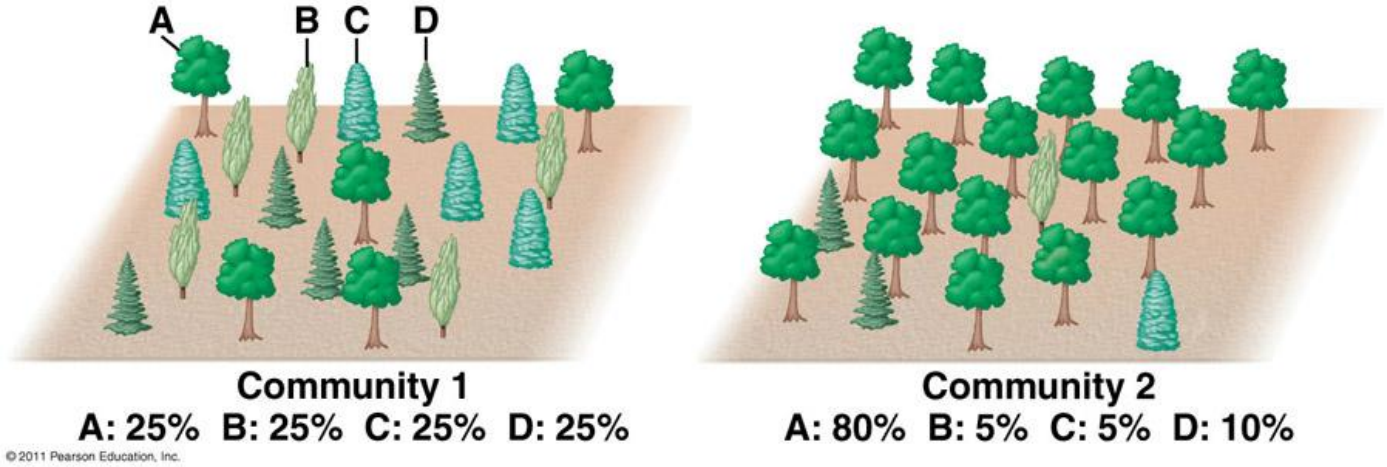
# Abundance Estimation: MetaPhlAn2 and Quikr

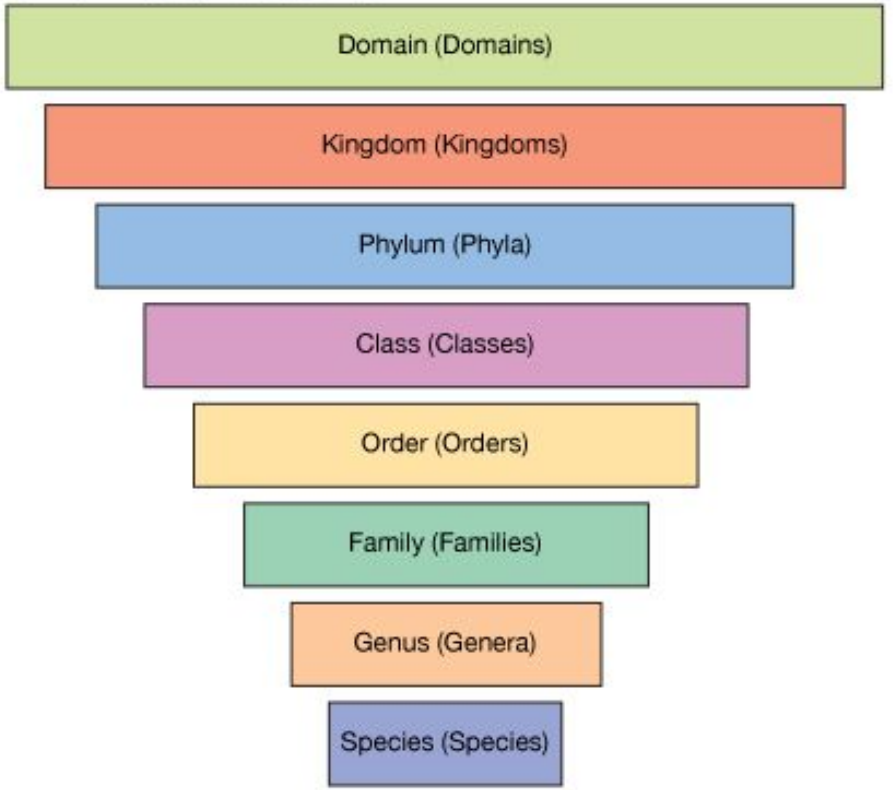
By: Moon Kim and Ariana Entezari

# Abundance Estimation

- Relative abundance - relative representation of a species in an ecosystem
- Species abundance - number of individuals per species
- Species richness - number of species in a community

