
EpivizQuindex Documentation

Release latest

Yifan*Yang*

Aug 26, 2022

Contents

1	Contents	3
1.1	EpivizQuindex	3
1.2	Usage	3
1.3	Contributing	4
1.4	License	8
1.5	Contributors	9
1.6	Changelog	9
1.7	epivizquindex	9
2	Indices and tables	11
	Python Module Index	13
	Index	15

This is the documentation of **EpivizQuindex**.

Note: This is the main page of your project's [Sphinx](#) documentation. It is formatted in [reStructuredText](#). Add additional pages by creating `rst`-files in `docs` and adding them to the [toctree](#) below. Use then [references](#) in order to link them from this page, e.g. [Contributors](#) and [Changelog](#).

It is also possible to refer to the documentation of other Python packages with the [Python domain syntax](#). By default you can reference the documentation of [Sphinx](#), [Python](#), [NumPy](#), [SciPy](#), [matplotlib](#), [Pandas](#), [Scikit-Learn](#). You can add more by extending the `intersphinx_mapping` in your Sphinx's `conf.py`.

The pretty useful extension [autodoc](#) is activated by default and lets you include documentation from docstrings. Docstrings can be written in [Google style](#) (recommended!), [NumPy style](#) and [classical style](#).

1.1 EpivizQuindex

Genomic analysis pipelines and workflows often use specialized file formats for manipulating and quickly finding data on potential genomic regions of interest. These file formats contain an index as part of the specification and allows users to perform random access queries. When we have a collection of these files, it's time consuming to read every single file and extract the data for a region of interest. The goal with Quindex approach is to “index the index” from these files and provide fast access to large collections of genomic data across files.

1.2 Usage

To import the package, simply run:

```
from epivizquindex import EpivizQuindex
```

1.2.1 Create the index

Define the genome range, and set the path to a folder where you want to hold the index:

base_path should be a folder. If the path does not exist, Quindex will create the path.

Add files to index with a simple function call:

```
f1 = "/path_to_your_file/some.bigwig"
f2 = "/path_to_your_file/someOther.bigwig"
# adding file to index
index.add_to_index(f1)
index.add_to_index(f2)
```

1.2.2 Performe in-memory query

Once the index is created, invoke the query in a specific chromosome and range:

```
index.query("chr2", 0, 900000)
```

You can also specify which file you are looking for:

```
index.query("chr2", 0, 900000, file = f1)
```

1.2.3 Store and load computed index to disk

Store the index to disk and load index to memory with `to_disk()` and `from_disk()`. The path is the `base_path` parameter when creating the index.

```
# storing the precomputed index
index.to_disk()
# reading a precomputed index
index = EpivizQuindex.EpivizQuindex(genome, base_path=base_path)
index.from_disk()
```

1.2.4 Perform search without loading

We can also perform search without loading the index to memory:

```
memory = False
index = EpivizQuindex.EpivizQuindex(genome, base_path=base_path)
index.from_disk(load = memory)
index.query("chr2", 0, 900000, in_memory = memory)
```

1.2.5 Note

This project has been set up using PyScaffold 4.2.3. For details and usage information on PyScaffold see <https://pyscaffold.org/>.

1.3 Contributing

Welcome to EpivizQuindex contributor's guide.

This document focuses on getting any potential contributor familiarized with the development processes, but [other kinds of contributions](#) are also appreciated.

If you are new to using [git](#) or have never collaborated in a project previously, please have a look at [contribution-guide.org](#). Other resources are also listed in the excellent [guide created by FreeCodeCamp](#)¹.

Please notice, all users and contributors are expected to be **open, considerate, reasonable, and respectful**. When in doubt, [Python Software Foundation's Code of Conduct](#) is a good reference in terms of behavior guidelines.

1.3.1 Issue Reports

If you experience bugs or general issues with EpivizQuindex, please have a look on the [issue tracker](#). If you don't see anything useful there, please feel free to fire an issue report.

Tip: Please don't forget to include the closed issues in your search. Sometimes a solution was already reported, and the problem is considered **solved**.

New issue reports should include information about your programming environment (e.g., operating system, Python version) and steps to reproduce the problem. Please try also to simplify the reproduction steps to a very minimal example that still illustrates the problem you are facing. By removing other factors, you help us to identify the root cause of the issue.

1.3.2 Documentation Improvements

You can help improve EpivizQuindex docs by making them more readable and coherent, or by adding missing information and correcting mistakes.

EpivizQuindex documentation uses [Sphinx](#) as its main documentation compiler. This means that the docs are kept in the same repository as the project code, and that any documentation update is done in the same way was a code contribution.

When working on documentation changes in your local machine, you can compile them using `tox`:

```
tox -e docs
```

and use Python's built-in web server for a preview in your web browser (`http://localhost:8000`):

```
python3 -m http.server --directory 'docs/_build/html'
```

1.3.3 Code Contributions

Submit an issue

Before you work on any non-trivial code contribution it's best to first create a report in the [issue tracker](#) to start a discussion on the subject. This often provides additional considerations and avoids unnecessary work.

