

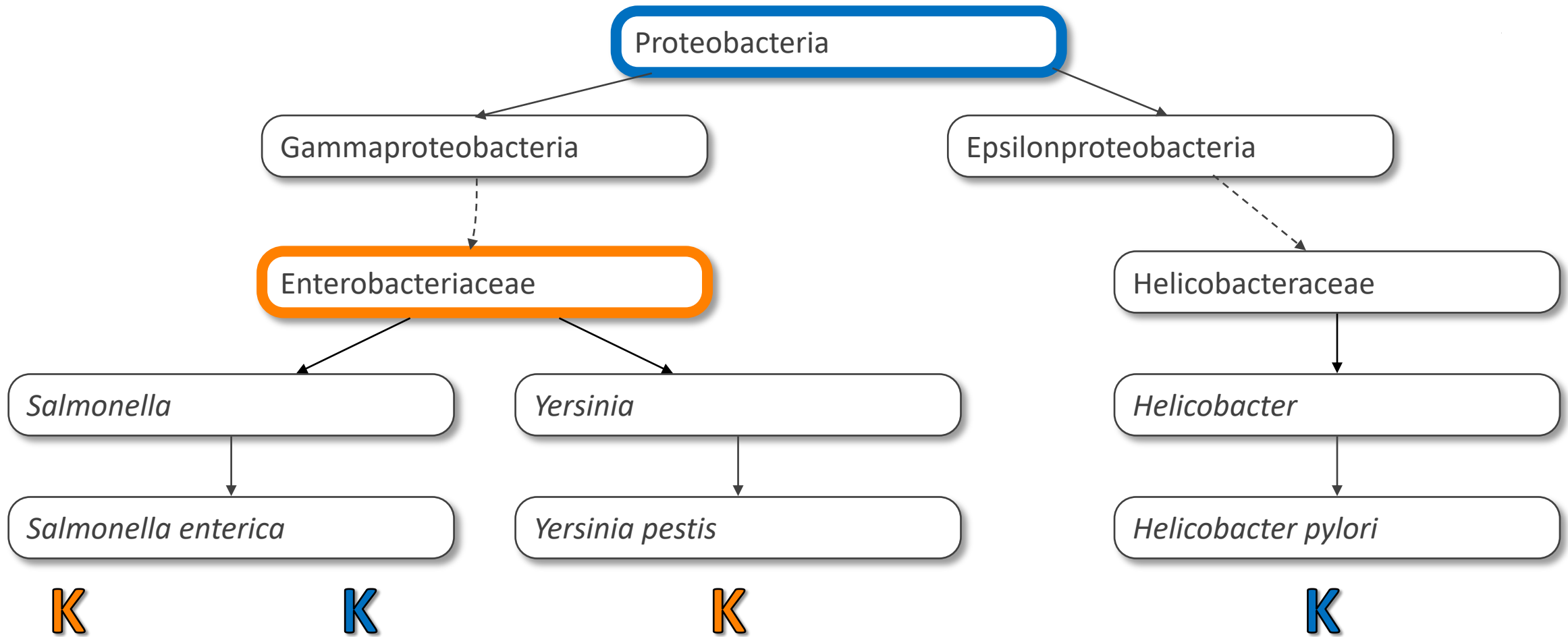
Fast Metagenomic Profiling Methods

JENNIFER LU

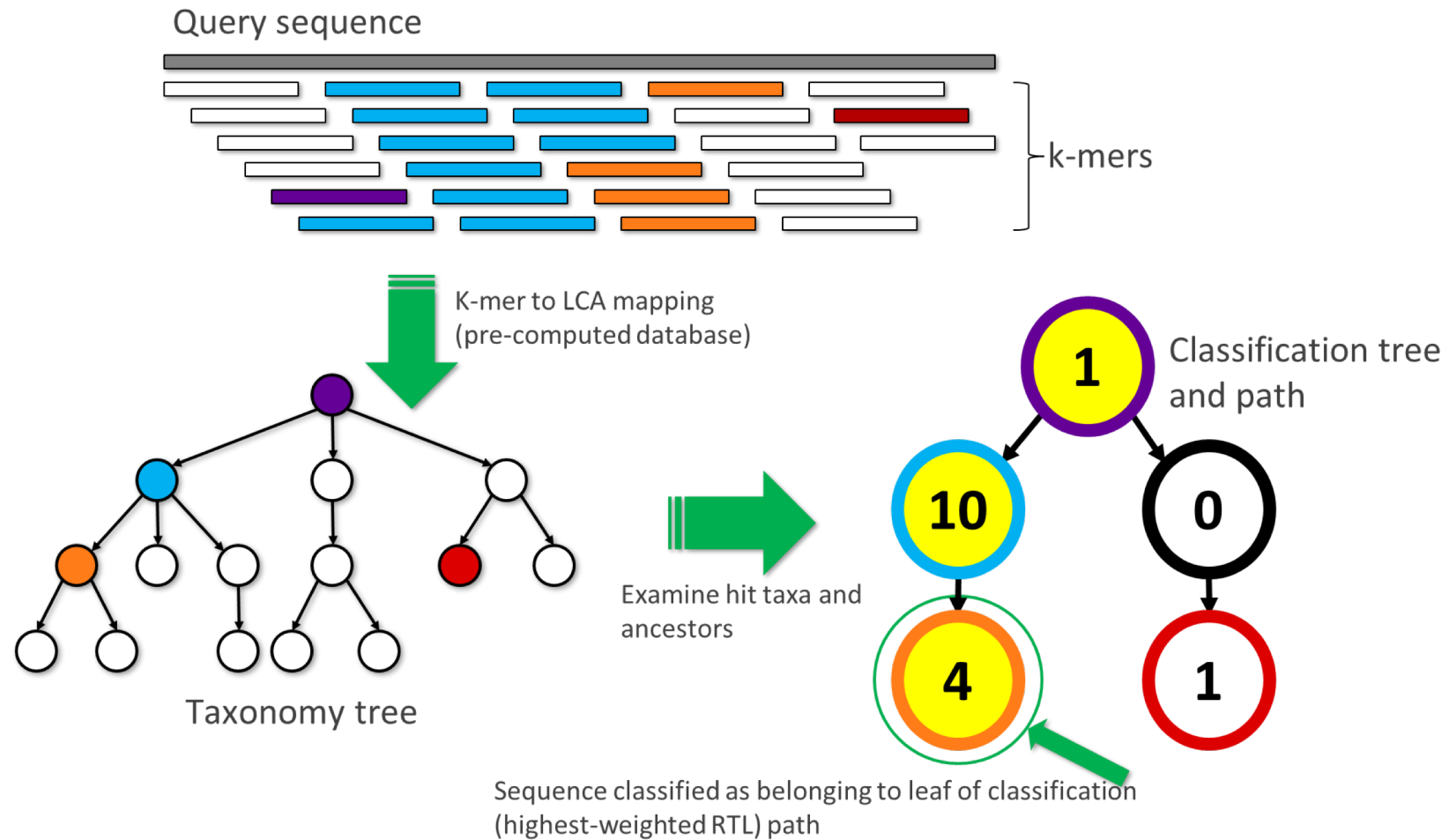
PHD CANDIDATE, BIOMEDICAL ENGINEERING

SALZBERG LAB, JOHNS HOPKINS UNIVERSITY

How does Kraken work?



How does Kraken work?

