa Irej Qureshi[2], Hanne C Winther-Larsen[2], Trine B Rounge[1,2,3]

Affiliations:

1. Department of Tumor Biology, Oslo University Hospital, Norway
2. Section for Pharmacology and Pharmaceutical Biosciences, Department of Pharmacy, University of Oslo, Norway
3. Department for Research, Cancer Registry of Norway, Norwegian Institute of Public Health - Cancer Registry of Norway, Norway

Corresponding author: Trine B. Rounge. e-mail: t.b.rounge@farmasi.uio.no

Supplementary Information

Supplementary Text 1. Script and database modifications required for FunOMIC

FunOMIC

1. follow the setup instructions on https://manichanh.vhir.org/funomic/

2. when using provided BacteriaDB index database, bowtie2 crushes. This can be fixed by creating another index database with bacteria of your choice. For example, HumGut (https://arken.nmbu.no/~larssn/humgut/)

   (build database using *bowtie2-build bacteria.fasta indexname --large-index*) and provide path to this index instead

3. (pandas versions conflict – if pandas v ≥ 1.3.0) - in file coverageFilter.py line 27 change *error_bad_lines=False* to *on_bad_lines='skip'*

4. (pandas versions conflict – if pandas v ≥ 1.4.0) - in file buglist.py comment out line 62 and paste the following instead:

```
counts_series = [
        kingdom_counts.kingdom,
        phylum_counts.combined,
        class_counts.combined,
        order_counts.combined,
        family_counts.combined,
        genus_counts.combined,
        species_counts.combined]

        abundance_series = [
        kingdom_counts.Abundance,
        phylum_counts.Abundance,
        class_counts.Abundance,
        order_counts.Abundance,
        family_counts.Abundance,
        genus_counts.Abundance,
        species_counts.Abundance]

        counts_series = pd.concat(counts_series)
        abundance_series = pd.concat(abundance_series)
        new = pd.concat([counts_series, abundance_series], axis=1)
```

Supplementary Text 2. Script modifications and unlisted dependencies required for MycobiomeScan 2.0 (HMS)

1. Download the tool from https://sourceforge.net/projects/hmscan/

2. After download, all the scripts are located in MScan2/tools. However, in MScan.sh and all related files, paths to the scripts are provided as MScan2/HMS/tools - therefore, create 'HMS' folder inside MScan2 and move all the files and folders into that folder

