

1 FASTQSim scripts used to generate the datasets for this study

To generate the Virus Even dataset:

```
sh FASTQspike.sh --nobackground --platform illumina 10 NC_008297.1.fasta True 10 NC_010563.1.fasta False 10 NC_010562.1.fasta
False 10 NC_00124.1.fasta True 10 NC_005082.1.fasta False 10 NC_006577.2.fasta False 10 NC_001802.1.fasta False 10 NC_001722
.1.fasta False 10 NC_006577.2.fasta False 10 NC_001586.1.fasta True 10 NC_001531.1.fasta True 10 NC_002023.1.fasta False 10
NC_002021.1.fasta False 10 NC_002022.1.fasta False 10 NC_002017.1.fasta False 10 NC_002019.1.fasta False 10 NC_002018.1.
fasta False 10 NC_002016.1.fasta False 10 NC_002020.1.fasta False 10 NC_002204.1.fasta False 10 NC_002205.1.fasta False 10
NC_002206.1.fasta False 10 NC_002207.1.fasta False 10 NC_002208.1.fasta False 10 NC_002209.1.fasta False 10 NC_002210.1.
fasta False 10 NC_002211.1.fasta False 10 NC_006307.1.fasta False 10 NC_006308.1.fasta False 10 NC_006309.1.fasta False 10
NC_006310.1.fasta False 10 NC_006311.1.fasta False 10 NC_006312.1.fasta False 10 NC_006306.1.fasta False 10 NC_005081.1.
fasta False 10 NC_005080.1.fasta False 10 NC_001608.3.fasta False 10 NC_003690.1.fasta False 10 NC_005079.1.fasta False 10
NC_005078.1.fasta False 10 NC_005062.1.fasta False 10 NC_006313.1.fasta False 10 NC_006317.1.fasta False 10 NC_004718.3.fasta
False 10 NC_001549.1.fasta False 10 NC_0064321.1.fasta False --plohistogram -o virus_fastqsim_illumina_even --threads 40
```

To generate the Virus Staggered dataset:

```
sh FASTQspike.sh --nobackground --platform illumina 1 NC_008297.1.fasta True 10 NC_010563.1.fasta False 10 NC_010562.1.fasta
False 0.925 NC_00124.1.fasta True 5 NC_005082.1.fasta False 0.5 NC_006577.2.fasta False 10 NC_001802.1.fasta False 5
NC_001722.1.fasta False 0.5 NC_006577.2.fasta False 20 NC_001586.1.fasta True 5 NC_001531.1.fasta True 5 NC_002023.1.fasta
False 5 NC_002021.1.fasta False 5 NC_002022.1.fasta False 5 NC_002017.1.fasta False 5 NC_002019.1.fasta False 5 NC_002018.1.
fasta False 5 NC_002016.1.fasta False 5 NC_002020.1.fasta False 3 NC_002204.1.fasta False 3 NC_002205.1.fasta False 3
NC_002206.1.fasta False 3 NC_002207.1.fasta False 3 NC_002208.1.fasta False 3 NC_002209.1.fasta False 3 NC_002210.1.fasta
False 3 NC_002211.1.fasta False 1 NC_006307.1.fasta False 1 NC_006308.1.fasta False 1 NC_006309.1.fasta False 1 NC_006310.1.
fasta False 1 NC_006311.1.fasta False 1 NC_006312.1.fasta False 1 NC_006306.1.fasta False 3 NC_005081.1.fasta False 3
NC_005080.1.fasta False 1 NC_001608.3.fasta False 1 NC_003690.1.fasta False 2 NC_005079.1.fasta False 2 NC_005078.1.fasta
False 5 NC_005062.1.fasta False 1 NC_006313.1.fasta False 1 NC_006317.1.fasta False 1 NC_004718.3.fasta False 1 NC_001549.1.
fasta False 100 NC_0064321.1.fasta False
--plohistogram -o virus_fastqsim_illumina_even --threads 40
```

To generate HMP Even dataset:

```
sh FASTQspike.sh --nobackground --platform illumina 18.85 NZ_DS264586.1.fasta True 3.14 NC_003909.8.fasta True 1.435 NC_009614.1.
fasta True 1.452 NC_009617.1.fasta True 1.249 NC_001263.1.fasta True 2.45 NC_001264.1.fasta True 2.37 NC_017316.1.fasta True
1.615 NC_000913.3.fasta True 4.496 NC_000915.1.fasta True 3.95 NC_008530.1.fasta True 2.547 NC_003210.1.fasta True 4.047
NC_009515.1.fasta True 3.3 NC_003112.2.fasta True 2.929 AE017283.1.fasta True 1.197 NC_002516.2.fasta True 1.629 NC_007494
.2.fasta True 1.629 NC_009007.1 True 1.629 NC_007488.2.fasta True 1.629 NC_007489.1.fasta True 1.629 NC_007490.2.fasta True
1.629 NC_009008.1.fasta True 1.629 NC_007493.2.fasta True 2.61 NC_0100079.1.fasta True 3.0008 NC_004461.1.fasta True 3.471
NC_004116.1.fasta True 3.689 NC_004350.2.fasta True 3.47 NC_003028.3.fasta True --plohistogram -o hmp_fastqsim_illumina_even
--threads 40
```

