

NPLinker Community Meeting

2024-07-02

netherlands
eScience center

Agenda

- 16:00 Short Introduction
- 16:05 Announcements and upcoming events
- 16:10 NPLinker development
- 16:25 Monthly plan
- 16:30 Q&A
- 17:00 .



Announcements and upcoming events?



NPLinker Development

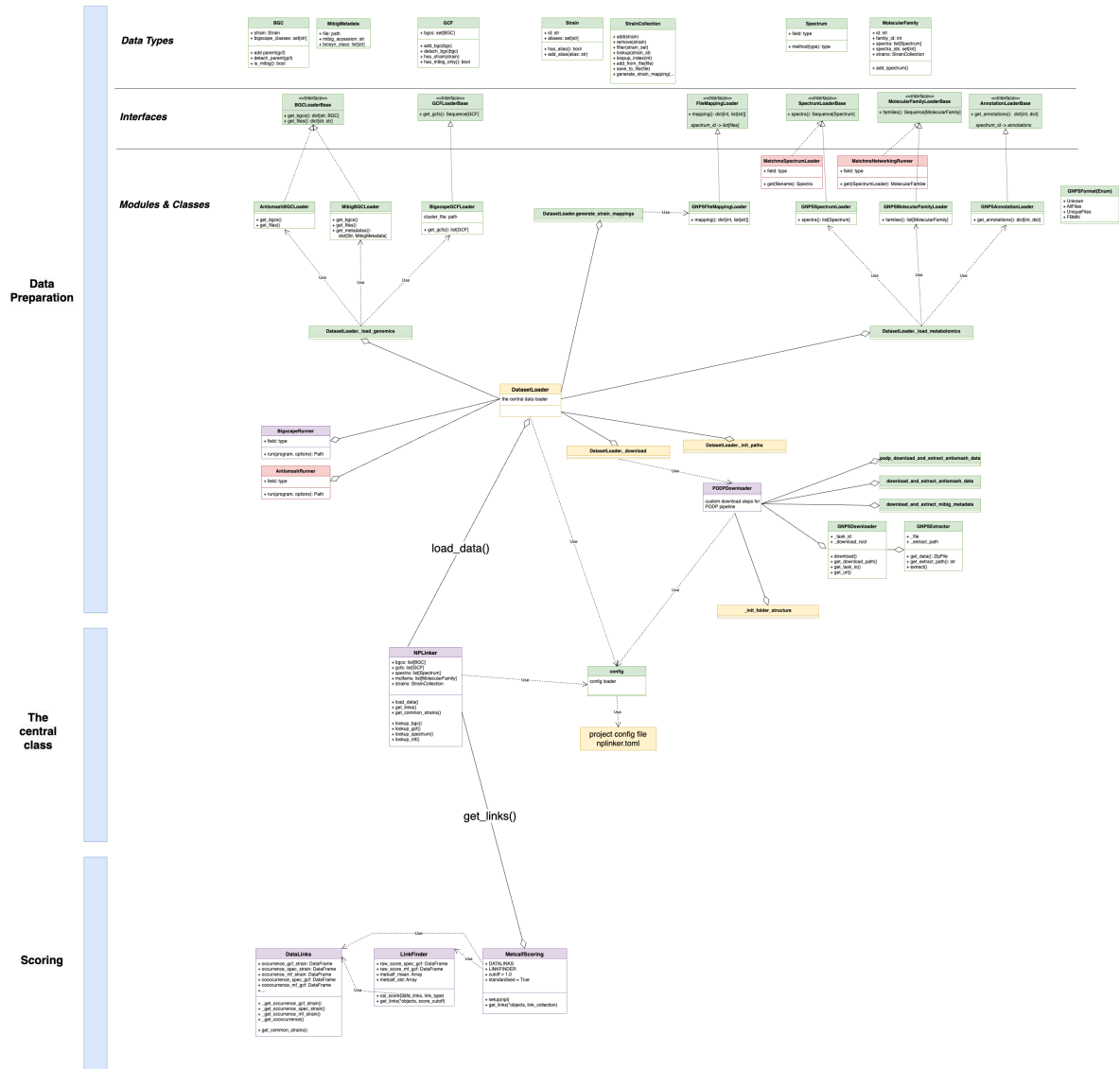
New features & changes

Contributors: Cunliang Geng, Giulia Crocioni, Arjan Draisma

Refactoring progress

Updated: 2024-02-05

Ideas Todo In progress Done



Refactor data preparation (✅ Q1)

