fastools Documentation

Release latest

Nov 13, 2021

Contents:

This package provides various tools for the analysis and manipulation of FASTA and FASTQ files.

Please see ReadTheDocs for the latest documentation.

CHAPTER 1

Installation

The software is distributed via PyPI, it can be installed with `pip`:

```
pip install fastools
```

1.1 From source

The source is hosted on GitHub, to install the latest development version, use the following commands.

```
git clone https://github.com/jfjlaros/fastools.git
cd fastools
pip install .
```

CHAPTER 2

Usage

This package provides two separate command line interfaces, one for splitting FASTA files on substring occurrence and the main interface that provides a large number of small conversion and manipulation procedures.

2.1 `fastools`

The `fastools` command line interface provides a large number of elementary procedures that can be chained to get more complex behaviour. The elementary procedures can be accessed as subcommands of the main program. To get the full list of subcommands, type:

```
fastools -h
```

To get help on a specific subcommand, e.g., `sanitise`, type:

```
fastools sanitise -h
```

As mentioned above, more complex behaviour can be obtained by chaining elementary command by using UNIX pipes. For example, Fastools has the subcommand `gen` for generating a random FASTA record and the subcommand `fa2fq` to convert a FASTA file to a FASTQ file. To combine these two subcommands, we do the following:

```
fastools gen - name description 60 | fastools fa2fq - output.fq
```

This produces a FASTQ file containing one random sequence. Similarly a dash (−) can always be used instead of a file name to use standard input or standard output.

2.1.1 Automatic detection of input formats

Some subcommands, like `lenfilt`, accepts both FASTA and FASTQ files as input. The output format will be set to the same type as the input format. So to use this command with FASTA files, we use:

```
fastools lenfilt -l 25 input.fa small.fa large.fa
```

and for FASTQ, we can use:

```
fastools lenfilt -l 25 input.fq small.fq large.fq
```

In both cases, sequences larger than 25 are written to the `large` file, the other sequences are written to the `small` file.

2.2 split_fasta

The `split_fasta` program splits a FASTA file based on the occurrence of markers. For more information, use the *help* option:

```
split_fasta -h
```

CHAPTER 3

Library

All public functions in the `fastools` module are directly usable from other programs. To access the library, simply import `fastools`:

```
from fastools import gen

# Make a random FASTA record and write it to `output.fa`.
handle = open('output.fa', 'w')
gen(10, handle, 'name', 'description')
```

CHAPTER 4

Contributors

- Jeroen F.J. Laros <J.F.J.Laros@lumc.nl> (Original author, maintainer)
- Martijn Vermaat <martijn@vermaat.name> (Contributor)
- Rick H. de Leeuw <R.H.de_Leeuw@lumc.nl> (Contributor)
- Sander H.B. Bollen <A.H.B.Bollen@lumc.nl> (Contributor)

Find out who contributed:

```
git shortlog -s -e
```