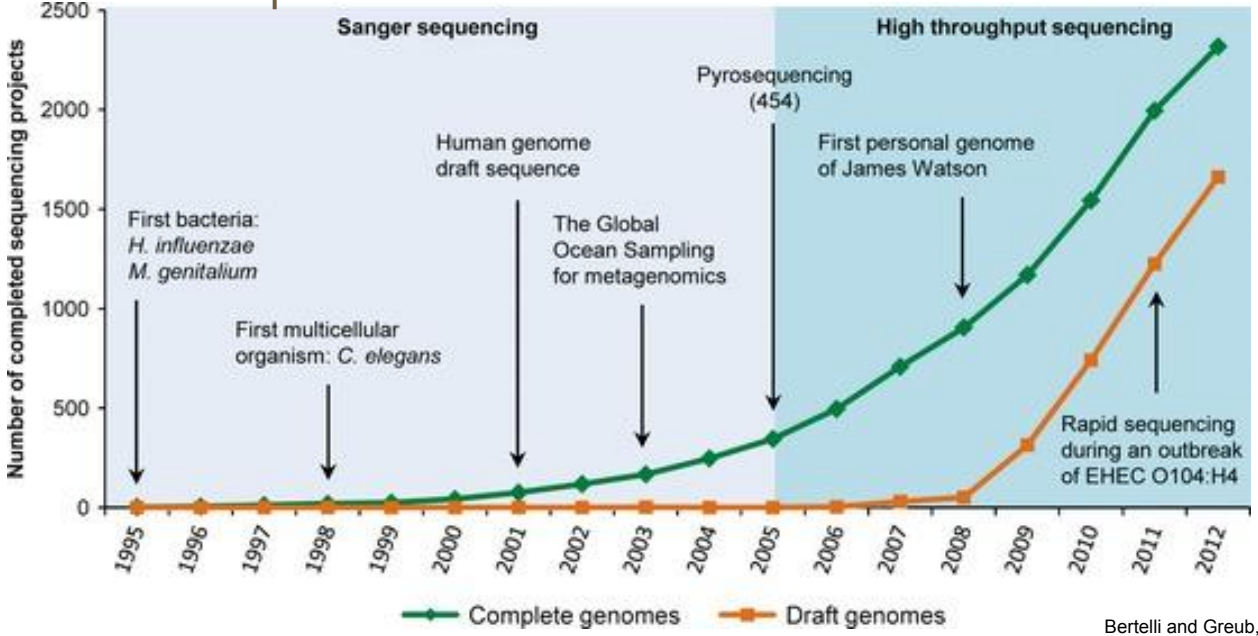


# Taxonomic Classifications



# Urban Dataset

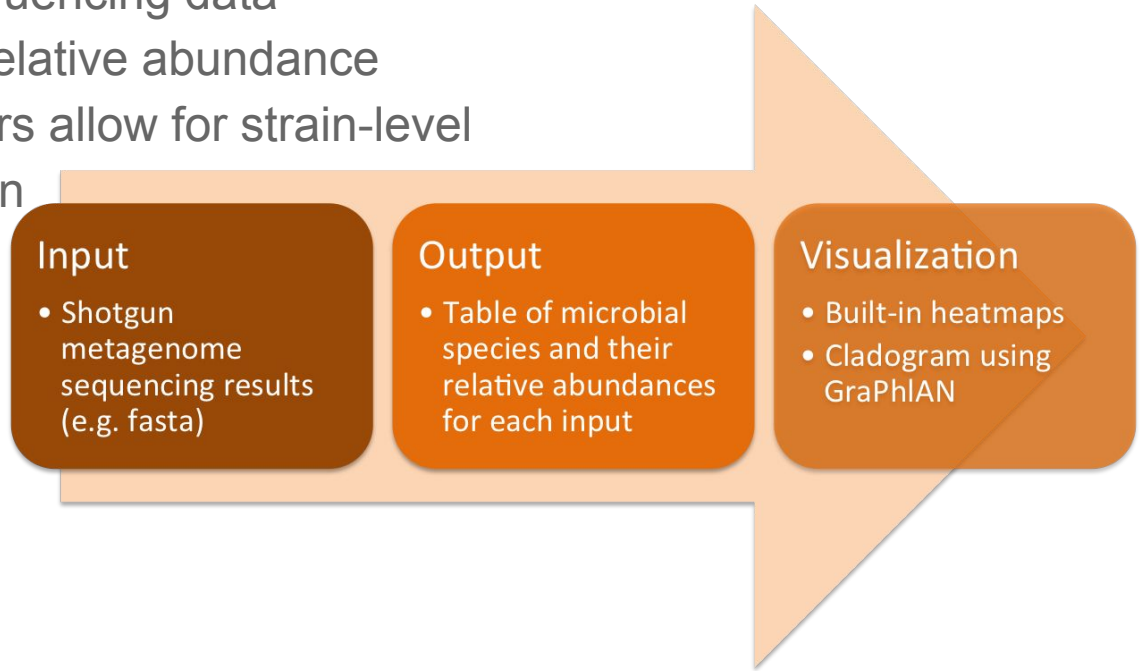
- Whole genome shotgun sequencing data - the whole genome is cut into small DNA fragments that are sequenced and reassembled based on overlap between sequences