Create an environment

Before you start coding, we recommend creating an isolated [virtual environment](#) to avoid any problems with your installed Python packages. This can easily be done via either [virtualenv](#):

¹ Even though, these resources focus on open source projects and communities, the general ideas behind collaborating with other developers to collectively create software are general and can be applied to all sorts of environments, including private companies and proprietary code bases.

```
virtualenv <PATH TO VENV>
source <PATH TO VENV>/bin/activate
```

or [Miniconda](#):

```
conda create -n EpivizQuindex python=3 six virtualenv pytest pytest-cov
conda activate EpivizQuindex
```

Clone the repository

1. Create an user account on GitHub if you do not already have one.
2. Fork the project [repository](#): click on the *Fork* button near the top of the page. This creates a copy of the code under your account on GitHub.
3. Clone this copy to your local disk:

```
git clone git@github.com:YourLogin/EpivizQuindex.git
cd EpivizQuindex
```

4. You should run:

```
pip install -U pip setuptools -e .
```

to be able to import the package under development in the Python REPL.

5. Install [pre-commit](#):

```
pip install pre-commit
pre-commit install
```

EpivizQuindex comes with a lot of hooks configured to automatically help the developer to check the code being written.

Implement your changes

1. Create a branch to hold your changes:

```
git checkout -b my-feature
```

and start making changes. Never work on the main branch!

2. Start your work on this branch. Don't forget to add [docstrings](#) to new functions, modules and classes, especially if they are part of public APIs.
3. Add yourself to the list of contributors in `AUTHORS.rst`.
4. When you're done editing, do:

```
git add <MODIFIED FILES>
git commit
```

to record your changes in [git](#).

Please make sure to see the validation messages from [pre-commit](#) and fix any eventual issues. This should automatically use [flake8/black](#) to check/fix the code style in a way that is compatible with the project.

Important: Don't forget to add unit tests and documentation in case your contribution adds an additional feature and is not just a bugfix.

Moreover, writing a [descriptive commit message](#) is highly recommended. In case of doubt, you can check the commit history with:

```
git log --graph --decorate --pretty=oneline --abbrev-commit --all
```

to look for recurring communication patterns.

5. Please check that your changes don't break any unit tests with:

```
tox
```

(after having installed `tox` with `pip install tox` or `pipx`).

You can also use `tox` to run several other pre-configured tasks in the repository. Try `tox -av` to see a list of the available checks.

Submit your contribution

1. If everything works fine, push your local branch to GitHub with:

```
git push -u origin my-feature
```

2. Go to the web page of your fork and click "Create pull request" to send your changes for review.

Troubleshooting

The following tips can be used when facing problems to build or test the package:

1. Make sure to fetch all the tags from the upstream [repository](#). The command `git describe --abbrev=0 --tags` should return the version you are expecting. If you are trying to run CI scripts in a fork repository, make sure to push all the tags. You can also try to remove all the egg files or the complete egg folder, i.e., `.eggs`, as well as the `*.egg-info` folders in the `src` folder or potentially in the root of your project.
2. Sometimes `tox` misses out when new dependencies are added, especially to `setup.cfg` and `docs/requirements.txt`. If you find any problems with missing dependencies when running a command with `tox`, try to recreate the `tox` environment using the `-r` flag. For example, instead of:

```
tox -e docs
```

Try running:

```
tox -r -e docs
```

3. Make sure to have a reliable `tox` installation that uses the correct Python version (e.g., 3.7+). When in doubt you can run:

```
tox --version
# OR
which tox
```

If you have trouble and are seeing weird errors upon running `tox`, you can also try to create a dedicated [virtual environment](#) with a `tox` binary freshly installed. For example:

```
virtualenv .venv
source .venv/bin/activate
.venv/bin/pip install tox
.venv/bin/tox -e all
```

4. [Pytest](#) can drop you in an interactive session in the case an error occurs. In order to do that you need to pass a `--pdb` option (for example by running `tox -- -k <NAME OF THE FALLING TEST> --pdb`). You can also setup breakpoints manually instead of using the `--pdb` option.

1.3.4 Maintainer tasks

Releases

If you are part of the group of maintainers and have correct user permissions on [PyPI](#), the following steps can be used to release a new version for `EpivizQuindex`:

1. Make sure all unit tests are successful.
2. Tag the current commit on the main branch with a release tag, e.g., `v1.2.3`.
3. Push the new tag to the upstream [repository](#), e.g., `git push upstream v1.2.3`
4. Clean up the `dist` and `build` folders with `tox -e clean` (or `rm -rf dist build`) to avoid confusion with old builds and Sphinx docs.
5. Run `tox -e build` and check that the files in `dist` have the correct version (no `.dirty` or `git` hash) according to the `git` tag. Also check the sizes of the distributions, if they are too big (e.g., `> 500KB`), unwanted clutter may have been accidentally included.
6. Run `tox -e publish -- --repository pypi` and check that everything was uploaded to [PyPI](#) correctly.

