**SUPPLEMENTAL MATERIAL**

**Compositional analysis**

Here, we outline the commands used for each species classifier, in addition to PanPhlAn, and we describe how these parameters deviated from the default settings. Commands are highlighted in grey.

**CLARK**

ls \*.fasta | awk -F '.fasta' '{print "classify\_metagenome.sh -O "$0" -R "$0".clark\_out -m 0"}' > run\_CLARK.sh

sh run\_CLARK.sh

ls \*.csv | awk -F '.csv' '{print "/shared/software/clarke/1.2.3/estimate\_abundance.sh -F "$0" -D $DIR\_DB -a 0.1 -c 1 -g 0.05 > "$0".abundances"}' > run\_CLARK\_Abundances.sh

sh run\_CLARK\_Abundances.sh

Description: The CLARK classification step was run with full mode execution. The CLARK estimate abundances step was run with minAbundance 0.1, minConfidenceScore 1, and minGamma 0.05.

**Kaiju**

ls \*.fa | awk -F '.fa' '{print "/shared/software/kaiju/kaiju/bin/kaiju -t /data/tgsc1/aaron/kaijudb/nr/nodes.dmp -f /data/tgsc1/aaron/kaijudb/nr/kaiju\_db.fmi -i "$0" -o "$0"\_kaiju.nr.out -z 10 –m 33 -x -v"}' > run\_Kaiju.sh

sh run\_Kaiju.sh

ls \*.nr.out | awk -F '.nr.out' '{print "kaijuReport -u -m 0.1 -t /data/tgsc1/aaron/kaijudb/nr/nodes.dmp -n /data/tgsc1/aaron/kaijudb/nr/names.dmp -r species -i "$0" -o "$0".species.summary"}' > Kaiju\_Report.sh

sh Kaiju\_Report.sh

Description: The Kaiju classification step was run using the SEG low complexity filter, and the minimum match length was set to 22. Reads were mapped against the RefSeq database. The Kaiju report step was run using minAbundance 0.1. Only classified reads were reported.

**Kraken**

ls \*.fasta | awk -F '.fasta' '{print "kraken --threads 10 --preload --db $KRAKEN\_DIR/krakken\_db "$0" > "$0"\_kraken\_out"}' > run\_Kraken.sh

sh run\_Kraken.sh

ls \*kraken\_out | awk -F 'kraken\_out' '{print "kraken-filter --db $KRAKEN\_DIR/krakken\_db --threshold 0.5 "$0" > "$0".filtered"}' > run\_Kraken\_Filter.sh

sh run\_Kraken\_Filter.sh

ls \*filtered | awk -F 'filtered' '{print "kraken-mpa-report --db $KRAKEN\_DIR/krakken\_db "$0" > "$0"\_mpa"}' > run\_Kraken\_Report.sh; sh run\_Kraken\_Report.sh

Description: Kraken results were filtered using a threshold set to 0.5 to remove low confidence classifications.

**MetaPhlAn2**

MetaPhlAn2 was run using default parameters (<https://bitbucket.org/biobakery/biobakery/wiki/metaphlan2>).

**SLIMM**

ls \*trimmed.fastq | awk -F 'trimmed.fastq' '{print "bowtie2 -x $DB\_DIR/slimm\_db/AB\_5K\_indexed\_ref\_genomes\_bowtie2/AB\_5K -U "$0" | samtools view -bSF4 - > "$1"\_mapped\_reads.bam"}' > bowtie2\_map.sh

sh bowtie2\_map.sh

ls \*bam | awk -F 'bam' '{print "slimm -m $DB\_DIR/slimm\_db/slimmDB\_5K "$0""}' > run\_SLIMM.sh

sh run\_SLIMM.sh

Description: Bowtie 2 ([1](#_ENREF_1)) was used to trimmed fastq reads against the slimmDB\_5K reference database.

**PanPhlAn**

ls -d \*.fasta | awk -F '.fasta' '{print "/shared/software/panphlan/b40c003/panphlan\_map.py -c $pangenome\_index -i "$0" -o map\_results"}' > run\_PanPhlAn.sh

sh run\_PanPhlAn.sh

panphlan\_profile.py -c $pangenome\_index -i map\_results --add\_strains --min\_coverage 1 --left\_max 1.70 --right\_min 0.30 --o\_dna result\_gene\_presence\_absence.csv --strain\_hit\_genes\_perc percent\_match.txt

Description: The PanPhlAn profiling step was run with a –min\_coverage set to 1, --left\_max set to 1.7, and –right\_min set to 0.3. These parameters increase the tool’s sensitivity.

**REFERENCES**

1. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nature methods **9:**357-359.