Jul 29, 2022

Support Protocol 2: Metagenotyping with a Custom Collection of Genomes

In 1 collection

miriam.goldman [1,2], chunyu.zhao [3,4]

[1]Data Science and Biotechnology, Gladstone Institutes, San Francisco, CA, USA,;

[2]Biomedical Informatics, University of California San Francisco, San Francisco, CA;

[3]Data Science, Chan Zuckerberg Biohub, San Francisco, CA, USA,;

[4]Data Science and Biotechnology, Gladstone Institutes, San Francisco, CA, USA

dx.doi.org/10.17504/protocols.io.dm6gpbp65lzp/v1

miriam.goldman

ABSTRACT

This protocol describes how to build a MIDASDB from a custom collection of genomes and perform SNV metagenotyping with it. Other MIDAS2 commands can also be run with the new database. Single-sample SNV metagenotyping is shown as an example. There are three steps: construct a custom rep-genome database for a collection of representative genomes of interest, build a bowtie2 index, and execute run_snps command with the prebuilt bowtie2 index.

DOI

dx.doi.org/10.17504/protocols.io.dm6gpbp65lzp/v1

PROTOCOL CITATION

miriam.goldman , chunyu.zhao 2022. Support Protocol 2: Metagenotyping with a Custom Collection of Genomes. **protocols.io** https://dx.doi.org/10.17504/protocols.io.dm6gpbp65lzp/v1

COLLECTIONS

MIDAS 2 Protocol

CREATED

Jul 28, 2022

LAST MODIFIED

Jul 29, 2022

PROTOCOL INTEGER ID

67837

PARENT PROTOCOLS

Part of collection

MIDAS 2 Protocol

1 Download the example genome collection folder from Zenodo to the work directory (midas2_protocol)

```
wget https://zenodo.org/record/6774976/files/midasdb_custom.zip
unzip midasdb_custom.zip
```

We have prepared two genomes for two species from the 21 NCBI genomes in the HMP mock community: *Staphylococcus aureus* (GCF_000013425.1) and *Staphylococcus epidermidis* (GCF_006094375.1) for this protocol.

- genomes.tsv: the table-of-contents file specifying the assignment of representative genomes to species. We randomly assign each species a six-digit species_id.
- cleaned_imports/: the FASTA file of each representative genome, saved in the directory <species>/<genome>/<genome>.fna.
- metadata.tsv: taxonomic assignment of the randomly assigned species_id.

MIDAS2 reserves the --midasdb_name newdb for building any new MIDASDB, and the custom MIDASDB will be built at --midasdb_dir midasdb_custom

2 Construct rep-genome component of MIDASDB
Annotate all the genomes and build the files needed for the rep-genome database. These commands should be executed in the work directory midas2_protocol.

Citation: miriam.goldman , chunyu.zhao Support Protocol 2: Metagenotyping with a Custom Collection of Genomes https://dx.doi.org/10.17504/protocols.io.dm6gpbp65lzp/v1

```
midas2 annotate_genome --species all \
  --midasdb_name newdb --midasdb_dir midasdb_custom \
  --debug --force
```

```
midas2 build_midasdb --generate_gene_feature \
  --genomes all \
  --midasdb_name newdb --midasdb_dir midasdb_custom \
  --debug --force
```

There are two command-line parameters that users need to pass:
--debug: keep the local file after successfully build the database
--force: re-build the database even if one already exists locally

3 Build one bowtie2 index with the representative genomes.

```
midas2 build_bowtie2db \
  --midasdb_name newdb --midasdb_dir midasdb_custom \
  --species_list 100001,100002 \
  --bt2_indexes_name repgenomes \
  --bt2_indexes_dir bt2_index_custom \
  --num_cores 8
```

We build therep-genome bowtie2 index for the two species specified via --species_list to the local directory bt2_index_custom/. Note we need to provide the custom MIDASDB midasdb_custom/ to –-midasdb_dir.
Users can also specify --bt2_indexes_name pangenomes specify to build the bowtie2 index for pangenomes.

4 Execute run_snps with the rep-genome database. For each sample, this code performs single-sample SNV calling for all the species in the bowtie2 database without any species filters (--select_threshold=-1). The number of CPUs used is specified via --num_cores 8.

Citation: miriam.goldman , chunyu.zhao Support Protocol 2: Metagenotyping with a Custom Collection of Genomes
https://dx.doi.org/10.17504/protocols.io.dm6gpbp65lzp/v1

```
for sample_name in SRR172902 SRR172903
do
   midas2 run_snps \
   --sample_name ${sample_name} \
   -1 reads/${sample_name}.fastq.gz \
   --midasdb_name newdb --midasdb_dir midasdb_custom \
   --prebuilt_bowtie2_indexes bt2_index_custom/repgenomes \
   --prebuilt_bowtie2_species bt2_index_custom/repgenomes.species \
   --select_threshold=-1 \
   --num_cores 8 midas2_output_custom
done
```

5 Confirm midas2 run_snps has finished successfully.
Once the single-sample SNV analysis is complete without any reported error, check for the output files (see Basic Protocol 3).

Citation: miriam.goldman , chunyu.zhao Support Protocol 2: Metagenotyping with a Custom Collection of Genomes
https://dx.doi.org/10.17504/protocols.io.dm6gpbp65lzp/v1