Bertelli and Greub, 2013

# MetaPhlAn2

- Profiles the composition of microbial communities from metagenomic shotgun sequencing data
- Outputs an estimation of relative abundance
- Further subspecies markers allow for strain-level analyses using StrainPhlAn




https://bitbucket.org/biobakery/biobakery/wiki/metaphlan2

# MetaPhlAn2

- Relies on unique clade-specific marker genes identified from about 17,000 reference genomes
  - ~13,500 bacterial and archaeal, ~3,500 viral, and ~110 eukaryotic reference genomes
- Infers the presence and read coverage of about 1 million clade-specific markers from over 7,500 species to detect the taxonomy in a microbiome sample and estimate relative abundance
- Analysis speed of 25,000 reads-per-second

# Using MetaPhlAn2 on Proteus

- Clone the repository:
  - $ hg clone https://bitbucket.org/biobakery/metaphlan2
- Check that you have the prerequisites:
  - Python 2.7 or newer
  - BowTie2 (used by MetaPhlAn2 for aligning sequence reads)
- Basic command line usage:
  - metaphlan2.py metagenome.fastq --input_type fastq > profiled_metagenome.txt
- For large datasets, submit a job on Proteus
  - qsub file_name

# Using MetaPhlAn2 on Proteus

```
usage: metaphlan2.py --input_type
                     {fastq,fasta,multifasta,multifastq,bowtie2out,sam}
                     [--mpa_pkl MPA_PKL] [--bowtie2db METAPHLAN_BOWTIE2_DB]
                     [--bt2_ps BowTie2 presets] [--bowtie2_exe BOWTIE2_EXE]
                     [--bowtie2out FILE_NAME] [--no_map] [--tmp_dir]
                     [--tax_lev TAXONOMIC_LEVEL] [--min_cu_len]
                     [--min_alignment_len] [--ignore_viruses]
                     [--ignore_eukaryotes] [--ignore_bacteria]
                     [--ignore_archaea] [--stat_q]
                     [--ignore_markers IGNORE_MARKERS] [--avoid_disqm]
                     [--stat] [-t ANALYSIS TYPE] [--nreads NUMBER_OF_READS]
                     [--pres_th PRESENCE_THRESHOLD] [--clade] [--min_ab] [-h]
                     [-o output file] [--sample_id_key name]
                     [--sample_id value] [-s sam_output_file]
                     [--biom biom_output] [--mdelim mdelim] [--nproc N] [-v]
                     [INPUT_FILE] [OUTPUT_FILE]
```

- input_type
- bowtie2db
- bowtie2_exe
- bowtie2out
- tax_lev
- ignore_eukaryotes, ignore_viruses, etc.
- ignore_markers
- nproc

# Output of MetaPhlAn2

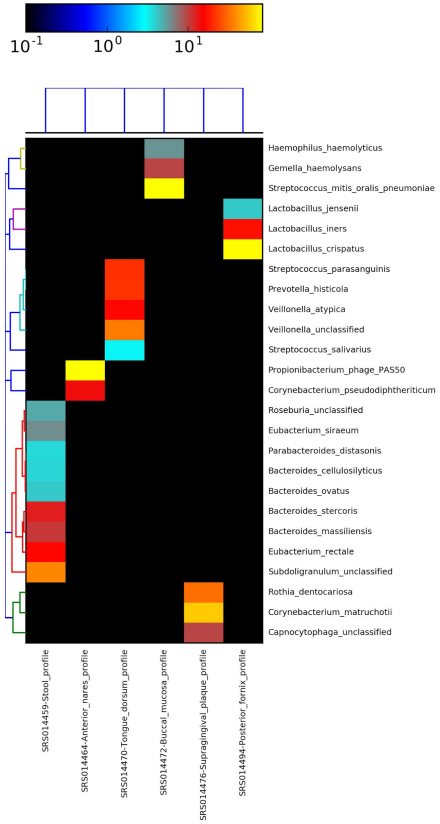
- Output .txt files of relative abundance for each sample
- Can be merged with python code that is part of MetaPhlAn2 package
  - $ python utils/merge_metaphlan_tables.py metaphlan_output1.txt metaphlan_output2.txt > output/merged_abundance_table.txt

```
#SampleID       Metaphlan2_Analysis
k__Bacteria     100.0
k__Bacteria|p__Firmicutes       64.91753
k__Bacteria|p__Bacteroidetes    35.08247
k__Bacteria|p__Firmicutes|c__Clostridia 64.91753
k__Bacteria|p__Bacteroidetes|c__Bacteroidia     35.08247
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales        64.91753
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales    35.08247
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Ruminococcaceae     37.7397
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae  31.34317
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Eubacteriaceae      22.08827
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae     5.08956
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Porphyromonadaceae      3.7393
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Ruminococcaceae|g__Subdoligranulum  37.7397
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides   31.34317
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Eubacteriaceae|g__Eubacterium       22.08827
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Roseburia        5.08956
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Porphyromonadaceae|g__Parabacteroides   3.7393
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Ruminococcaceae|g__Subdoligranulum|s__Subdoligranulum_unclassified  37.7397
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Eubacteriaceae|g__Eubacterium|s__Eubacterium_rectale       16.00116
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_stercoris  12.82765
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_massiliensis      10.61295
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Eubacteriaceae|g__Eubacterium|s__Eubacterium_siraeum       6.08711
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Roseburia|s__Roseburia_unclassified      5.08956
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_ovatus     4.08051
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_cellulosilyticus    3.82206
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Porphyromonadaceae|g__Parabacteroides|s__Parabacteroides_distasonis     3.7393
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Eubacteriaceae|g__Eubacterium|s__Eubacterium_rectale|t__Eubacterium_rectale_unclassified    16.00116
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_stercoris|t__Bacteroides_stercoris_unclassified
12.82765
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_massiliensis|
t__Bacteroides_massiliensis_unclassified        10.61295
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Eubacteriaceae|g__Eubacterium|s__Eubacterium_siraeum|t__Eubacterium_siraeum_unclassified    6.08711
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_ovatus|t__Bacteroides_ovatus_unclassified  4.08051
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_cellulosilyticus|
t__Bacteroides_cellulosilyticus_unclassified    3.82206
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Porphyromonadaceae|g__Parabacteroides|s__Parabacteroides_distasonis|
t__Parabacteroides_distasonis_unclassified      3.7393
```