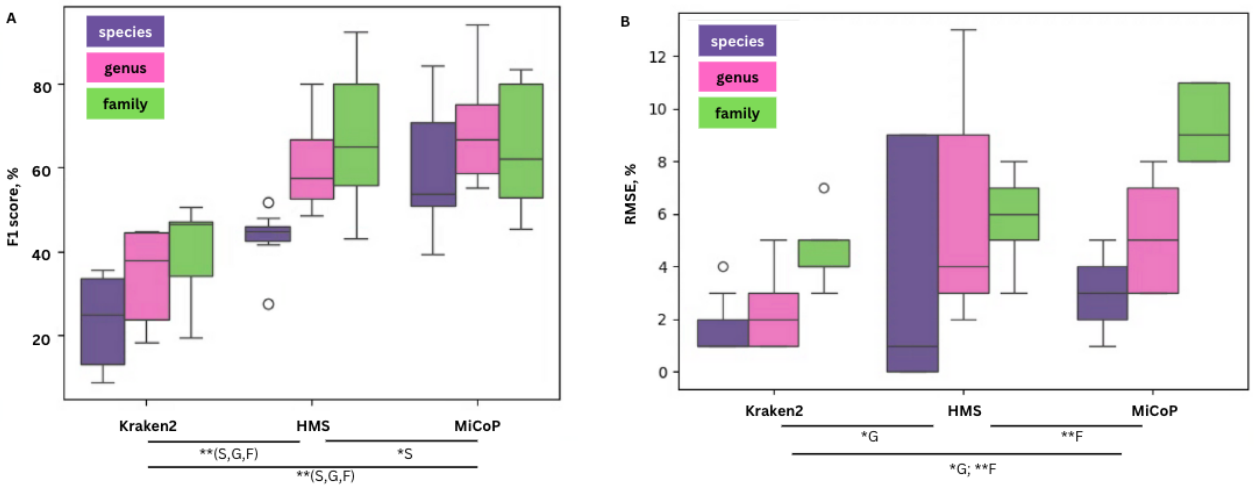
3. install Java

4. In R, install *ggplot2* and *stringr* libraries

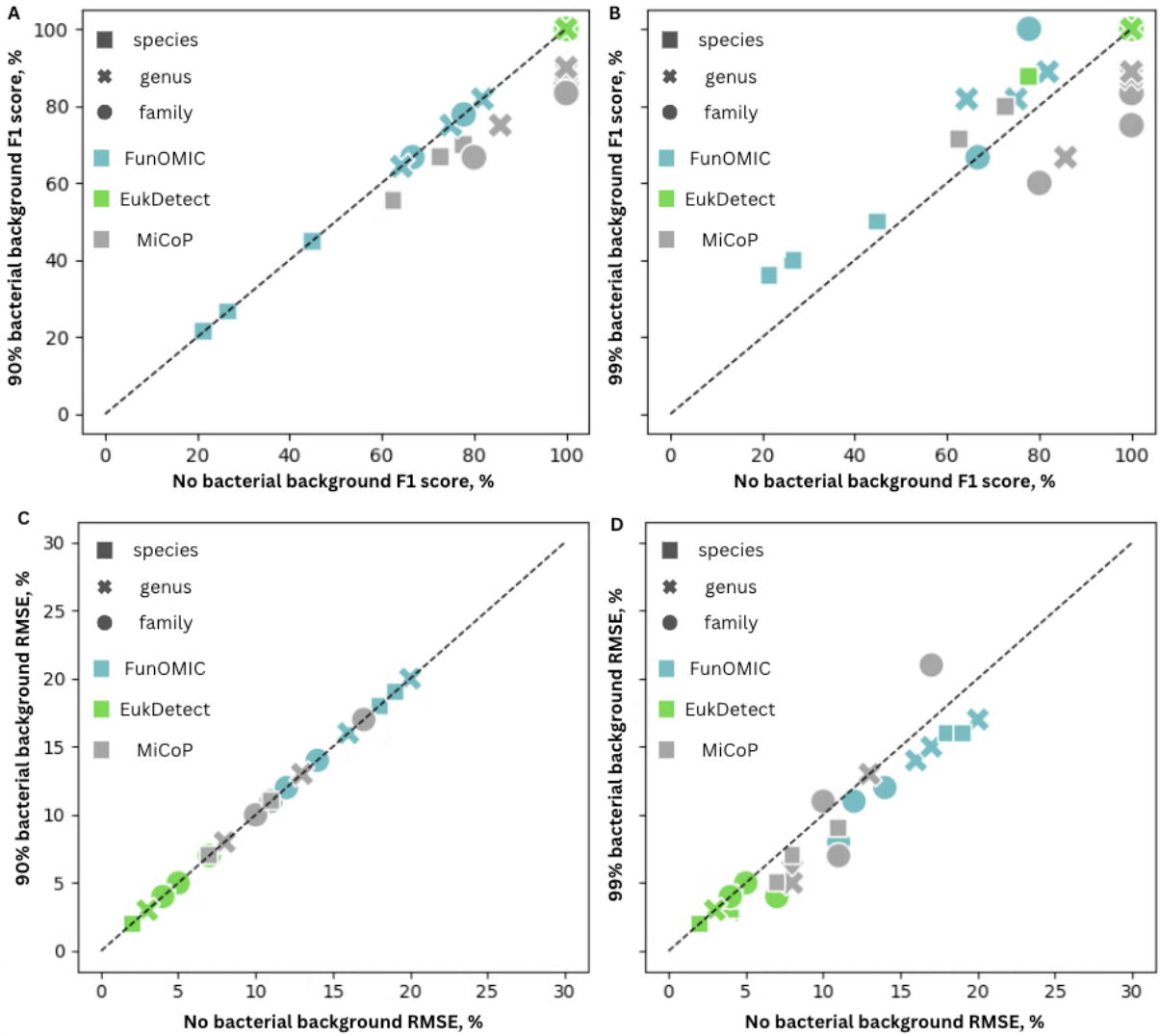
5. Download Normalizing_table.txt file from the first version of HMS and save the file in MScan2/HMS/var folder

6. In /HMS/tools/bmtagger.sh update to local paths in lines 38, 40 and 41.

7. In HMS/MScan.sh, change the output parameter in line 86 (first use of bmtagger.sh) from *-o $OUTPUT_DIR/$OUTPUT_DIR-filter-human* to *-o $OUTPUT_DIR/$OUTPUT_DIR-filter-bacteria*
   (In this line, the reads are filtered from bacterial reads and should be save into the -bacteria file which will further be used as an input in line 88 and filtered for human reads)



Supplementary Figure 1. Identification accuracy (A) and RMSE (B) by Kraken2, HMS and MiCoP using the same fungal RefSeq reference database on a species (purple), genus (pink) and family (green) levels



Supplementary Figure 2. Identification accuracy (A, B) and RMSE (C, D) by FunOMIC (blue), MiCoP (grey) and EukDetect (green) on a species (square), genus (x) and family (circle) levels in mock communities with 10 fungal species and bacterial background at 90 % (A, C) and 99 % (B, D) level (y-axis) vs no bacterial background (x-axis)