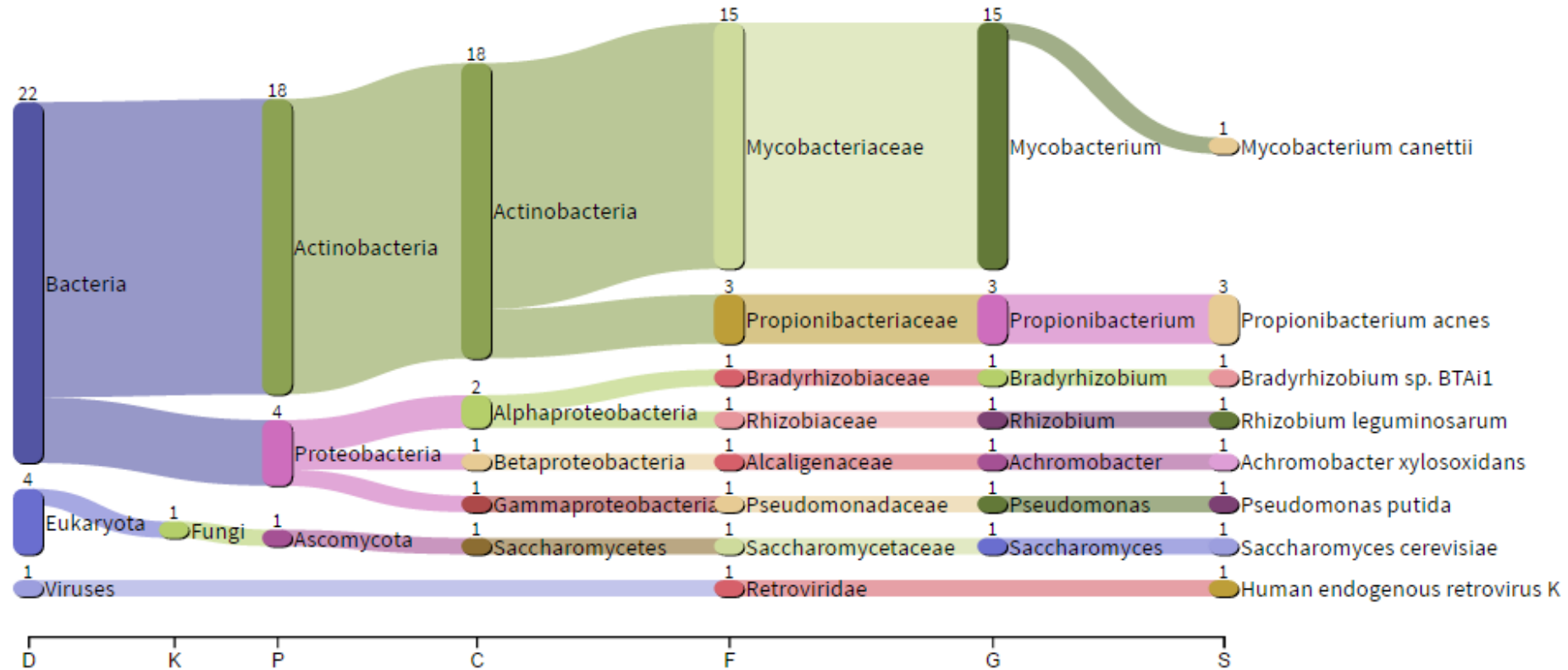


D. Wood and S.L. Salzberg, *Genome Biology* 2014

Kraken Output: Text Report

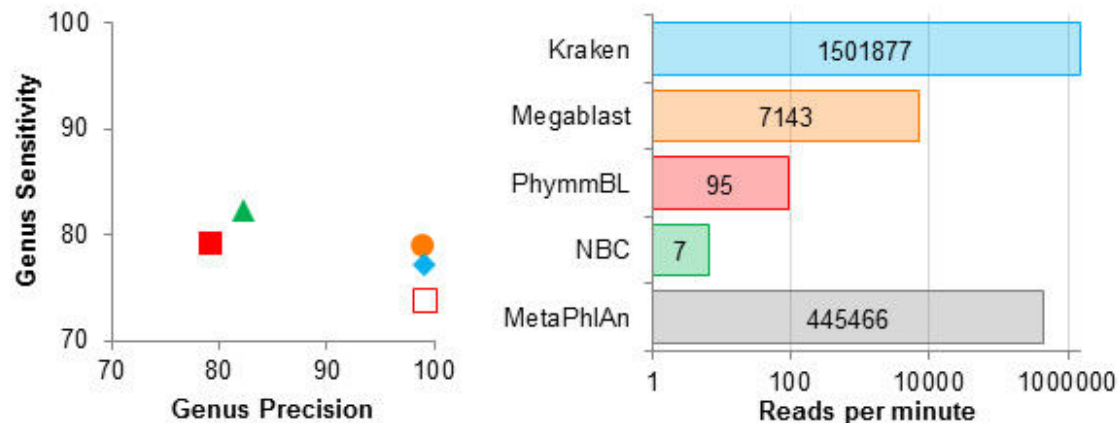
27.45	7319764	1567	P	1239	Firmicutes
20.93	5581780	2613	C	91061	Bacilli
16.04	4276646	1127	O	1385	Bacillales
10.82	2884169	830	F	186817	Bacillaceae
10.81	2883248	1623	G	1386	Bacillus
7.74	2063734	635215	-	86661	Bacillus cereus group
4.08	1086754	28739	S	1396	Bacillus cereus
0.7	187882	20115	S	1428	Bacillus thuringiensis
0.57	151306	151097	S	1392	Bacillus anthracis
0.01	1934	0	S	86662	Bacillus weihenstephanensis
0	643	0	S	580165	Bacillus cytotoxicus
1.06	281618	0	S	79880	Bacillus clausii
1.01	269485	0	S	86665	Bacillus halodurans
1	266768	1270	-	653685	Bacillus subtilis group
1	265462	0	S	1423	Bacillus subtilis
0	23	0	S	492670	Bacillus methylotrophicus
0	13	0	S	1402	Bacillus licheniformis
0	20	0	S	1408	Bacillus pumilus