To generate HMP Staggered dataset:

```
sh FASTQspike.sh --nobackground --platform illumina 11.291 NZ_DS264586.1.fasta True 5.396 NC_003909.8.fasta True 4.593 NC_009614.1.
fasta True 4.41 NC_009617.1.fasta 4.41 NC_001263.1.fasta 4.38 NC_001264.1.fasta True 4.067 NC_017316.1.fasta True 3.959
NC_000913.3.fasta True 2.49 NC_000915.1.fasta True 2.428 NC_008530.1.fasta True 2.26 NC_003210.1.fasta True 2.038 NC_009515
.1.fasta True 0.586 NC_003112.2.fasta True 0.12 AE017283.1.fasta True 0.037 NC_002516.2.fasta True 0.035 NC_007494.2.fasta
True 0.033 NC_009007.1.fasta True 0.033 NC_007488.2.fasta True 0.033 NC_007489.1.fasta True 0.033 NC_007490.2.fasta True
0.033 NC_009008.1.fasta True 0.033 NC_007493.2.fasta True 0.033 NC_0100079.1.fasta True 0.026 NC_004461.1.fasta True 0.006
NC_004461.1.fasta True 0.006 NC_004116.1.fasta True 0.001 NC_004350.2.fasta True 0.001 NC_003028.3.fasta True --plohistogram
-o hmp_fastqsim_illumina_staggered --threads 40
```

To generate Bacteria dataset:

```
sh FASTQspike.sh --nobackground --platform illumina 3 NC_016795.1.fasta 3 NC_016777.1.fasta True 3 NC_006932.1.fasta True 3
NC_006933.1.fasta True 5 CP001488.1.fasta True 5 CP001489.1.fasta True 5 NC_007618.1.fasta True 5 NC_007624.1.fasta True 5
NC_003317.1.fasta True 5 NC_003318.1.fasta True 5 NC_017244.1.fasta True 5 NC_017245.1.fasta True 0.00003 NC_004310.3.fasta
True 0.00003 NC_004311.2.fasta True 0.00003 NC_010169.1.fasta True 0.00003 NC_010167.1.fasta True 0.00003 CP006961.1.fasta
True 0.00003 CP006962.1.fasta True 5 NC_011528.1.fasta True 1 NC_010117.1.fasta True 10 NC_002971.3.fasta True 10 CP007025
.1.fasta True 5 NC_011750.1.fasta True 3 NC_018658.1.fasta True 0.00002 NC_011740.1.fasta True 0.00002 NZ_CM001142.1.fasta
True 1 AJ536088.1.fasta True 1 JX983649.1.fasta True 3 NC_021870.1.fasta True 3 NC_003198.1.fasta True 150 CP007448.1.fasta
True 200 NC_008800.1.fasta True 100 NC_015224.1.fasta True 5 NC_017168.1.fasta True 3 NC_010159.1.fasta True 50 NC_003143.1.
fasta True 25 NC_004088.1.fasta True 10 NC_009381.1.fasta True 2 NC_009708.1.fasta True 1 NC_006155.1.fasta True 0.5
NC_010634.1.fasta True 0.00002 NC_010465.1.fasta True --plohistogram -o bacteria_fastqsim_illumina --threads 40
```

To generate Human dataset:

```
sh FASTQspike.sh —nobackground —platform illumina 10 Grch38.fasta False —plothistogram —o human —threads 40
```

2 Commands used to run the six evaluation tools, using the "Bacteria" dataset as an example input

Evaluation algorithms were executed with the following commands, using the "Bacteria" dataset as an example. The same commands were used to characterize the five other datasets, replacing Bacteria.fastq with the respective input fastq files.

```
MetaScope:
fastq=Bacteria.fastq
outputf=Bacteria.out
sequencer=illumina
host="Eukaryota;Metazoa;Chordata;Craniata;Vertebrata;Euteleostomi;Mammalia;Eutheria;Euarchontoglires;
Primates;Haplorrhini;Catarrhini;Hominidae;Homo."
cores=60
metascope $fastq $outputf.dir $outputf $sequencer $host $cores
```

```
Kraken
Note: MinkKraken DB downloaded from https://ccb.jhu.edu/software/kraken/
kraken --db kraken_standard_db/ --threads 60 --fastq-input Bacteria.fastq > Bacteria.out
kraken-report --db kraken_standard_db/ Bacteria.out >Bacteria.out.report
```

```
MetaPhyler:
Note: Accepts FASTA format (not FASTQ). The FASTQ input files were converted to FASTA
prior to analysis with MetaPhyler, this step was not included in the runtime computation.
```

```
runMetaphyler.pl Bacteria.fa blastx Bacteria.out 60
```

```
MetaCV:
Note: cvdb_2059 database used from http://sourceforge.net/projects/metacv/files/cvdb_2059/
metacv classify /data/metacv_db/cvk6_2059 Bacteria.fastq Bacteria.out
--threads=60
metacv res2table /data/metacv_db/cvk6_2059 Bacteria.out Bacteria.out.2 --threads=60
```

```
MetaPhlAn:
Note: mpa database downloaded from https://bitbucket.org/nsegata/metaphlan/src/f353151d84e3/blastdb/
./metaphlan.py --bowtie2db bowtie2db/mpa --nproc 60 Bacteria.fq --bowtie2out metaphlan_bacteria.bt2out.txt
./metaphlan.py --input_type blastout metaphlan_bacteria.bt2out.txt > Bacteria.out
```

```
LMAT:
Note: the LMAT prebuilt databases were used.
./run_cs.sh --overwrite --ilst=Bacteria.out.flst --odir=Bacteria.out
--filesun=Bacteria.fa.kML+Human.v4-14.20.g10.db.lo.rl_output.0.30.fastsummary

./run_gl.sh --db_file = gene.20mer.db --ilst=Bacteria.out/rl_output.flst --odir=Bacteria.out
--filesun=Bacteria.fa.kML+Human.v4-14.20.g10.db.lo.rl_output.0.30.fastsummary --overwrite --threads=60
```