- Arranging data
- Loading data

version
2.0.0-a2

✅ Refactor NPLinker class

Refactor scoring methods

- ✅ Metcalf scoring method

New logging

- use the standard module `logging` to replace the class `LogConfig`
- added function `setup_logging` to easily setup the logging
- added a new library `rich` to colourise log message and to replace the `tdqm` library



```
(npl3.9) → nplinker_local_mode_example python run.py
[2024-07-01 16:52:09] INFO Configuration: nplinker.py:85
{'LOG': {'use_console': True,
        'level': 'DEBUG'},
 'MIBIG': {'version': '3.1',
          'to_use': True},
 'BIGSCAPE': {'cutoff': '0.30',
              'parameters': {'--mibig': '
                              --clans-off '
                              --mix '
                              --include_singletons '
                              --cutoffs '
                              '0.30'},
              'SCORING': {'methods': ['metcalf']},
              'ROOT_DIR':
PosixPath('/Users/clgeng/github/nplinker/nplinker_local_mode_exa
mple'),
              'MODE': 'local'}

[2024-07-01 16:52:09] INFO Using downloaded and verified file: utils.py:174
/Users/clgeng/github/nplinker/nplinker_local_mode_example/download
s/mibig_json_3.1.tar.gz

[2024-07-01 16:52:09] INFO Extracting utils.py:420
/Users/clgeng/github/nplinker/nplinker_local_mode_example/download
s/mibig_json_3.1.tar.gz to
