AMA

Requirements

- Bioconda ([Go to the official documentation](#)).

It is critical to include the 'bioconda channel' in addition to the other channels as indicated in the [official manual](#).

Installation

- Use the `requirements.txt` provided in the package and create a virtual environment.

- Activate the environment and run the setup script.

```
$ cd AMA.v1.2.0/
$ conda create --name env --file requirements-linux-64.txt -y
$ conda activate env
$ perl setup.pl
```

This tool was tested on "Ubuntu 20.04.4 LTS" with "conda 4.11.0" using the installation procedure mentioned above.

Troubleshooting

- Error installing Bioconda:

  The user may need to use `conda search <tool name>` to find the supported tool and then manually install it by using `conda install <tool name>`. Note: On macOS and Linux, the supported packages and their dependencies aren't always the same. Even when all of the requirements are completely aligned, the set of available versions isn't necessarily the same.

- Error using CPAN modules:

  Before beginning the setup, Mac OS users must ensure that they have write permission to the `/Users/\*/.cpan/` directory, and Linux users must ensure that their CPAN is properly configured.

Example usage

- To get usage info: `perl ama.pl --help`

- Command: `perl ama.pl --input example/SraRunInfo.csv --output src/main/test/ --label test_run --mode screen --config conf.txt`

Configuration file

Configuration file `conf.txt` requires the absolute path of the tools, blast database and bowtie index.

Before starting the analysis, please specify the path to your preferred *blast database* (`db_path`) and *bowtie index* (`bowtie_index_path`) in the configuration file.

The user can choose between *blastn* or *bowtie2* by changing the 'execute flag' to either 0 or 1 in the configuration file while leaving the rest of the parameters to default values. By default, both the tools are enabled *ie.* `execute = 1`. For additional information, please refer to the `example` directory provided in this package.

Note: If the user wishes to use a different installation than Bioconda, the user can manually install and specify the path of the tools in the configuration.

Pipeline parameters

- `--input` The user can provide input in either of the following ways:

  - A single SRA run accession. eg: `perl ama.pl --input SRR12548227 --output src/main/test/ --label test_run --mode screen --config conf.txt`

  - A list of run accessions in a text file (1 run accession per line). eg: `perl ama.pl --input example/list.txt --output src/main/test/ --label test_run --mode screen --config conf.txt`

  - The SRA runInfo exported directly from the NCBI-SRA web portal. Goto the SRA homepage and search for the desired keyword. Export the `SraRunInfo.csv` by clicking 'Send to' => File => RunInfo). eg: `perl ama.pl --input example/SraRunInfo.csv --output src/main/test/ --label test_run --mode screen --config conf.txt`

- `--output` eg: `src/main/test` The output directory to store the results.

- `--label` eg: `test_run` Provide any name for the analysis. This prefix will be used to create a parent working directory to store all the results, summary and log files.

- `--mode` eg: `screen`. Choose one of the three modes to run the pipline. The `screen` mode will only download a fraction of the dataset per SRA-run accession and analyse the file as per the given configuration. The `full` mode will execute the pipeline by

downloading the complete fastq file per SRA-run accession. The **both** option searches for samples using a fraction of the data that meet the minimum alignment cutoff from either 'bowtie2' or 'blastn', and then automatically performs alignment by downloading the entire fastq file. The **summary** mode will generate a unified alignment summary by examining the output files created by either screen-mode or full-mode.

- **--config** eg: **conf.txt** Pipeline configuration.

List of Perl modules and tools incorporated in the pipeline

- Perl modules:
    - Config::Simple
    - Parallel::ForkManager
    - Log::Log4perl
    - Getopt::Long
    - Text::CSV
    - Text::Fuzzy
- Tools:
    - NCBI EDirect utilities
    - NCBI SRA Toolkit
    - FastQC
    - Trimmomatic
    - FASTX-Toolkit
    - NCBI Blast
    - Bowtie2
    - Samtools