**Supplementary Methods and scripts used for Ublast\_blastx approach**

Methods based on original article by (Yang *et al.*, 2014)

**Quick search for candidate sequences from MRG database using USEARCH *ublast* command** (Edgar, 2010)**:**

ls \*.fa | parallel –gnu “usearch -ublast {} -db “your\_database.fasta” -evalue 1e-5 -accel 0.5 -blast6out {/.}.ublast\_output\_file.txt -threads 4"

**Extract sequence ids contained in column 1 from candidate sequence output files from *ublast***

awk '{if(FILENAME!=prev){close(prev)};print $1 > (FILENAME".ids.txt");prev=FILENAME}' \*.txt (or whatever file name is from previous step)

**Extract ublasted candidate sequences from original fastas** <https://bioinf.shenwei.me/seqkit/usage/>**:**

parallel 'cat {} | seqkit grep -f {.}.ids.txt > {.}.output.fa' ::: \*fa

**Create a protein database for use with blast:**

makeblastdb -in BacMet2\_EXP\_database.fasta -dbtype prot -input\_type fasta -out db/blast/BacMet2\_aa -hash\_index

**Create a protein database for use with Diamond blastx** (Buchfink *et al.*, 2015)**:**

diamond makedb --in BacMet2\_EXP\_database.fasta -d BacMet2

**Diamond blastx all candidate sequences originally derived from USEARCH ublast:**

ls \*.fa | parallel --gnu "diamond blastx -q {} -d …/BacMet2.dmnd --max-target-seqs 1 --threads 4 --evalue 0.000001 -o {/.}.diamond.tab"

awk '{if(FILENAME!=prev){close(prev)};print $1 > (FILENAME".diamond\_blastx\_ids.txt");prev=FILENAME}' \*.tab

**Extract final target sequences:**

awk '{if(FILENAME!=prev){close(prev)};print $1 > (FILENAME".ids.txt");prev=FILENAME}' \*.tab

**Get counts of each gene per read**

ls \*.tab | parallel --gnu "cut -f 2 {} > {/.}.outputfile.txt"

ls \*.txt | parallel --gnu "sort {} | uniq -c | sort -n > {/.}.ids\_counts"

**Get sequence ids from original BacMet and save to file**

grep '>' …/BacMet2\_EXP\_database.fasta > file.txt

**Prodigal**

ls \*.fasta | parallel prodigal -i {} -o {.}.genes -a proteins{.}.faa -p meta

**Use BBMAP to split large FASTA files for usearch 32-bit version (**<https://jgi.doe.gov/data-and-tools/bbtools/>)

ls \*.fasta| parallel --gnu "partition.sh in={} out=out\_dir/{/.}part%.fasta ways=5"

**Recombine or concatenate usearch-blastx search results**:

1. touch filename\_finish.txt
2. cat file\_name\_mergedpart\* >> filename\_finish.txt

**References:**

Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**: 59–60.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Yang, Y., Jiang, X.-T., and Zhang, T. (2014) Evaluation of a Hybrid Approach Using UBLAST and BLASTX for Metagenomic Sequences Annotation of Specific Functional Genes. *PLoS ONE* **9**:.