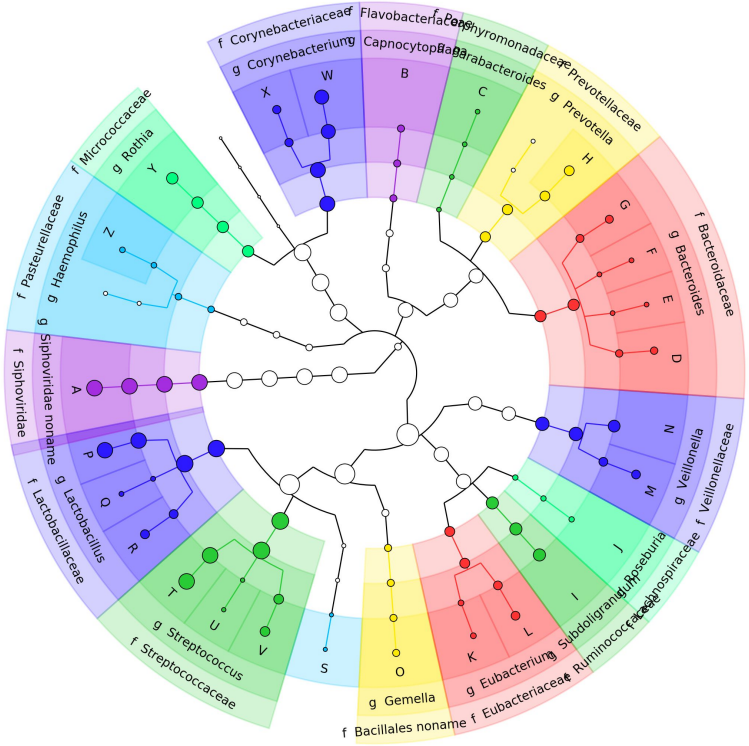
| | ID | SRR3545898 | SRR3545910 |
|---|---|---|---|
| 2 | k__Archaea | 0.03389 | 0 |
| 3 | k__Archaea\|p__Euryarchaeota | 0.03389 | 0 |
| 4 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria | 0 | 0 |
| 5 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales | 0 | 0 |
| 6 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales\|f__Halobacteriaceae | 0 | 0 |
| 7 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales\|f__Halobacteriaceae\|g__Halobacterium | 0 | 0 |
| 8 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales\|f__Halobacteriaceae\|g__Halobacterium\|s__Halobacterium_unclassified | 0 | 0 |
| 9 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales\|f__Halobacteriaceae\|g__Halobiforma | 0 | 0 |
| 10 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales\|f__Halobacteriaceae\|g__Halobiforma\|s__Halobiforma_unclassified | 0 | 0 |
| 11 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales\|f__Halobacteriaceae\|g__Halococcus | 0 | 0 |
| 12 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales\|f__Halobacteriaceae\|g__Halococcus\|s__Halococcus_unclassified | 0 | 0 |
| 13 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales\|f__Halobacteriaceae\|g__Natrialba | 0 | 0 |
| 14 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales\|f__Halobacteriaceae\|g__Natrialba\|s__Natrialba_unclassified | 0 | 0 |
| 15 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales\|f__Halobacteriaceae\|g__Natronococcus | 0 | 0 |
| 16 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales\|f__Halobacteriaceae\|g__Natronococcus\|s__Natronococcus_unclassified | 0 | 0 |
| 17 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales\|f__Halobacteriaceae\|g__Natronorubrum | 0 | 0 |
| 18 | k__Archaea\|p__Euryarchaeota\|c__Halobacteria\|o__Halobacteriales\|f__Halobacteriaceae\|g__Natronorubrum\|s__Natronorubrum_unclassified | 0 | 0 |
| 19 | k__Archaea\|p__Euryarchaeota\|c__Methanobacteria | 0 | 0 |
| 20 | k__Archaea\|p__Euryarchaeota\|c__Methanobacteria\|o__Methanobacteriales | 0 | 0 |
| 21 | k__Archaea\|p__Euryarchaeota\|c__Methanobacteria\|o__Methanobacteriales\|f__Methanobacteriaceae | 0 | 0 |
| 22 | k__Archaea\|p__Euryarchaeota\|c__Methanobacteria\|o__Methanobacteriales\|f__Methanobacteriaceae\|g__Methanobrevibacter | 0 | 0 |
| 23 | k__Archaea\|p__Euryarchaeota\|c__Methanobacteria\|o__Methanobacteriales\|f__Methanobacteriaceae\|g__Methanobrevibacter\|s__Methanobrevibacter_unclassified | 0 | 0 |
| 24 | k__Archaea\|p__Euryarchaeota\|c__Methanococci | 0.03389 | 0 |
| 25 | k__Archaea\|p__Euryarchaeota\|c__Methanococci\|o__Methanococcales | 0.03389 | 0 |
| 26 | k__Archaea\|p__Euryarchaeota\|c__Methanococci\|o__Methanococcales\|f__Methanocaldococcaceae | 0.03389 | 0 |
| 27 | k__Archaea\|p__Euryarchaeota\|c__Methanococci\|o__Methanococcales\|f__Methanocaldococcaceae\|g__Methanocaldococcaceae_unclassified | 0.03389 | 0 |
| 28 | k__Bacteria | 93.81197 | 91.14961 |
| 29 | k__Bacteria\|p__Acidobacteria | 0.08657 | 0.40452 |
| 30 | k__Bacteria\|p__Acidobacteria\|c__Acidobacteriia | 0.08657 | 0.40452 |