Kraken Output: Visualized

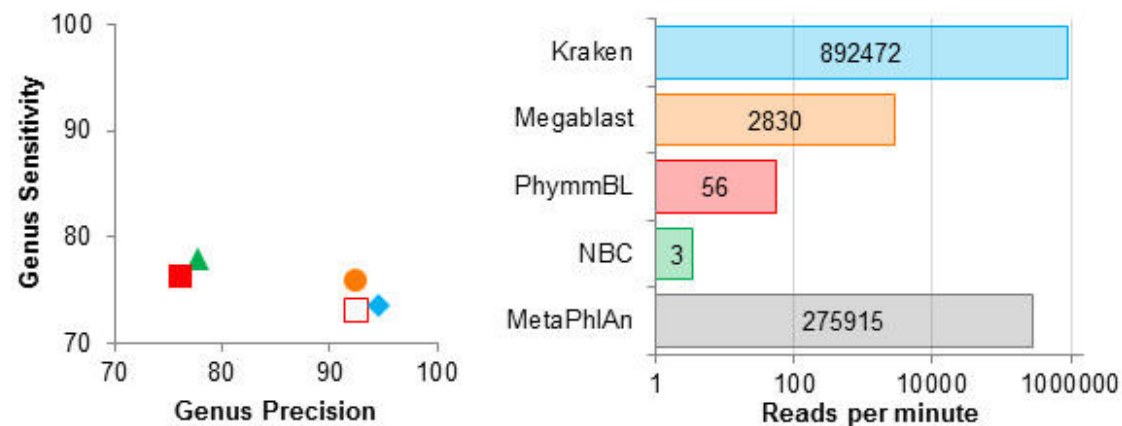


classifiers

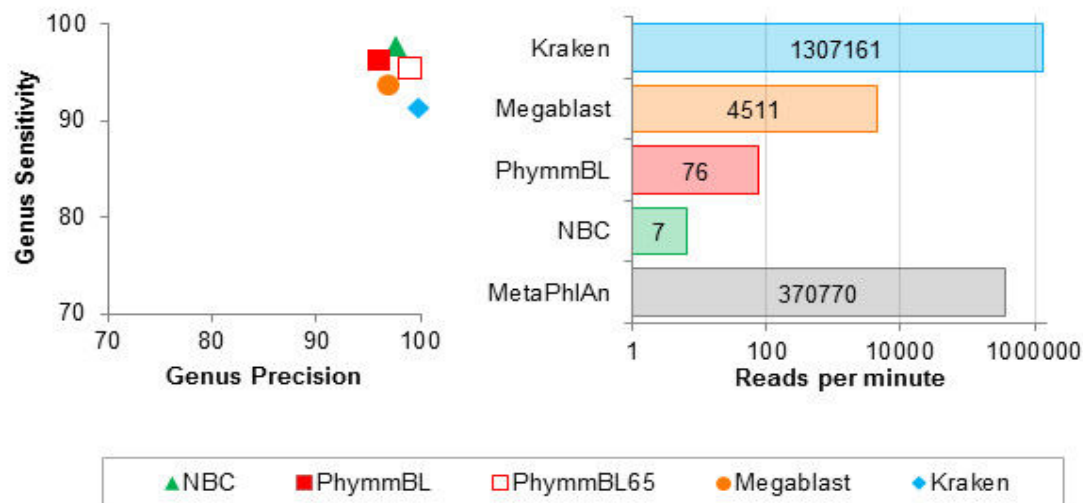
Tests on Hi-Seq Generated Data



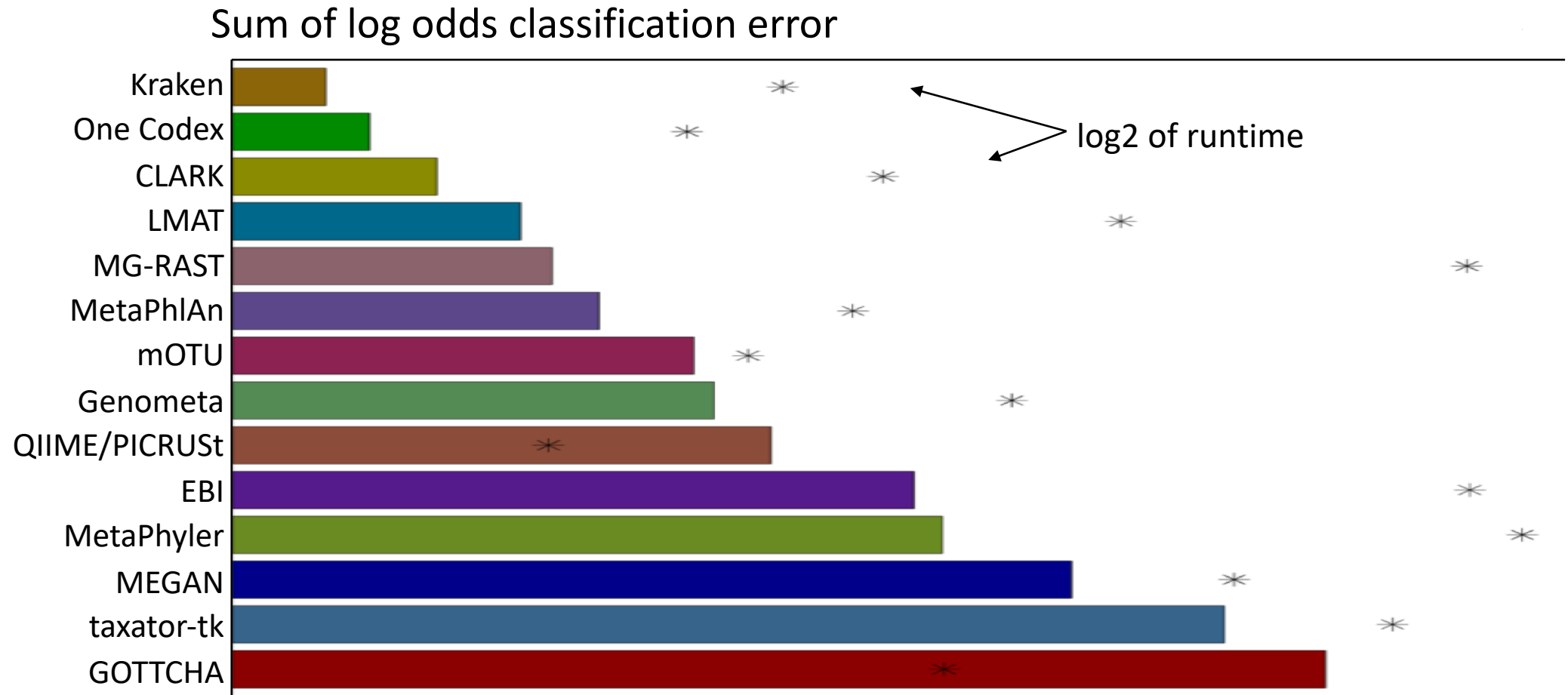
Tests on Mi-Seq Generated Data

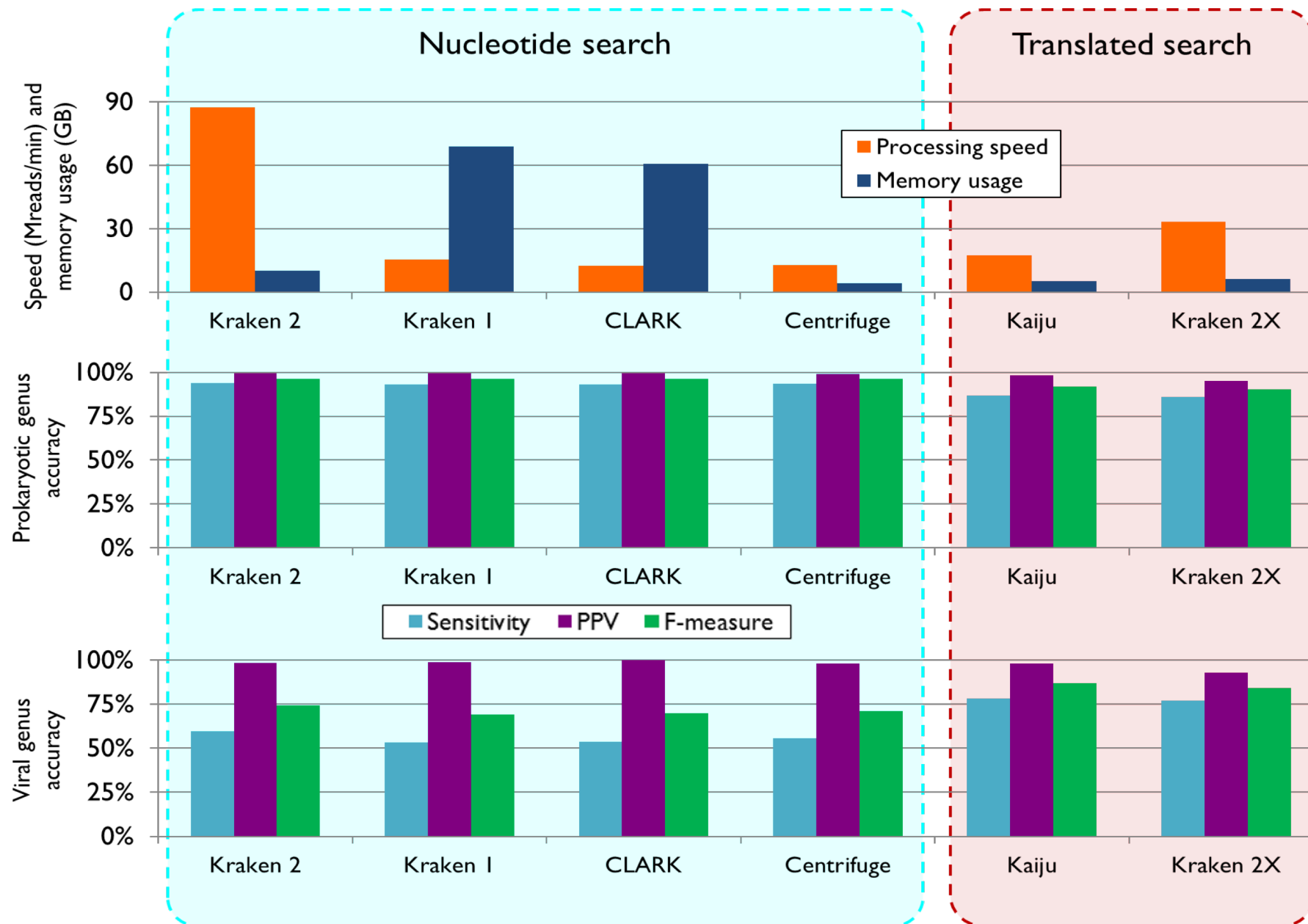


Tests on Simulated Metagenomic Data with high error rates



The first “unbiased, comprehensive benchmark of metagenome analysis tools in which the authors are not involved in any of the tools tested.”





Command Lines

Building:

```
> kraken2-build --download-library bacteria --db ${DATABASE} [--use-ftp, default: rsync]  
> kraken2-build --build --db ${DATABASE} --threads 4
```

Running:

```
> kraken2 --db ${DATABASE} --report mysample.report --threads 4 mysample.fasta >  
    mysample.kraken
```

Abundance Estimation: Bracken

```
> python est_abundance.py -i mysample.report -o mysample.bracken -k  
    database150mers.kmer_distrib
```

Visualization:

Pavian

MiniKraken Runtimes

```
> kraken --db ../m3taxworkshop/databases/kraken/minikraken1_8GB/ --threads 4 ../m3taxworkshop/data/1-datasets/hmp/stool_sample_subset_rep_set_filtered_final.fna > hmp.kraken
```

3032 sequences (1.54 Mbp) processed in 0.632s (287.6 Kseq/m, 145.69 Mbp/m).

