
peasel Documentation

Release 0.4.0-dev

Connor McCoy

August 01, 2015

1	Installation	3
2	Example Usage	5
2.1	Create an index file	5
2.2	Retrieving sequences from an index	5
2.3	Using a temporary index	5
3	API documentation	7
3.1	peasel Module	7
4	License	9
5	Indices and tables	11
	Python Module Index	13

Some python wrappers for a little bit of [Sean Eddy](#)'s excellent [Easel](#) library for sequence manipulation.

At present, it's just a Python API to the Simple Sequence Index (SSI) format for rapid sequence retrieval from large files.

Installation

peasel requires [Python 2.7](#), either [setuptools](#) or [distribute](#) and a working C compiler.

Development requires [Cython](#), tested with version 0.17.

To install:

```
python setup.py install
```

To run the unit tests:

```
python setup.py test
```

Example Usage

2.1 Create an index file

Use `peasel.create_ssi` to build a sequence index:

```
>>> import peasel
>>> peasel.create_ssi('my_big_sequence_file.fasta') # creates my_big_sequence_file.fasta.ssi
2 # Number of sequences indexed
```

2.2 Retrieving sequences from an index

Sequence-indexes support dict-like behavior:

```
>>> import peasel
>>> # Open the index
>>> index = peasel.open_ssi('my_big_sequence_file.fasta')
>>> index['sequence1']
<EaselSequence 0x7f38735b80f0 [name="sequence1";description="";length=5]>
>>> index.get('sequence1')
<EaselSequence 0x7f38735b8108 [name="sequence1";description="";length=5]>
>>> print index.get('missing_sequence')
None
```

2.3 Using a temporary index

If you'd prefer not to litter the filesystem with `.ssi` files, use the `temp_ssi` context manager:

```
>>> import peasel
>>> with peasel.temp_ssi('my_big_sequence_file.fasta') as index:
...     index['sequence1']
...
<EaselSequence 0x7ff15065a0f0 [name="sequence1";description="";length=5]>
```

API documentation

3.1 peasel Module

peasel

class `peasel.EaselSequence`

Wrapper for the Easel ESL_SQ object

__len__ (*self*)

`x.__len__() <==> len(x)`

Length of the sequence

__getitem__ (*self*, *s*)

Slice the sequence

Parameters *s* (*slice*) – Slice to get, e.g. `s[1:3]`

Returns *EaselSequence* sliced to the specified residues.

create (*name*, *residues*, *accession*, *description*)

Create a sequence

Parameters

- **name** (*str*) – Sequence name
- **seq** (*str*) – Sequence residues
- **acc** (*str*) – Sequence accession number
- **desc** (*str*) – Sequence description

Returns A new *EaselSequence*

name

Sequence identifier

seq

Sequence

`peasel.create_ssi` (*file_path*, *ssi_name=None*, *sq_format=SQFILE_UNKNOWN*)

Create a Simple Sequence Index for a file.

Parameters

- **file_path** – Path to the sequence file
- **ssi_path** – Path to the sequence SSI file. If not given, `.ssi` is appended to `file_path`.
- **sq_format** – File format.

`peasel.open_ssi` (*file_path*, *ssi_path=None*, *sq_format=SQFILE_UNKNOWN*)

Open a simple sequence index for a file.

Parameters

- **file_path** (*str*) – Path to the sequence file
- **ssi_path** (*str*) – Path to the sequence SSI file. If not given, `.ssi` is appended to `file_path`.
- **sq_format** – File format.

`peasel.read_fasta(path)`

Read sequences in FASTA format from a file.

Parameters **path** (*str*) – Path to file containing sequences in FASTA format.

Returns A generator of *EaselSequence* objects.

`peasel.read_seq_file(path, sq_format=SQFILE_UNKNOWN)`

Read sequences from `path`. This is a generator function.

Parameters **path** (*str*) – Path to sequence file

Returns Generator of *EaselSequence* objects.

`peasel.write_fasta(sequences, fp)`

Writes *sequences* to the open file handle `fp`

Parameters

- **sequences** – Iterable of *EaselSequence* objects
- **fp** – Open file-like object

License

Distributed under the GPLv3. Easel source code is distributed under the Janelia Farm License, included in the `easel-src` subdirectory.

Indices and tables

- `genindex`
- `modindex`
- `search`

p

peasel, [7](#)

Symbols

`__getitem__()` (peasel.EaselSequence method), 7

`__len__()` (peasel.EaselSequence method), 7

C

`create()` (peasel.EaselSequence method), 7

`create_ssi()` (in module peasel), 7

E

`EaselSequence` (class in peasel), 7

N

`name` (peasel.EaselSequence attribute), 7

O

`open_ssi()` (in module peasel), 7

P

`peasel` (module), 7

R

`read_fasta()` (in module peasel), 8

`read_seq_file()` (in module peasel), 8

S

`seq` (peasel.EaselSequence attribute), 7

W

`write_fasta()` (in module peasel), 8