| 31 | k__Bacteria\|p__Acidobacteria\|c__Acidobacteriia\|o__Acidobacteriales | 0.08657 | 0.40452 |
| 32 | k__Bacteria\|p__Acidobacteria\|c__Acidobacteriia\|o__Acidobacteriales\|f__Acidobacteriaceae | 0.08657 | 0.40452 |
| 33 | k__Bacteria\|p__Acidobacteria\|c__Acidobacteriia\|o__Acidobacteriales\|f__Acidobacteriaceae\|g__Acidobacteriaceae_unclassified | 0 | 0.30263 |
| 34 | k__Bacteria\|p__Acidobacteria\|c__Acidobacteriia\|o__Acidobacteriales\|f__Acidobacteriaceae\|g__Granulicella | 0.08657 | 0.1019 |
| 35 | k__Bacteria\|p__Acidobacteria\|c__Acidobacteriia\|o__Acidobacteriales\|f__Acidobacteriaceae\|g__Granulicella\|s__Granulicella_unclassified | 0.08657 | 0.1019 |
| 36 | k__Bacteria\|p__Acidobacteria\|c__Acidobacteriia\|o__Acidobacteriales\|f__Acidobacteriaceae\|g__Terriglobus | 0 | 0 |
| 37 | k__Bacteria\|p__Acidobacteria\|c__Acidobacteriia\|o__Acidobacteriales\|f__Acidobacteriaceae\|g__Terriglobus\|s__Terriglobus_unclassified | 0 | 0 |
| 38 | k__Bacteria\|p__Actinobacteria | 90.68817 | 60.7167 |
| 39 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria | 90.68817 | 60.7167 |
| 40 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales | 90.23097 | 55.39621 |
| 41 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae | 0.19683 | 0.46461 |
| 42 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinobaculum | 0 | 0 |
| 43 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinobaculum\|s__Actinobaculum_unclassified | 0 | 0 |
| 44 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces | 0.19683 | 0.46461 |
| 45 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces\|s__Actinomyces_georgiae | 0 | 0 |
| 46 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces\|s__Actinomyces_georgiae\|t__GCF_000277685 | 0 | 0 |
| 47 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces\|s__Actinomyces_graevenitzii | 0 | 0 |
| 48 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces\|s__Actinomyces_graevenitzii\|t__Actinomyces_graevenitzii_unclassified | 0 | 0 |
| 49 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces\|s__Actinomyces_johnsonii | 0 | 0 |
| 50 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces\|s__Actinomyces_johnsonii\|t__Actinomyces_johnsonii_unclassified | 0 | 0 |
| 51 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces\|s__Actinomyces_massiliensis | 0 | 0 |
| 52 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces\|s__Actinomyces_massiliensis\|t__Actinomyces_massiliensis_unclassified | 0 | 0 |
| 53 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces\|s__Actinomyces_naeslundii | 0.00774 | 0.07167 |
| 54 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces\|s__Actinomyces_naeslundii\|t__GCF_000285995 | 0.00774 | 0.07167 |
| 55 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces\|s__Actinomyces_neuii | 0 | 0.01052 |
| 56 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces\|s__Actinomyces_neuii\|t__GCF_000296485 | 0 | 0.01052 |
| 57 | k__Bacteria\|p__Actinobacteria\|c__Actinobacteria\|o__Actinomycetales\|f__Actinomycetaceae\|g__Actinomyces\|s__Actinomyces_odontolyticus | 0.04211 | 0.0635 |

