Application 1: metagenomics read classification with NCBI/refseq bacterial genome database

Ready-built indexes:

20mer index: https://drive.google.com/open?id=0BxgO-FKbbXRIYWREa2NwejlVYUU

25mer index: https://drive.google.com/open?id=0BxgO-FKbbXRIY1pRaHJsYVg5dTQ

31mer index: https://drive.google.com/open?id=0BxgO-FKbbXRIa0Flc3Q4bWtycGM

If you want to build index with your own reference sequences,

Step1: Preparing Kmer files for each reference sequence using jellyfish : http://www.cbcb.umd.edu/software/jellyfish/

Step1.1: get Kmer count file:

Command: jellyfish count –o <path_to_bacterial_rawKmerCountFile> -m <Kmer_length> -t <threads_num> -s <bf_size> -C <path_to_bacterial_referenceSeqFastaFile>

Step1.2: dump to human-readable format

Command: jellyfish dump –t –c –o <path_to_bacterial_readableKmerCountFile> <path_to_bacterial_rawKmerCountFile>

Step1.3: put all readable Kmer count files into the same directory <path_to_bacterial_reference_seq_Kmer_file_dir> and rename them as 1.Kmer, 2.Kmer, …, m.Kmer, and generated a taxonomy info file like: https://drive.google.com/open?id=0BxgO-FKbbXRIZlV3ZzBBdlFpMTQ

There are three columns for each taxonomic rank in the file: the 1st column is a reissued id from 0 to m-1, where m is the total taxon num in that taxonomic rank. The 2nd column lists taxon ids and the 3rd column lists taxon scientific names. Each raw represents a species and its associated taxonomy info.

Step2: run "make build"

Step3: ./build <bacterial_ reference_seq_associated_taxonomy_info_file (generated in Step1.3)> <path_to_bacterial_reference_seq_Kmer_file_dir> .Kmer <Kmer_length> 6 <path_to_bacterial_index> <path_to_a_temp_dir_for_intermediate_files>

Classification:

Step1: run "make assignMetagenomicsRead_allTaxoRank_12_w2"

Step2:
```
./assignMetagenomicsRead_allTaxoRank_12_w2 <path_to_bacterial_index>
<path_to_output_results_dir> <Kmer_length> <threads_num> <fa_or_fq>
<SE_or_PE> <bacterial_speciesId2taxoInfo_file> <NCBI_names_file>
<readFile_singleEnd or readFile_end1> (<readFile_end2>)
```

<bacterial_speciesId2taxoInfo_file> can be downloaded at: https://drive.google.com/open?id=0BxgO-FKbbXRIc3FkLVFvMlpVVGM

<NCBI_names_file> can be downloaded at: https://drive.google.com/open?id=0BxgO-FKbbXRIUFI2dHlBMXZhdTA

Application 2: metagenomics read classification with NCBI/refseq human + virus genome database

Ready-built indexes:

20mer index: https://drive.google.com/open?id=0BxgO-FKbbXRITFlzUDRjdjEydzg

If you want to build index with your own reference sequences,

Step1: Preparing Kmer files for each reference sequence using jellyfish :
http://www.cbcb.umd.edu/software/jellyfish/

Step1.1: get Kmer count file:

Command: jellyfish count –o <path_to_human_virus _rawKmerCountFile> -m <Kmer_length> -t <threads_num> -s <bf_size> -C <path_to_human_virus _referenceSeqFastaFile>

Step1.2: dump to human-readable format

Command: jellyfish dump –t –c –o <path_to_human_virus_readableKmerCountFile> <path_to_human_virus_rawKmerCountFile>

Step1.3: put all readable Kmer count files into the same directory and rename them as 1.Kmer, 2.Kmer, …, m.Kmer, and generated a taxonomy info file like: https://drive.google.com/open?id=0BxgO-FKbbXRILVJKdVlBcXNTcGc

There are three columns for each taxonomic rank in the file: the 1st column is a reissued id from 0 to m-1, where m is the total taxon num in that taxonomic rank. The 2nd column lists taxon ids and the 3rd column lists taxon scientific names. Each raw represents a species and its associated taxonomy info.

Step2: run "make build"

Step3: ./build <human_virus_reference_seq_associated_taxonomy_info_file (generated in Step1.3)> <path_to_ human_virus_reference_seq_Kmer_file_dir> .Kmer <Kmer_length> 6 <path_to_ human_virus_index> <path_to_a_temp_dir_for_intermediate_files>

Classification:

Step1: run "make `assignMetagenomicsRead_allTaxoRank_13_w2`"

Step2:
```
./assignMetagenomicsRead_allTaxoRank_13_w2 <path_to_human_virus_index>
<path_to_output_results_dir> <Kmer_length> <threads_num> <fa_or_fq>
<SE_or_PE> <human_virus_speciesId2taxoInfo_file> <NCBI_names_file>
<readFile_singleEnd or readFile_end1> (<readFile_end2>)
```

```
<human_virus_speciesId2taxoInfo_file> can be downloaded at:
https://drive.google.com/open?id=0BxgO-FKbbXRIM0tFR3pHclpOWmc
```

```
<NCBI_names_file> can be downloaded at:
https://drive.google.com/open?id=0BxgO-FKbbXRIUFI2dHlBMXZhdTA
```