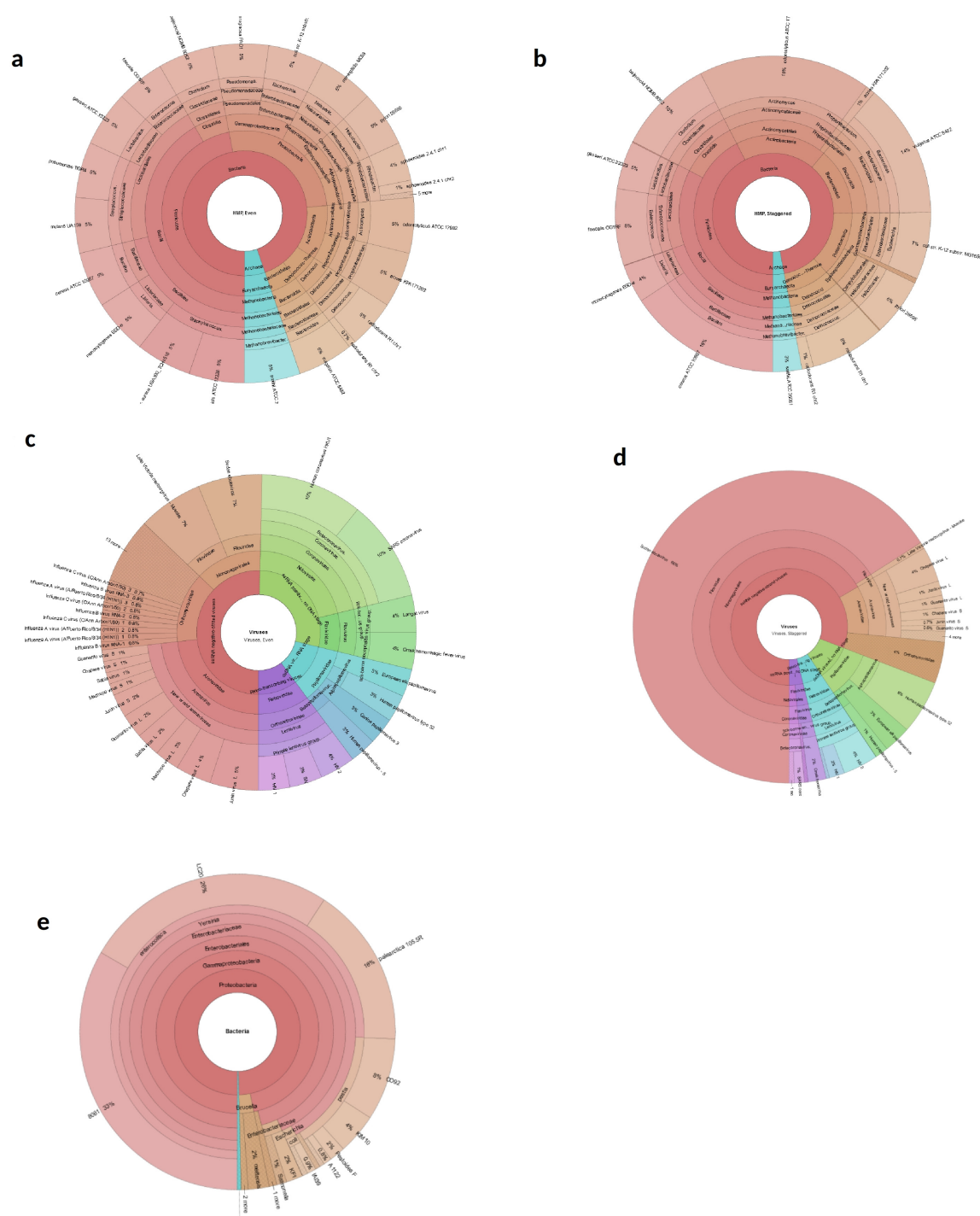


Fig. 1. FASTQSim in silico dataset composition to strain level. a. 20 bacteria from the Human Microbiome Project (HMP), even coverage levels. b. Same 20 bacteria from HMP, staggered coverage levels. c. 22 species of viruses across 11 genera, even coverage levels. d. Same 22 species of viruses, staggered coverage levels. e. 33 strains of bacteria representing 13 species and 5 genera. See Krona HTML files for a-e.