3030 sequences classified (99.93%)

2 sequences unclassified (0.07%)

```
> kraken-report -db ../m3taxworkshop/databases/kraken/minikraken1_8GB hmp.kraken > hmp.report
```

```
> kraken2 --db ../m3taxworkshop/databases/kraken/minikraken2_8GB/ --threads 4 ../m3taxworkshop/data/1-datasets/hmp/stool_sample_subset_rep_set_filtered_final.fna > hmp.kraken2
```

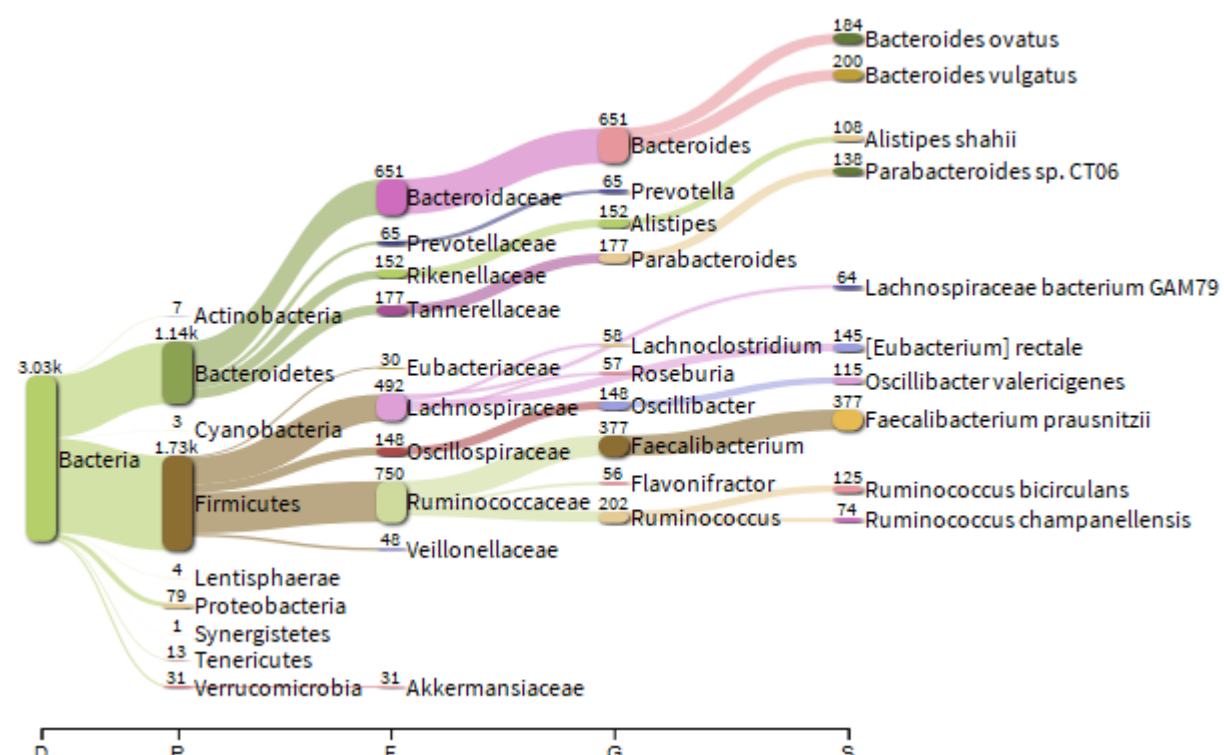
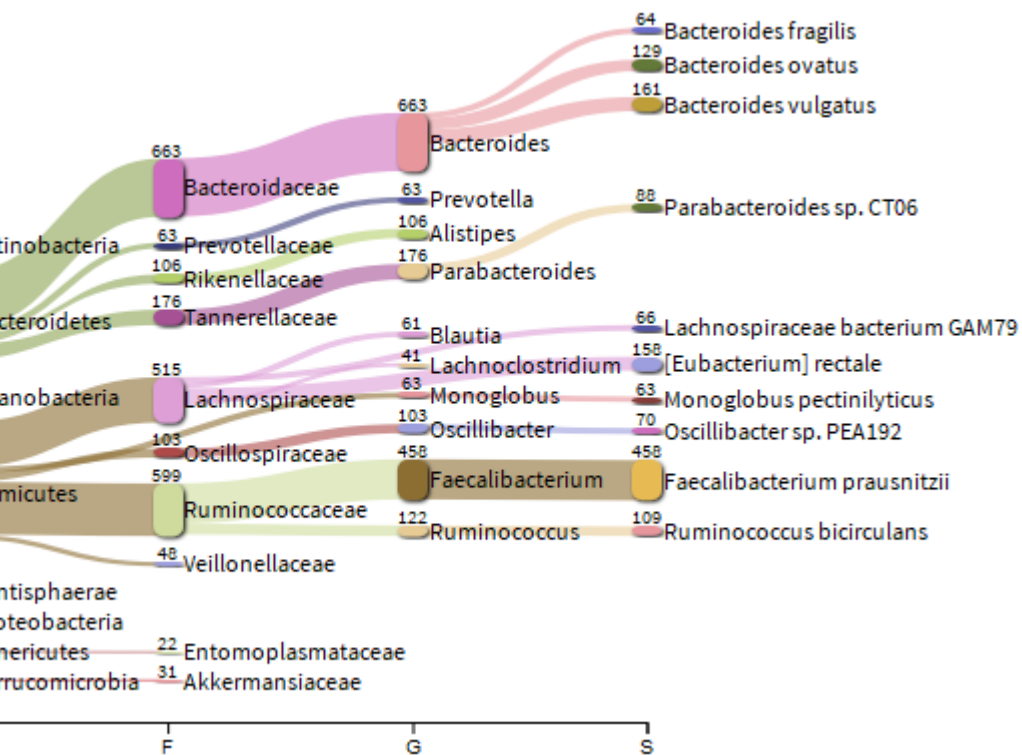
Loading database information... done.