1.4 License

The MIT License (MIT)

Copyright (c) 2022 Yifan_Yang

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

1.5 Contributors

- Yifan_Yang <yang7832@umd.edu>

1.6 Changelog

1.6.1 Version 0.1

- Feature A added
- FIX: nasty bug #1729 fixed
- add your changes here!

1.7 epivizquindex

1.7.1 epivizquindex package

Submodules

epivizquindex.EpivizQuindex module

epivizquindex.QuadTree module

```
class epivizquindex.QuadTree.Index (bbox=None, x=None, y=None, width=None,
                                     height=None, max_items=50, max_depth=30, disk=None,
                                     first_run=False)
```

Bases: epivizquindex.QuadTree._QuadTree

The wrapper of the root quad tree node, which represents a spatial index.

```
from_disk (f_path)
```

Constructs a quadtree index from a precomputed file and store it in the current node. Parameters: - **f_path**: a string containing the path to the precomputed index.

```
insert (item, bbox)
```

Inserts an item into the quadtree along with its bounding box. Parameters: - **item**: The item to insert into the index, which will be returned by the intersection method - **bbox**: The spatial bounding box tuple of the item, with four members (xmin,ymin,xmax,ymax)

```
intersect (bbox, in_memory=False, debug=False)
```

Intersects an input boundingbox rectangle with all of the items contained in the quadtree. Parameters: - **bbox**: A spatial bounding box tuple with four members (xmin,ymin,xmax,ymax) - **in_memory** (optional): A flag for using in_memory search with respect to file based search. - **debug** (optional): A flag that allows extra output when debugging. Returns: - A list of inserted items whose bounding boxes intersect with the input bbox.

```
read_header ()
```

```
to_disk (path)
```

Converts a quadtree index to file format and output it to disk. Parameter: - **path**: a string that contains the path to which the tree will be stored at.

`epivizquindex.QuadTree.box_contains(box1, box2)`

Determin whether the box 1 contains box 2.

Parameters: - **box1 (tuple)**: bottom left, top right coordinate of a bounding box. - **box2 (tuple)**: bottom left, top right coordinate of a bounding box.

Returns: - **box_contains (bool)**: whether the box 1 contains box 2.

`epivizquindex.QuadTree.box_intersect(box1, box2)`

Determin whether the two bounding box intersect.

Parameters: - **box1 (tuple)**: bottom left, top right coordinate of a bounding box. - **box2 (tuple)**: bottom left, top right coordinate of a bounding box.

Returns: - **box_intersect (bool)**: Whether the two bounding box intersect.

epivizquindex.sample module

epivizquindex.utils module

`epivizquindex.utils.get_genome(t)`

Get the range of chromosomes of the given type. Currently these are read from hgdownload.cse.ucsc.edu.

Parameters: - **t (str)**: type of gene.

Returns: - **genome (dict)**: Dictionary of the range of the chromosome for the given type.

`epivizquindex.utils.hcoords(x, chromLength, dims=2)`

Returns hilbert space coordinate of the input query.

Parameters: - **x (int)**: A integer representing the query location in a chromosome. - **chromLength (int)**: The total length of the chromosome. - **dims (int)**: dimension of the hilbert space, by default we use 2d hilbert space.

Returns: - **x (int)**: x coordinate in hilbert space. - **y (int)**: y coordinate in hilbert space. - **hlevel (int)**: level of the hilbert space.

`epivizquindex.utils.range2bbox(hlevel, query, dims=2, margin=0)`

Convert the input query chromosome range to query bounding box in hilbert space.

Parameters: - **hlevel (int)**: level of the hilbert space, calculated by the size of the chromosome. - **query (dict)**: dictionary of a query range, with “start” and “end” representing the query range in a chromosome. - **dims (int)**: dimension of the hilbert space, by default we use 2d hilbert space. - **margin (int)**: margin of the query box in hilbert space.

Returns: - **bbbox (tuple)**: 4 integers representing the lower left and top right coordinate. The order is lower left x, y, then top right x, y.

Module contents

CHAPTER 2

Indices and tables

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

e

`epivizquindex`, [10](#)

`epivizquindex.QuadTree`, [9](#)

`epivizquindex.utils`, [10](#)

B

`box_contains()` (*in module `epivizquindex.QuadTree`*), 9
`box_intersect()` (*in module `epivizquindex.QuadTree`*), 10

E

`epivizquindex` (*module*), 10
`epivizquindex.QuadTree` (*module*), 9
`epivizquindex.utils` (*module*), 10

F

`from_disk()` (*`epivizquindex.QuadTree.Index` method*), 9

G

`get_genome()` (*in module `epivizquindex.utils`*), 10

H

`hcoords()` (*in module `epivizquindex.utils`*), 10

I

`Index` (*class in `epivizquindex.QuadTree`*), 9
`insert()` (*`epivizquindex.QuadTree.Index` method*), 9
`intersect()` (*`epivizquindex.QuadTree.Index` method*), 9

R

`range2bbox()` (*in module `epivizquindex.utils`*), 10
`read_header()` (*`epivizquindex.QuadTree.Index` method*), 9

T

`to_disk()` (*`epivizquindex.QuadTree.Index` method*), 9