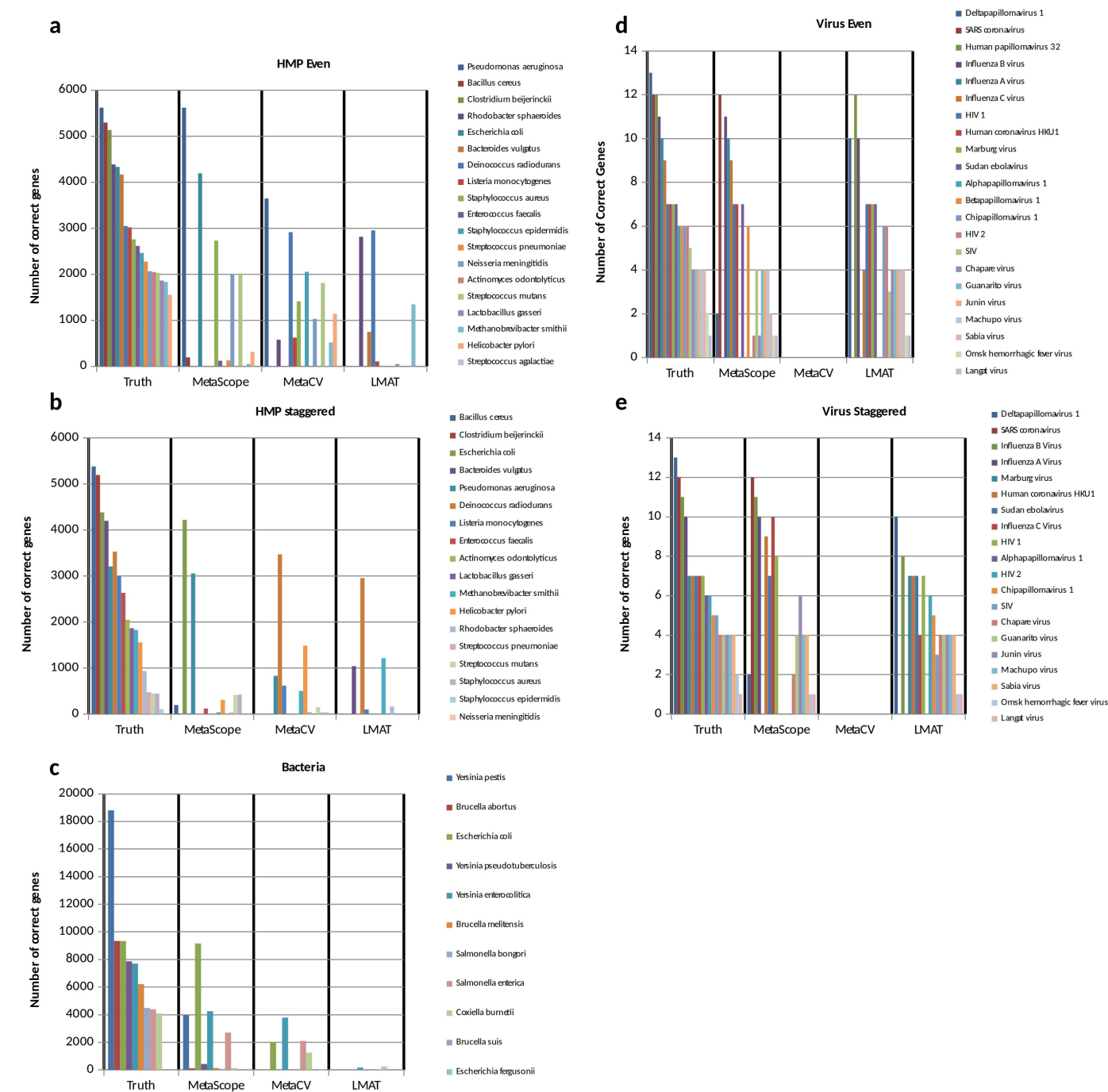


Fig. 2. False positive organisms identified to the genus and species level by the 6 metagenomic algorithms. Heatmap color scales are log10 (number of incorrectly assigned false positive reads) for a genus or species. a. All false positive genera identified in the bacterial dataset. b. All false positive species identified in the bacterial dataset. c. 20 false positive genera for the bacterial dataset with the most assigned reads. d. 20 false positive species for the bacterial dataset with the most assigned reads. e. All false positive genera identified in the virus even dataset. f. All false positive species identified in the virus even dataset. g. 20 false positive genera for the virus even dataset with the most assigned reads. h. 20 false positive species for the virus even dataset with the most assigned reads.

Table 1. Source organisms and coverage levels for HMP Even and HMP Staggered datasets.