# Further Visualization

## Heatmap




## GraPhlAn



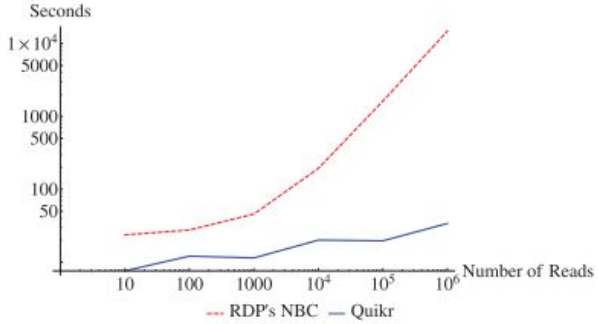
# Quikr

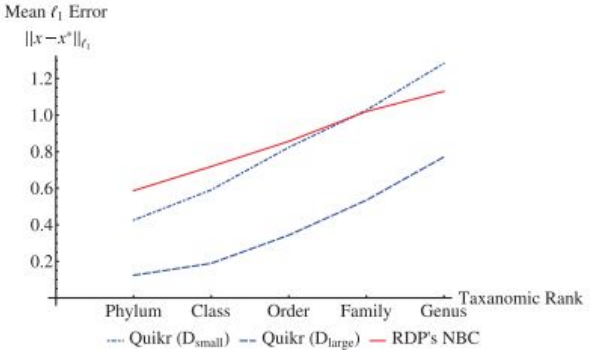
- Quadratic K-mer based iterative reconstruction method
- Developed by Dr. Rosen
- Reconstructs all taxonomic concentrations of a bacterial community simultaneously - as opposed to read by read classification. Leads to much improved runtimes over traditional methods.
- Unaffected by the presence of chimeras

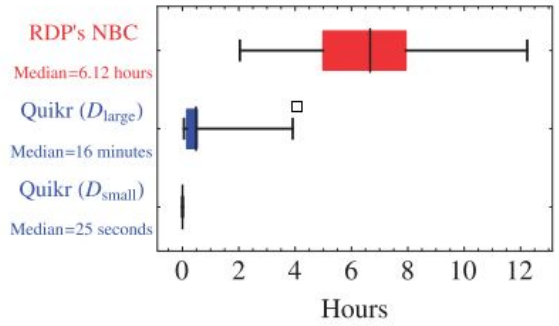
# Sensing Matrix

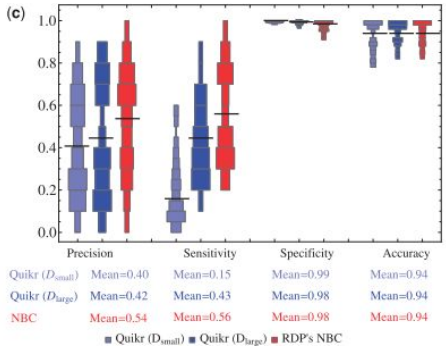
- Frequency of k-mers in a 16s database, and then reconstructs the concentration of the bacteria by solving an undetermined system of linear equations under a sparsity assumption.
- To solve the linear equation, a typical non negative least squares method is employed (MATLAB)
- Assumption that a sample will not contain any bacteria that is not in the database.

# Quikr Metrics



# Using Quikr

- Package available  in MATLAB, Octave, Python, and C
- Git clone https://github.com/EESI/quikr.git
- For use with MATLAB, need dna_utils (https://github.com/EESI/dna-utils.git)
- GCC 4.7 or newer - not available on Proteus
- Recommend using Quikr on a machine with SUDO privileges

# Using Quikr

- Create the sensing matrix first
  - quikr_train -i ~/gg_13_5_otus/rep_set/97_otus.fasta -o 97_sensing.matrix.gz -k 6 -v
  - Green Genes 97%
- Run Quikr
  - multifasta_to_otu -i ~/urban/fastas -s ~/urban/97_sensing.matrix.gz -k 6 -l 10000 -j 60 -o otus.txt -v
  - OTU table compatible with QIIME

```
multifasta_to_otu's arguments:
-i, --input-directory the directory containing the samples' fasta files of
reads (note each file should correspond to a separate sample)
-f, --sensing-fasta location of the fasta file database used to create the sensing matrix (fasta format)
-s, --sensing-matrix location of the sensing matrix. (sensing from quikr_train)
-k, --kmer specify what size of kmer to use. (default value is 6)
-l, --lambda lambda value to use. (default value is 10000)
-j, --jobs specifies how many jobs to run at once. (default value is the number of CPUs)
-o, --output the OTU table, with NUM_READS_PRESENT for each sample which
is compatible with QIIME's convert_biom.py (or a sequence table if not OTU's)
-v, --verbose verbose mode.
-V, --version  print version.
```

