
Covidor

Release 0.1.0

Thomas Cokelaer

Feb 09, 2022

CONTENTS

1	Overview	3
1.1	third-party tools	3
1.2	Installation	3
1.3	Testing	3
2	User guide and reference	5
2.1	User Guide	5
2.2	References	5
	Python Module Index	9
	Index	11

Python version Python 3.7, 3.8, 3.9

Source See <http://github.com/sequana/covidor>.

Issues Please fill a report on [github](#)

Platform This is currently only available for Linux distribution with bash shell (contributions are welcome to port the tool on MacOSX and other platforms)

OVERVIEW

```
pip install covidor
```

1.1 third-party tools

kraken2 and art_illumina are required for the simulation and analysis.

```
pip install damona
# for kraken2 only:
damona install kraken:2.0.9
# to install kraken2 and art_illumina altogether, there are both provided in the sequana_
↳ tools package:
damona install sequana_tools
```

1.2 Installation

For Users:

```
pip install covidor
```

For developers:

```
git clone git@github.com:sequana/covidor.git
cd covidor
pip install .
```

1.3 Testing

```
pytest -v ./tests
```


USER GUIDE AND REFERENCE

2.1 User Guide

Table of Contents

- *User Guide*
 - *Getting help*

2.1.1 Getting help

Coming soon

2.2 References

covidor.kmers

covidor.distance

`covidor.main`

2.2.1 kmer module

```
class KmerOverlap(kmer_len=35)
    from covidor import KmerOverlap ko = KmerOverlap() for fasta in fastas:
        ko.enumerate_unique_kmers(fasta)
    names = self.kmer_dict.keys() ko.get_list_specific_kmers(name)
    enumerate_unique_kmers(fasta_files)
        populate kmers for each fasta
    get_list_specific_kmers(name)
        Returns kmer specific (unique) to a fasta
```

get_proportion_kmer_specific(*name*)

Gives same results as `get_specific_kmers` but normalised and for a given name

get_proportion_reads_specific(*name*, *N=100*, *read_length=75*, *paired=False*)

```
import glob
fastas = glob.glob("databases/genomes/*fasta")
kmer_dict = get_all_kmers()
expected_percentage = get_proportion_reads_specific(kmer_dict, "NAME_ONE_FASTA")
```

get_specific_kmers()

Return dataframe with number of unique kmers in each fasta

2.2.2 distance module

class Distance

compute_pdistance(*alignment*, *mode='spike'*)

Compute distance between genomes on the spike gene

This takes as input the output of **mafft** tool on a set of input fasta file:

```
cat *fasta > in.fasta
mafft --auto in.fasta > alignment.fasta
```

Then:

```
from covidor.distance import Distance
d = Distance()
d.compute_pdistance("alignment.out")
```

distance is computed as the number of bases that are different along 2 sequences, normalised by the length (and multiplied by 100). Each sequence is identified by the starting and ending 20-bases long sequences.

plot(*alignment*, *mode='spike'*)

```
d = Distance()
d.plot("alignment.out")
```

2.2.3 Covidor standalone (script module)

covidor

This is the main entry point

```
covidor [OPTIONS] COMMAND [ARGS]...
```

Options

--version

Show the version and exit.

analyse

Example of mix of wuhan and beta (PE):

```
covidor --file1 4520_S4_L001_R1_001.fastq.gz --file2 4520_S4_L001_R2_001.fastq.gz
```

Example of omicron (SE):

```
fastq-dump SRR17673561 covidor analyse --file1 SRR17673561.fastq --db databases/default/
```

```
covidor analyse [OPTIONS]
```

Options

--file1 <file1>

Required

--file2 <file2>

--tag <tag>

--db <db>

--factor <factor>

stats

```
covidor stats [OPTIONS]
```

Options

--tag <tag>

--rl <rl>

--paired

--factor <factor>

PYTHON MODULE INDEX

C

`covidor.distance`, 6

`covidor.kmers`, 5

Symbols

--db
 covidor-analyse command line option, 7
 --factor
 covidor-analyse command line option, 7
 covidor-stats command line option, 7
 --file1
 covidor-analyse command line option, 7
 --file2
 covidor-analyse command line option, 7
 --paired
 covidor-stats command line option, 7
 --rl
 covidor-stats command line option, 7
 --tag
 covidor-analyse command line option, 7
 covidor-stats command line option, 7
 --version
 covidor command line option, 6

C

compute_pdistance() (*Distance method*), 6
 covidor command line option
 --version, 6
 covidor.distance
 module, 6
 covidor.kmers
 module, 5
 covidor-analyse command line option
 --db, 7
 --factor, 7
 --file1, 7
 --file2, 7
 --tag, 7
 covidor-stats command line option
 --factor, 7
 --paired, 7
 --rl, 7
 --tag, 7

D

Distance (*class in covidor.distance*), 6

E

enumerate_unique_kmers() (*KmerOverlap method*), 5

G

get_list_specific_kmers() (*KmerOverlap method*), 5
 get_proportion_kmer_specific() (*KmerOverlap method*), 5
 get_proportion_reads_specific() (*KmerOverlap method*), 6
 get_specific_kmers() (*KmerOverlap method*), 6

K

KmerOverlap (*class in covidor.kmers*), 5

M

module
 covidor.distance, 6
 covidor.kmers, 5

P

plot() (*Distance method*), 6