Taxonomy	Organism	NCBI accession	Platform	HMP Even Coverage	HMP Even % Enrichment	HMP Even Reads	HMP Staggered Coverage	HMP Staggered Reads
Bacteria;Actinobacteria;Actinomycetales;Actinomycetaceae;Actinomycetes	Actinomycetes odontolyticus ATCC 17982 ScfId021 genomic scaffold, whole genome shotgun sequence	NZ_DS264586.1	Illumina	18.85	0.050402832580679	60471	11.291	217512
Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;	Bacillus cereus ATCC 10987, complete genome	NC_003909.8	Illumina	3.14	0.0503253166899214	60378	5.396	193298
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides	Bacteroides vulgatus ATCC 8482 chromosome, complete genome	NC_009614.1	Illumina	1.435	0.0503253166899214	60378	4.593	169164
Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium	Clostridium beijerinckii NCIMB 8052 chromosome, complete genome	NC_009617.1	Illumina	1.452	0.0503111471184926	60361	4.41	120792
Bacteria;Deinococcus-Thermus;Deinococci;Deinococcales;Deinococcaceae;Deinococcus.	Deinococcus radiodurans R1 chromosome 1, complete sequence	NC_001263.1	Illumina	1.249	0.043558929580564	52260	4.41	94067
Bacteria;Deinococcus-Thermus;Deinococci;Deinococcales;Deinococcaceae;Deinococcus.	Deinococcus radiodurans R1 chromosome 2, complete sequence	NC_001264.1	Illumina	2.45	0.00679555975641673	8153	4.38	14624
Bacteria;Firmicutes;Bacilli;Lactobacillales;Enterococcaceae;Enterococcus	Enterococcus faecalis OG1RF chromosome, complete genome	NC_017316.1	Illumina	2.37	0.0435839347066149	52290	4.067	96640
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;Escherichia.	Escherichia coli str. K-12 substr. MG1655, complete genome	NC_000913.3	Illumina	1.615	0.050321149168913	60373	3.959	84474
Bacteria;Proteobacteria;Epsilonproteobacteria;Campylobacteriales;Helicobacteriaceae;Helicobacter.	Helicobacter pylori 26695 chromosome, complete genome	NC_000915.1	Illumina	4.496	0.0503378192529469	60393	2.49	72463
Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Lactobacillus.	Lactobacillus gasseri ATCC 33323 chromosome, complete genome	NC_008530.1	Illumina	3.95	0.0502194616563062	60251	2.428	60409
Bacteria;Firmicutes;Bacilli;Bacillales;Listeriaceae;Listeria.	Listeria monocytogenes EGD-e chromosome, complete genome	NC_003210.1	Illumina	2.547	0.0503553228411824	60414	2.26	48371
Archaea;Euryarchaeota;Methanobacteria;Methanobacteriales;Methanobacteriaceae;Methanobrevibacter.	Methanobrevibacter smithii ATCC 35061 chromosome, complete genome	NC_009515.1	Illumina	4.047	0.0503553228411824	60414	2.038	36272
Bacteria;Proteobacteria;Betaproteobacteria;Neisseriales;Neisseriaceae;Neisseria.	Neisseria meningitidis MC58 chromosome, complete genome	NC_003112.2	Illumina	3.3	0.0503494883117706	60407	0.586	2
Bacteria;Actinobacteria;Propionibacteriales;Propionibacteriaceae;Propionibacterium.	Propionibacterium acnes KPA171202, complete genome	AE017283.1	Illumina	2.929	0.0503403197655519	60396	0.12	12095
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas.	Pseudomonas aeruginosa PAO1 chromosome, complete genome	NC_002516.2	Illumina	1.197	0.0503478213033672	60405	0.037	6058
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 chromosome 2, complete sequence	NC_007494.2	Illumina	1.629	0.0103146144959717	12375	0.035	252
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 plasmid A, partial sequence	NC_009007.1	Illumina	1.629	0.00124942279834033	1499	0.033	31
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 plasmid B, complete sequence	NC_007488.2	Illumina	1.629	0.00124525527733185	1494	0.033	32
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 plasmid C, complete sequence	NC_007489.1	Illumina	1.629	0.00114856878993527	1378	0.033	28
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 plasmid D, complete sequence	NC_007490.2	Illumina	1.629	0.00110355956304376	1324	0.033	28
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 plasmid E, partial sequence	NC_009008.1	Illumina	1.629	0.000409250563032088	491	0.033	10
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 chromosome 1, complete sequence	NC_007493.2	Illumina	1.629	0.0348663142610902	41831	0.033	849
Bacteria;Firmicutes;Bacilli;Bacillales;Staphylococcus.	Staphylococcus aureus subsp. aureus USA300_TCH1516 chromosome, complete genome	NC_010079.1	Illumina	2.61	0.0503444872865604	60401	0.033	602
Bacteria;Firmicutes;Bacilli;Bacillales;Staphylococcus.	Staphylococcus epidermidis ATCC 12228 chromosome, complete genome	NC_004461.1	Illumina	3.0008	0.050361990874796	60422	0.026	123
Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Streptococcus.	Streptococcus agalactiae 2603V/R chromosome, complete genome	NC_004116.1	Illumina	3.471	0.0503436537823587	60400	0.006	18
Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Streptococcus.	Streptococcus mutans UA159 chromosome, complete genome	NC_004350.2	Illumina	3.689	0.0503461542949638	60403	0.001	603
Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Streptococcus.	Streptococcus pneumoniae TIGR4 chromosome, complete genome	NC_003028.3	Illumina	3.47	0.0503369857487452	60392	0.001	609

Table 2. Source organisms and coverage levels for the bacteria specificity dataset.