# Quikr's OTU Table

- Does not give taxonomy
- Raw count given

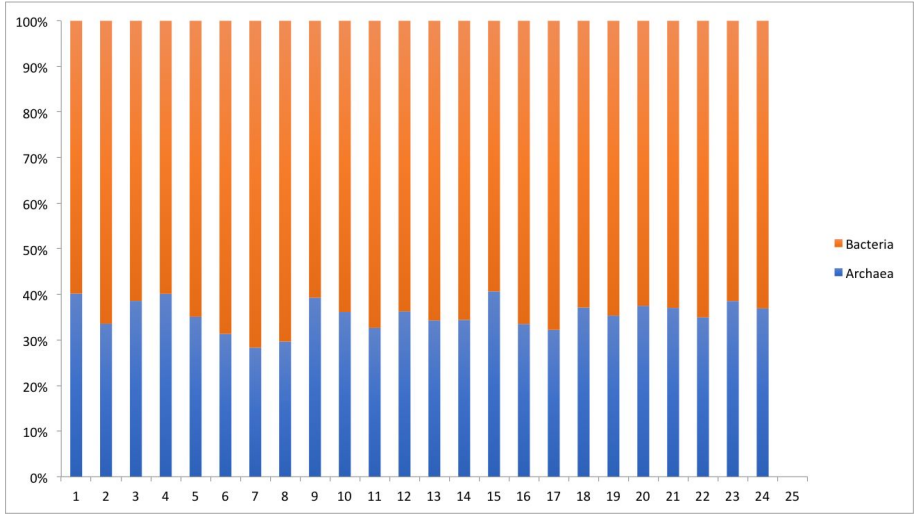
| #OTU_ID | SRR3546365.fna | SRR3545955.fna | SRR3546356.fna | SRR3546373.fna | SRR3546361.fna | SRR3546382.fna | SRR35463 |
|---|---|---|---|---|---|---|---|
| 1107928 | 167928 | 0 | 9252 | 22202 | 0 | 0 | 0 | 0 | 6580 | 0 | 0 | 0 | 0 | 0 | 12634 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1052239 | 207797 | 0 | 16921 | 25805 | 0 | 0 | 0 | 0 | 5238 | 0 | 0 | 0 | 0 | 0 | 14127 | 0 | 0 | 0 | 0 | 0 | 0 |
| 654240 | 0 | 0 | 0 | 0 | 8502 | 13720 | 73690 | 23749 | 0 | 0 | 6092 | 0 | 3395 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 628601 | 52454 | 0 | 12828 | 9330 | 0 | 0 | 0 | 0 | 2921 | 2476 | 0 | 6683 | 0 | 0 | 5618 | 0 | 0 | 64942 | | | |
| 615282 | 142407 | 0 | 31480 | 25765 | 16485 | 0 | 0 | 0 | 6050 | 9313 | 0 | 32213 | 3452 | 0 | 12801 | 0 | 0 | | | | |
| 561830 | 0 | 0 | 0 | 0 | 0 | 0 | 4769 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 547183 | 117820 | 95381 | 41680 | 27177 | 121192 | 46365 | 66476 | 31795 | 4957 | 26411 | 37074 | 88948 | 45208 | | | | | | | | |
| 523246 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 522549 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1104 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1534 | 0 | 0 | 5793 | 0 | 0 |
| 516290 | 13387 | 13375 | 18613 | 10349 | 46591 | 1066 | 0 | 0 | 5786 | 12396 | 7155 | 33481 | 15009 | 1695 | | | | | | | |
| 515655 | 0 | 10932 | 0 | 0 | 0 | 19082 | 16381 | 7557 | 0 | 0 | 5806 | 11799 | 2572 | 573 | 0 | 80771 | 43203 | | | | |
| 514926 | 298747 | 0 | 52288 | 50774 | 0 | 0 | 0 | 0 | 13045 | 0 | 0 | 27646 | 0 | 0 | 26169 | 0 | 0 | 155995 | 1454 | | |
| 513826 | 0 | 708 | 0 | 0 | 20938 | 0 | 0 | 0 | 0 | 1350 | 985 | 0 | 6045 | 97 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 366392 | 360963 | 58108 | 94407 | 72824 | 26563 | 0 | 0 | 0 | 19782 | 15554 | 0 | 125110 | 6428 | 1421 | 35653 | | | | | | |
| 348942 | 0 | 37625 | 7510 | 0 | 43745 | 0 | 0 | 0 | 1595 | 10138 | 5132 | 24668 | 13209 | 1310 | 0 | 76604 | | | | | |
| 346735 | 0 | 0 | 0 | 0 | 3926 | 0 | 0 | 0 | 0 | 1173 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 316748 | 0 | 0 | 0 | 0 | 0 | 6607 | 18347 | 8441 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2721 | 0 | 0 | 0 | 0 | 0 | 0 |
| 310817 | 109095 | 0 | 1502 | 13579 | 0 | 0 | 0 | 0 | 1958 | 0 | 0 | 0 | 0 | 0 | 6548 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 303354 | 26897 | 0 | 8936 | 8420 | 0 | 0 | 0 | 0 | 2366 | 0 | 0 | 0 | 0 | 0 | 4179 | 0 | 0 | 0 | 0 | 1909 | | | |
| 300695 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 843 | 1090 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