/Users/clgeng/github/nplinker/nplinker_local_mode_example/mibig

[2024-07-01 16:52:09] INFO Loaded 26 non-MiBIG Strain objects loader.py:102
[2024-07-01 16:52:09] INFO Loaded 26 Strain objects in total loader.py:113
[2024-07-01 16:52:09] INFO ===== loader.py:127

Loading metabolomics data starts...
[2024-07-01 16:52:23] INFO 24652 Spectrum objects updated with Strain objects. utils.py:67
1283 Spectrum objects not updated with Strain objects.
[2024-07-01 16:52:23] INFO 29 MolecularFamily objects updated with Spectrum objects. utils.py:122
0 MolecularFamily objects not updated with Spectrum objects.
[2024-07-01 16:52:23] INFO 0 MolecularFamily objects have missing Spectrum objects.
Loading metabolomics data completed loader.py:154

===== loader.py:168

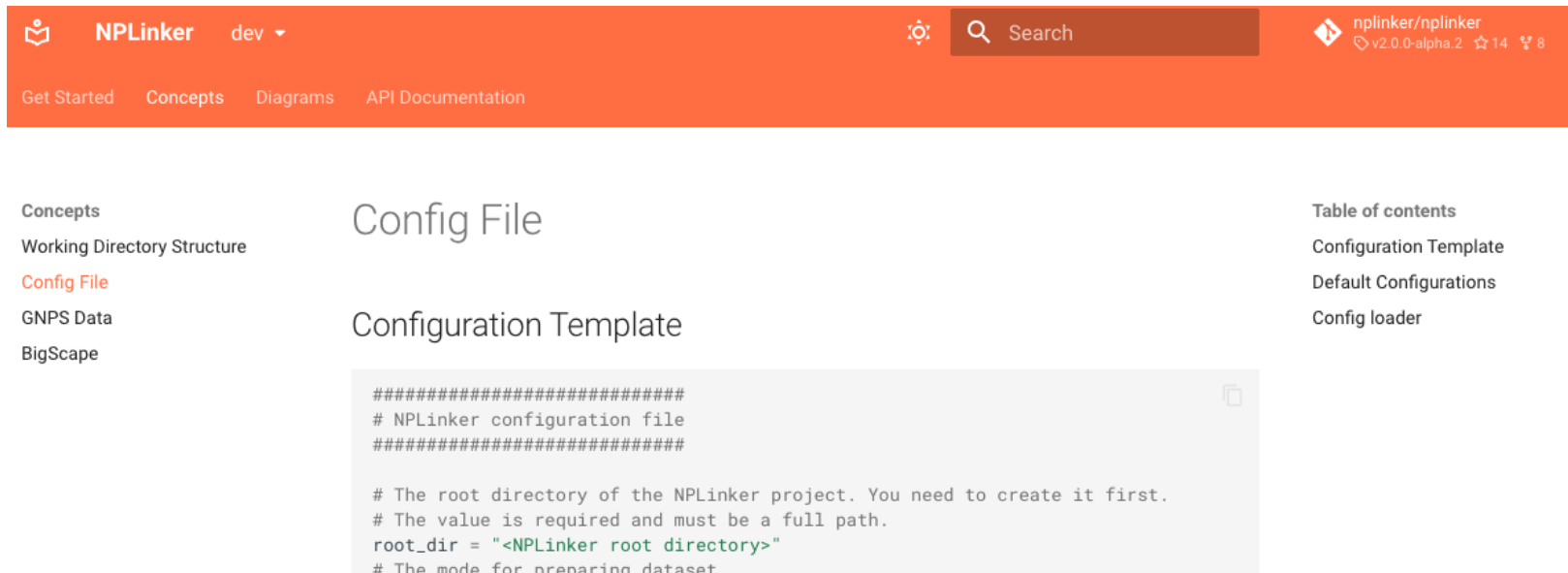
Loading genomics data starts...
[2024-07-01 16:52:23] INFO Parsing AntiSMASH directory... loader.py:171
[2024-07-01 16:52:24] INFO 370 BGC objects updated with Strain object. utils.py:110
0 BGC objects not updated with Strain object.
[2024-07-01 16:52:24] INFO Loading BigSCAPE cluster file loader.py:197
/Users/clgeng/github/nplinker/nplinker_local_mode_example/bigscap
e/mix_clustering_c0.30.tsv
[2024-07-01 16:52:24] INFO 64 GCF objects updated with BGC objects. utils.py:164
0 GCF objects not updated with BGC objects.
[2024-07-01 16:52:24] INFO 0 GCF objects have missing BGC objects.
Loading genomics data completed loader.py:223
```