Taxonomy	Organism	NCBI accession	Platform	HMP Even Coverage	HMP Even % Enrichment	HMP Even Reads	HMP Staggered Coverage	HMP Staggered Reads
Bacteria;Actinobacteria;Actinomycetales;Actinomycetaceae;Actinomycetes	Actinomycetes odontolyticus ATCC 17982 ScfId021 genomic scaffold, whole genome shotgun sequence	NZ_DS264586.1	Illumina	18.85	0.050402832580679	60471	11.291	217512
Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;	Bacillus cereus ATCC 10987, complete genome	NC_003909.8	Illumina	3.14	0.0503253166899214	60378	5.396	193298
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides	Bacteroides vulgatus ATCC 8482 chromosome, complete genome	NC_009614.1	Illumina	1.435	0.0503253166899214	60378	4.593	169164
Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiaceae;Clostridium	Clostridium beijerinckii NCIMB 8052 chromosome, complete genome	NC_009617.1	Illumina	1.452	0.0503111471184926	60361	4.41	120792
Bacteria;Deinococcus-Thermus;Deinococci;Deinococcales;Deinococcaceae;Deinococcus.	Deinococcus radiodurans R1 chromosome 1, complete sequence	NC_001263.1	Illumina	1.249	0.043558929580564	52260	4.41	94067
Bacteria;Deinococcus-Thermus;Deinococci;Deinococcales;Deinococcaceae;Deinococcus.	Deinococcus radiodurans R1 chromosome 2, complete sequence	NC_001264.1	Illumina	2.45	0.00679555975641673	8153	4.38	14624
Bacteria;Firmicutes;Bacilli;Lactobacillales;Enterococcaceae;Enterococcus	Enterococcus faecalis OG1RF chromosome, complete genome	NC_017316.1	Illumina	2.37	0.0435839347066149	52290	4.067	96640
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;Escherichia.	Escherichia coli str. K-12 substr. MG1655, complete genome	NC_000913.3	Illumina	1.615	0.050321149168913	60373	3.959	84474
Bacteria;Proteobacteria;Epsilonproteobacteria;Campylobacteriales;Helicobacteriaceae;Helicobacter.	Helicobacter pylori 26695 chromosome, complete genome	NC_000915.1	Illumina	4.496	0.0503378192529469	60393	2.49	72463
Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Lactobacillus.	Lactobacillus gasseri ATCC 33323 chromosome, complete genome	NC_008530.1	Illumina	3.95	0.0502194616563062	60251	2.428	60409
Bacteria;Firmicutes;Bacilli;Bacillales;Listeriaceae;Listeria.	Listeria monocytogenes EGD-e chromosome, complete genome	NC_003210.1	Illumina	2.547	0.0503553228411824	60414	2.26	48371
Archaea;Euryarchaeota;Methanobacteria;Methanobacteriales;Methanobacteriaceae;Methanobrevibacter.	Methanobrevibacter smithii ATCC 35061 chromosome, complete genome	NC_009515.1	Illumina	4.047	0.0503553228411824	60414	2.038	36272
Bacteria;Proteobacteria;Betaproteobacteria;Neisseriales;Neisseriaceae;Neisseria.	Neisseria meningitidis MC58 chromosome, complete genome	NC_003112.2	Illumina	3.3	0.0503494883117706	60407	0.586	2
Bacteria;Actinobacteria;Propionibacteriales;Propionibacteriaceae;Propionibacterium.	Propionibacterium acnes KPA171202, complete genome	AE017283.1	Illumina	2.929	0.0503403197655519	60396	0.12	12095
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas.	Pseudomonas aeruginosa PAO1 chromosome, complete genome	NC_002516.2	Illumina	1.197	0.0503478213033672	60405	0.037	6058
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 chromosome 2, complete sequence	NC_007494.2	Illumina	1.629	0.0103146144959717	12375	0.035	252
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 plasmid A, partial sequence	NC_009007.1	Illumina	1.629	0.00124942279834033	1499	0.033	31
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 plasmid B, complete sequence	NC_007488.2	Illumina	1.629	0.00124525527733185	1494	0.033	32
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 plasmid C, complete sequence	NC_007489.1	Illumina	1.629	0.00114856878993527	1378	0.033	28
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 plasmid D, complete sequence	NC_007490.2	Illumina	1.629	0.00110355956304376	1324	0.033	28
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 plasmid E, partial sequence	NC_009008.1	Illumina	1.629	0.000409250563032088	491	0.033	10
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter.	Rhodobacter sphaeroides 2.4.1 chromosome 1, complete sequence	NC_007493.2	Illumina	1.629	0.0348663142610902	41831	0.033	849
Bacteria;Firmicutes;Bacilli;Bacillales;Staphylococcus.	Staphylococcus aureus subsp. aureus USA300_TCH1516 chromosome, complete genome	NC_010079.1	Illumina	2.61	0.0503444872865604	60401	0.033	602
Bacteria;Firmicutes;Bacilli;Bacillales;Staphylococcus.	Staphylococcus epidermidis ATCC 12228 chromosome, complete genome	NC_004461.1	Illumina	3.0008	0.050361990874796	60422	0.026	123
Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Streptococcus.	Streptococcus agalactiae 2603V/R chromosome, complete genome	NC_004116.1	Illumina	3.471	0.0503436537823587	60400	0.006	18
Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Streptococcus.	Streptococcus mutans UA159 chromosome, complete genome	NC_004350.2	Illumina	3.689	0.0503461542949638	60403	0.001	603
Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Streptococcus.	Streptococcus pneumoniae TIGR4 chromosome, complete genome	NC_003028.3	Illumina	3.47	0.0503369857487452	60392	0.001	609
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Moraxellaceae;Acinetobacter;Acinetobacter calcoaceticus/baumannii complex.	Acinetobacter baumannii ATCC 17978 (g-proteobacteria)	NC_009085.1	Illumina	0.50341820073115	603978	2417132		