# MetaPhlAn2 vs Quikr

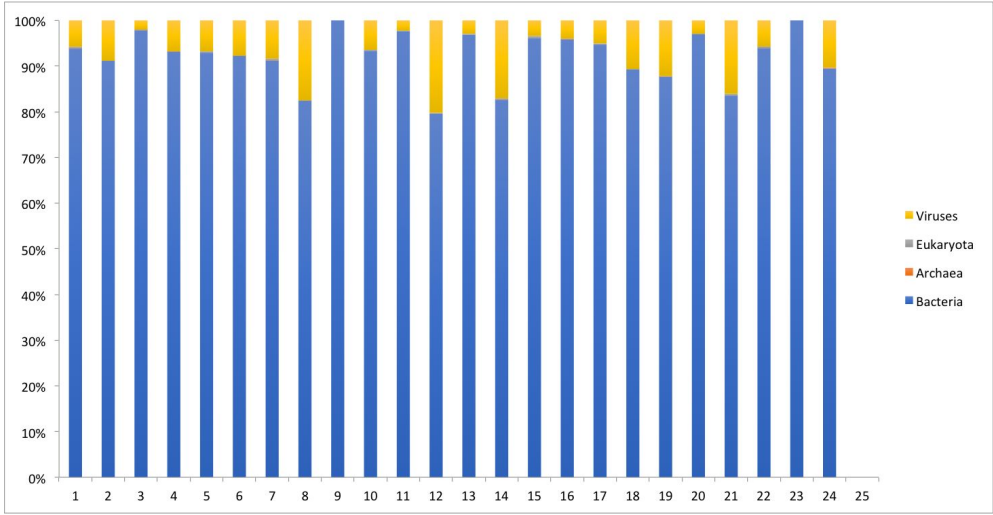
- Classifications
  - MetaPhlAn2 - bacteria, archaea, viruses, and eukaryotes
  - Quikr - bacteria and archaea
- Run Time for a Large Dataset
  - MetaPhlAn2 - ~4 hours
  - Quikr - ~5 minutes
- Urban Dataset Results
  - MetaPhlAn2 - 1498 results ranging from comparison of kingdoms to species
  - Quikr - 198 results, not mores specific than genus
    - Does not give abundances of kingdoms
- Abundances
  - MetaPhlAn2 - relative abundance
  - Quikr - species abundance

# Comparison of Taxonomic Classifications

Quikr




MetaPhlAn2



# Conclusion

- Quikr and MetaPhlAn2 have their advantages and disadvantages
  - Use Quikr for fast processing
  - MetaPhlAn2 gives a comprehensive output that can easily be further analyzed
- Consider refining arguments for input based on samples

# References


Bertelli, C., and G. Greub. "Rapid Bacterial Genome Sequencing: Methods and Applications in Clinical Microbiology." *Clinical Microbiology and Infection* 19.9 (2013): 803-13.

Farmer, Kay H. "Population Size, Density, and Abundance." *A.P.E.S.* N.p., 2011.

"MetaPhlAn2 Tutorial." *BitBucket*. 31 Oct. 2016. <https://bitbucket.org/biobakery/biobakery/wiki/metaphlan2#rst-header-create-taxonomic-profiles>.

“MetaPhlAn V2.0." *MetaPhlAn V2.0 | The Huttenhower Lab*.

Segata, Nicola, Levi Waldron, Annalisa Ballarini, Vagheesh Narasimhan, Olivier Jousson, and Curtis Huttenhower. "Metagenomic Microbial Community Profiling Using Unique Clade-specific Marker Genes." *Nature Methods* 9.8 (2012): 811-14.

Senavirathne, Gayan, Jiaquan Liu, Miguel A. Lopez, Jr., Jeungphill Hanne, Juana Martin-Lopez, Jong-Bong Lee, Kristine E. Yoder, and Richard Fishel. "MetaPhlAn2 for Enhanced Metagenomic Taxonomic Profiling." *Nature Methods* 12.10 (2015): 902-03.

"What Is Shotgun Sequencing?" *Your Genome*. 17 Nov. 2014.

Koslicki, David, Foucart, Simon, Rosen, Gail. “Quikr: a method for rapid reconstruction of bacterial communities via compressive sensing.” *Advance Access* Vol. 28 no. 17 (2013): 2096-2102.