Specify config file explicitly

Now you must provide the config file explicitly when initializing the NPLinker app.

```
from nplinker import NPLinker

# create an instance of NPLinker
npl = NPLinker("nplinker.toml")
```



The screenshot shows the NPLinker documentation website. The top navigation bar is orange and contains the NPLinker logo, a dropdown menu set to 'dev', a search bar, and a GitHub repository link for 'nplinker/nplinker' showing version 'v2.0.0-alpha.2' with 14 stars and 8 forks. Below the navigation bar, there are links for 'Get Started', 'Concepts', 'Diagrams', and 'API Documentation'. The 'Concepts' section is expanded, showing links for 'Working Directory Structure', 'Config File' (highlighted in red), 'GNPS Data', and 'BigScape'. The 'Config File' section is titled 'Config File' and contains a 'Configuration Template' section. The template shows a TOML configuration file structure with comments and a placeholder for the root directory.

```
#####
# NPLinker configuration file
#####

# The root directory of the NPLinker project. You need to create it first.
# The value is required and must be a full path.
root_dir = "<NPLinker root directory>"
# The mode for preparing dataset.
```



Refactored scoring (metcalf)

- Removed the classes LinkFinder, DataLinks, ObjectLink and LinkCollection
- Added classes `LinkGraph` to represent links
- Added classes `Score` to represent scores
- Added abstract base class for scoring methods
- Updated class `MetcalfScoring`
- Updated class `NPLinker`

```
# compute the links for the first 3 GCFs using metcalf scoring method
link_graph = npl.get_links(npl.gcfs[:3], "metcalf") +

# get links as a list of tuples
link_graph.links

# get the link data between two objects or entities
link_graph.get_link_data(npl.gcfs[0], npl.spectra[0])
```



Updated docs

NPLinker

Get Started

Concepts

Diagrams

Dataset Arranger

Dataset Loader

API Documentation

Get Started

Concepts

Diagrams

API Documentation

API Documentation

NPLinker

Dataset Arranger

Dataset Loader

Genomics Data

Metabolomics Data

Strain Data

Scoring

Schemas

General Utilities

NPLinker

mod

nplinker

class

NPLinker

NPLinker

(config_file: str | PathLike)

Main class for the NPLinker application.

Attributes:

Name	Type	Description
config		The configuration object for the current NPLinker application.
root_dir	str	The path to the root directory of the current NPLinker application.
output_dir	str	The path to the output directory of the current NPLinker application.
bgcs	list[BGC]	A list of all BGC objects.
gcfs	list[GCF]	A list of all GCF objects.

Table of contents

mod

nplinker

class

NPLinker

attr

config

attr

root_dir

attr

output_dir

attr

bgcs

attr

gcfs

attr

spectra

attr

mfs

attr

mibig_bgcs

attr

strains

attr

product_types

attr

chem_classes

attr

class_matches

attr

scoring_methods

meth

load_data

meth

get_links

meth

lookup_bgc

meth

lookup_gcf

meth

lookup_spectrum

meth

lookup_mf

meth

save_data

func

setup_logging

Integrating BigScape V2

Support for running BigScape V2 is coming soon

```
def test_run_bigscape_v1(self, tmp_path):
    result = run_bigscape(
        antismash_path=DATA_DIR,
        output_path=tmp_path,
        extra_params="--help",
        version=1,
    )

    assert result is True

def test_run_bigscape_v2(self, tmp_path):
    result = run_bigscape(
        antismash_path=DATA_DIR,
        output_path=tmp_path,
        extra_params="--help",
        version=2,
    )

    assert result is True
```



Monthly plan

What will happen before next meeting

Develop the webapp

NPLinker

1 Import dataImportDocAbout

2 Choose the tab according to the data you want to start with, e. g. Genomics data

Genomics -> MetabolomicsMetabolomics -> Genomics

3 Filter data using filters (optional)

Genomics filter

GCF IDType

GCF ID

Enter one or more GCF IDs

Type

Enter one or more compound types

Foldable tab

4a Inspect the data (optional)

4b Select data

2 GCFs selected

GCF ID	# BGC
<input checked="" type="checkbox"/> 1	3
<input checked="" type="checkbox"/> 2	1
<input checked="" type="checkbox"/> 3	10
<input checked="" type="checkbox"/> 4	6

Rows per page 20 <>

5 Set scoring methods and parameters (optional)

Scoring

MetcalfRosetta

RAWSTANDARDIZED

Cutoff1

Spectral score cutoff0

BGC score cutoff0

Logical operation between scoring methodsANDOR

This format indicates hyperlinks

All cutoff forms display default value or valid value range

The operation option appears only when users set two or more scoring methods

6 Click to show the links found

Show spectra

7 Inspect the results

Candidate Links

Select columns

GCF ID	# links	Top 1 spectrum	Product	Score
1	300	100	xxx	0.8
4	20	20	yyy	0.6
3	400	45	zzz	0.5

Rows per page 3 <>

Need your tests and feedback on the refactored NPLinker

Questions? Ideas? Feedback?

Need your tests and feedback on the refactored NPLinker

Questions? Ideas? Feedback?