Table 3. Source organisms and coverage levels for Virus Even and Virus Staggered datasets.

Taxonomy	Organism	NCBI accession	Platform	Virus Even Coverage	Virus Even Reads	Virus Even % Enrichment	Virus Staggered Coverage	Virus Staggered Reads	Virus Staggered % Enrichment
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza B virus RNA-1	NC_002204.1	Illumina	10x	1914	0.00835884513426995	3x	572	0.00250826587618288
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza B virus RNA-2	NC_002205.1	Illumina	10x	1870	0.00816668777486145	3x	559	0.00245125983354236
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza B virus RNA-3	NC_002206.1	Illumina	10x	1783	0.00778674026875827	3x	532	0.0023328626680582
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza B virus RNA-4	NC_002207.1	Illumina	10x	1523	0.00665126496316256	3x	456	0.00199959657262131
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza B virus RNA-5	NC_002208.1	Illumina	10x	1490	0.00650714694360618	3x	446	0.00195574577059014
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza B virus RNA-6	NC_002209.1	Illumina	10x	1261	0.00550705523213919	3x	378	0.00165756031677819
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza B virus RNA-7	NC_002210.1	Illumina	10x	963	0.00420562584341796	3x	289	0.00126728817870079
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza B virus RNA-8	NC_002211.1	Illumina	10x	886	0.00386935046445307	3x	266	0.0011664313340291
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza A virus (A/Puerto Rico/8/34 (H1N1)) segment 1	NC_002023.1	Illumina	10x	1893	0.00826713366727953	5x	940	0.0041219753909299
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza A virus (A/Puerto Rico/8/34 (H1N1)) segment 2	NC_002021.1	Illumina	10x	1893	0.00826713366727953	5x	940	0.0041219753909299
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza A virus (A/Puerto Rico/8/34 (H1N1)) segment 3	NC_002022.1	Illumina	10x	1806	0.00788718616117635	5x	897	0.00393341694219587
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza A virus (A/Puerto Rico/8/34 (H1N1)) segment 4	NC_002017.1	Illumina	10x	1440	0.0062887863079147	5x	716	0.00313971742543171
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza A virus (A/Puerto Rico/8/34 (H1N1)) segment 5	NC_002019.1	Illumina	10x	1267	0.00553325850842217	5x	630	0.00276260052796366
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza A virus (A/Puerto Rico/8/34 (H1N1)) segment 6	NC_002018.1	Illumina	10x	1143	0.00499172413190729	5x	569	0.00249511063557352
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza A virus (A/Puerto Rico/8/34 (H1N1)) segment 7	NC_002016.1	Illumina	10x	831	0.00362915376519244	5x	415	0.00181980828429352
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza A virus (A/Puerto Rico/8/34 (H1N1)) segment 8	NC_002020.1	Illumina	10x	722	0.00315312757938501	5x	360	0.00157862887312209
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza C virus (C/Ann Arbor/1/50) segment 1, partial sequence	NC_006307.1	Illumina	10x	1880	0.00821035990199975	1x	188	0.00082439507818598
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza C virus (C/Ann Arbor/1/50) segment 2, partial sequence	NC_006308.1	Illumina	10x	1832	0.00800073369173592	1x	184	0.000806854757373512
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza C virus (C/Ann Arbor/1/50) segment 3, partial sequence	NC_006309.1	Illumina	10x	1723	0.00752470750592849	1x	173	0.000758618875139226
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza C virus (C/Ann Arbor/1/50) segment 4, partial sequence	NC_006310.1	Illumina	10x	1592	0.00695260264041681	1x	159	0.000697227752295589
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza C virus (C/Ann Arbor/1/50) segment 5, partial sequence	NC_006311.1	Illumina	10x	1462	0.00638466498761895	1x	146	0.00064022170965507
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza C virus (C/Ann Arbor/1/50) segment 6, partial sequence	NC_006312.1	Illumina	10x	908	0.00396542914415732	1x	91	0.000399042298483639
Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;	Influenza C virus (C/Ann Arbor/1/50) segment 7, partial sequence	NC_006306.1	Illumina	10x	758	0.00331034723708288	1x	76	0.000333266095436886
Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Flavivirus; tick-borne encephalitis virus group; Kyasanur forest disease virus	Alkhurma hemorrhagic fever virus	NC_004355.1	Illumina	10x	8602	0.0375667637643627	1x	2577	0.0113003516834323
Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;Chpapillomavirus	Canine papillomavirus 3, complete genome	NC_008297.1	Illumina	10x	6292	0.0274785023954162	1x	58	0.000254334651780781
Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus; New world arenaviruses	Chupare virus segment L, complete sequence	NC_010563.1	Illumina	10x	5734	0.0250415977010992	10x	5729	0.0251221244836568
Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus; New world arenaviruses	Chupare virus segment S, complete sequence	NC_010562.1	Illumina	10x	2705	0.0118133103900992	10x	2699	0.0118353314682126
Viruses; dsDNA viruses, no RNA stage; Papillomaviridae; Deltapapillomavirus	European elk papillomavirus, complete genome	NC_001524.1	Illumina	10x	6531	0.0285222662340215	0.925x	6525	0.0286126483253379
Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus; New world arenaviruses	Guanarito virus segment L, complete genome	NC_005082.1	Illumina	10x	5713	0.0249498862341088	5x	2850	0.0124974785788832
Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus; New world arenaviruses	Guanarito virus segment S, complete genome	NC_005077.1	Illumina	10x	2694	0.0117652710510571	5x	1346	0.00590231795339537
Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales; Coronaviridae; Coronavirinae; Betacoronavirus.	Human coronavirus HKU1, complete genome	NC_006577.2	Illumina	10x	24108	0.105284764105005	0.5x	1205	0.00528402164475588
Viruses; Retro-transcribing viruses; Retroviridae;Orthoretrovirinae; Lentivirus; Primate lentivirus group.	Human immunodeficiency virus 1, complete genome	NC_001802.1	Illumina	10x	7401	0.0323217412950533	10x	3699	0.0162204116713295
Viruses; Retro-transcribing viruses; Retroviridae;Orthoretrovirinae; Lentivirus; Primate lentivirus group.	Human immunodeficiency virus 2, complete genome	NC_001722.1	Illumina	10x	8352	0.0364749605859053	5x	8348	0.03660664953562
Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;Betapapillomavirus.	Human papillomavirus - 5, complete genome	NC_001531.1	Illumina	10x	6248	0.0272863450360077	5x	3120	0.0136814502337248
Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;Alphapapillomavirus	Human papillomavirus type 32, complete genome	NC_001586.1	Illumina	10x	6420	0.0280375056227864	20x	12825	0.0562386536049744
Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus; New world arenaviruses	Junin virus segment L, complete genome	NC_005080.1	Illumina	10x	5740	0.0250678009773822	3x	1718	0.00753356778895486
Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus; New world arenaviruses	Junin virus segment S, complete genome	NC_005081.1	Illumina	10x	2749	0.0120054677503177	3x	823	0.00360892100716522
Viruses; ssRNA negative-strand viruses; Mononegavirales;Filoviridae;	Lake Victoria marburgvirus - Musoke, complete genome	NC_001608.3	Illumina	10x	15392	0.067220138091266	1x	1539	0.00674863843259693
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;Flavivirus; tick-borne encephalitis virus group.	Langat virus, complete genome	NC_003690.1	Illumina	10x	8810	0.0384751440088392	1x	879	0.00385448549853977
Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus; New world arenaviruses	Machupo virus segment L, complete genome	NC_005079.1	Illumina	10x	5805	0.0253516698037811	2x	1158	0.00507792287520939
Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus; New world arenaviruses	Machupo virus segment S, complete genome	NC_005078.1	Illumina	10x	2771	0.012101546430022	2x	554	0.00242933443252677
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;Flavivirus;tick-borne encephalitis virus group.	Omsk hemorrhagic fever virus, complete genome	NC_005062.1	Illumina	10x	8685	0.0379292424196105	5x	4350	0.0190750988835586
Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus; New world arenaviruses	Sabia virus segment L, complete genome	NC_006313.1	Illumina	10x	5756	0.0251376763808035	1x	575	0.00252142111679223
Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus; New world arenaviruses	Sabia virus, complete genome	NC_006317.1	Illumina	10x	2713	0.0118482480926198	1x	273	0.00119712689545092
Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales; Coronaviridae; Coronavirinae; Betacoronavirus.	SARS coronavirus, complete genome	NC_004718.3	Illumina	10x	23961	0.104642783836072	1x	2390	0.0104803416854494
Viruses; Retro-transcribing viruses; Retroviridae;Orthoretrovirinae; Lentivirus; Primate lentivirus group.	Simian immunodeficiency virus, complete genome	NC_001549.1	Illumina	10x	7756	0.0338721018084628	1x	774	0.00339405207721249
Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae	Sudan ebolavirus, complete genome	NC_006432.1	Illumina	10x	15201	0.0663860004629245	100x	152050	0.666751444883927