3032 sequences (1.54 Mbp) processed in 0.147s (1239.1 Kseq/m, 627.59 Mbp/m).

3032 sequences classified (100.00%)

0 sequences unclassified (0.00%)

HMP MiniKraken1 v MiniKraken2



Kraken 2 Databases Supported

Download of Standard Genome Sets Supported:

- Refseq archaea, bacteria, plasmid, viral, fungi, plant, protozoa
- NCBI nr, nt, env_nr, env_nt, UniVec, UniVec_Core
- GRCh38 human

Special Databases: Kraken 2 also provides support for

- 16S databases: SILVA, Greengenes, and RDP
- Protein databases [similar to BLASTX].

Kraken 2 uses a reduced amino acid alphabet for protein databases to represent each amino acid with 4 bits. For classification, Kraken classifies all six frames of translation of the query sequence.

Kraken Resources

Kraken: <https://ccb.jhu.edu/software/kraken/>

Kraken Paper: <https://doi.org/10.1186/gb-2014-15-3-r46>

Pavian: <https://github.com/fbreitwieser/pavian>

Bracken: <https://ccb.jhu.edu/software/bracken/>

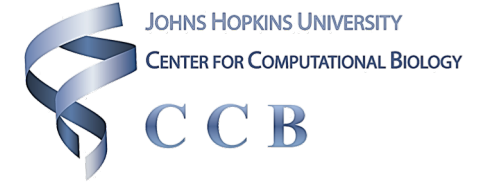
Bracken Paper: <https://peerj.com/articles/cs-104>

Kraken-HLL: <https://github.com/fbreitwieser/krakenhll>

Kraken2: <https://ccb.jhu.edu/software/kraken2/>

For more information: <https://ccb.jhu.edu/software/choosing-a-metagenomics-classifier/>

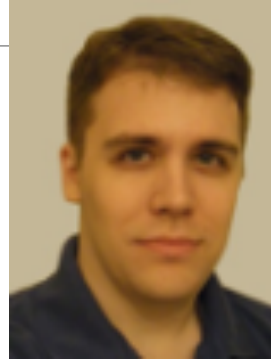
Kraken Team



Steven Salzberg, Ph.D.
Bloomberg Distinguished
Professor of Biomedical
Engineering, Computer
Science, and Biostatistics

Principal Investigator (PI)

<https://salzberg-lab.org>



Derrick Wood, Ph.D.
Principal Bioinformatics
Scientist at Personal
Genome Diagnostics

Developer of Kraken
v1/v2



Florian Breitwieser, Ph.D.
Postdoctoral Fellow in the
Center for Computational
Biology at JHUSOM

Developer of Centrifuge,
Pavian, and Kraken-HLL

[http://ccb.jhu.edu/people/
florian/](http://ccb.jhu.edu/people/florian/)



Jennifer Lu
Ph.D. Candidate in
Biomedical Engineering
at JHU/JHUSOM

Developer of Bracken,
Kraken v1/2 Point of
Contact

[http://ccb.jhu.edu/people/
jennifer.lu/](http://ccb.jhu.edu/people/jennifer.lu/)