
snpgenie Documentation

Damien Farrell

Oct 24, 2022

CONTENTS

1	Introduction	3
1.1	Command line tool	3
1.2	Current Features	3
1.3	Links	4
2	Installation	5
2.1	Linux	5
2.2	Windows	5
2.3	Mac OSX	5
3	Using snpgenie	7
3.1	Command Line	7
3.2	From Python	7
3.3	Desktop Application	8
4	snpgenie	9
4.1	snpgenie package	9
5	Indices and tables	11

Contents:

INTRODUCTION

snpgenie is a desktop and command line tool for microbial variant calling and phylogenetic analysis from raw read data. It is primarily written to be used with bacterial isolates of MBovis but can be applied to other species. This is in early stages of development. Anyone interested in using the software is encouraged to make suggestions on improving or adding features.

This software is written in Python and is developed with the Qt toolkit using PySide2. It was made on Ubuntu linux but is designed to also run on Windows 10 with a standalone application.

1.1 Command line tool

This tool works from the command line and via Python scripts. Unlike many other SNP calling pipelines, it is also designed to have a graphical user interface, which is in development.

1.2 Current Features

- load multiple fastq files and process together
- view fastq quality statistics
- trim reads
- align to reference
- view bam alignments
- call variants
- filter variants
- create SNP core multiple sequence alignment
- create phylogenetic tree

1.3 Links

<http://dmnfarrell.github.io/snpgenie>

INSTALLATION

2.1 Linux

With pip:

```
pip install -e git+https://github.com/dmnfarrell/snpgenie.git#egg=snpgenie
```

Install binary dependencies:

```
sudo apt install bcftools samtools bwa
```

2.2 Windows

A standalone installer will be used to deploy on windows.

2.3 Mac OSX

Not tested. You can try the Linux instructions possibly with bioconda for the binaries.

USING SNPGENIE

The program includes both a command line and graphical interface. Both will produce the same results.

3.1 Command Line

This will run the entire process based on a set of options given at the terminal:

```
-h, --help            show this help message and exit
-i FILE, --input FILE  input folder(s)
-l FILE, --labels FILE sample labels file, optional
-r FILE, --reference FILE
                        reference genome filename
-w, --overwrite        overwrite intermediate files
-m, --trim             trim fastq files
-q QUALITY, --quality QUALITY
                        trim quality
-t THREADS, --threads THREADS
                        cpu threads to use
-o FILE, --outdir FILE
                        Results folder
-v, --version          Get version
-s, --test             Do test run
```

Example:

```
snpgenie -r reference.fa -g reference.gff -i data_files -t 8 -o results
```

3.2 From Python

You can run a workflow from within Python:

```
from snpgenie import app
args = {'threads':8, 'outdir': 'results', 'labelsep':'-',
        'input':['/my/folder/',
                  '/my/other/folder'],
        'reference': None, 'overwrite':False}
```

(continues on next page)

(continued from previous page)

```
W = app.WorkFlow(**args)
st = W.setup()
W.run()
```

3.3 Desktop Application

This interactive tool is designed for those not comfortable with the command line and includes some additional features such as visualization of fastq qualities.

SNPGENIE

4.1 snpgenie package

4.1.1 Submodules

4.1.2 snpgenie.gui module

4.1.3 snpgenie.widgets module

4.1.4 snpgenie.tools module

4.1.5 snpgenie.app module

4.1.6 snpgenie.aligners module

4.1.7 snpgenie.plotting module

4.1.8 Module contents

INDICES AND TABLES

- genindex
- modindex
- search