

Daily Records

Caccone PostDoc

December, 2014

Contents

1	2014-12-19	2
1.1	Argot2 batch size reduction:	2
1.1.1	Problems:	2
2	2014-12-21	2
2.1	Argot2 batch size reduction:	2
2.1.1	Completed:	2
2.1.2	Problems:	3
2.2	Working on:	3
3	2014-12-22 (Monday)	3
3.1	Argot2 batch size reduction:	3
3.1.1	Files:	3
3.1.2	Next:	4
4	2014-12-23 (Tuesday)	4
4.1	Meeting with Dan	4
4.2	Argot2 batch size reduction:	4
5	2014-12-24 (Wednesday)	4
5.1	Argot batch size reduction	4
5.1.1	Storing in Python-friendly format (pandas.HDFStore)	5
5.1.2	File location	5
5.1.3	PROJECT COMPLETED	5
5.2	Installing PyTables	5
5.3	T: Functional Annotations of genes near SNPs of interest	5

6	2014-12-26 (Friday)	6
6.1	T: Functional Annotations of genes near SNPs of interest	6

1 2014-12-19

1.1 Argot2 batch size reduction:

tags = [argot2, ddRAD58, python, spartan]

- dumping the whole proteome on the online server seems to have broken it.
- I believe I will need to split the input up into around 3000 proteins per submission
- this will need some python code (maybe added to spartan?)

1.1.1 Problems:

tags = [python, pip, virtualenvwrapper, py279, admin]

- cant run ipython notebook bc of py279 from py278 issues.
 - must reinstall most/all python dependencies.
-

2 2014-12-21

2.1 Argot2 batch size reduction:

- see 2014-12-19
- finished updating py279 python requirements
- ipython seems to work again

2.1.1 Completed:

- writing filter functions for blast and hmmer outputs:
 - spartan/src/spartan/utils/blast/output.py:
 - * filter_for_argot(path, protein_names)
 - spartan/src/spartan/utils/hmmer/output.py:
 - * filter_for_argot(path, protein_names)

- created sublime text 3 snippet to make new ipython notebook file with meta info and template.
- wrote and tested split up logic:

```

    - spartan/src/spartan/utils/misc.py:
      * split_stream(stream, divisor)

```

2.1.2 Problems:

- coding the split up logic

2.2 Working on:

- split blastp and hmmer outputs by protein (~3000 per group)
-

3 2014-12-22 (Monday)

3.1 Argot2 batch size reduction:

- see 2014-12-21
- finished splitting
- files were zipped
- file pairs were submitted to Argot2

3.1.1 Files:

```

Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.0.hmmscan.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.1.hmmscan.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.2.hmmscan.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.3.hmmscan.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.4.hmmscan.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.5.hmmscan.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.6.hmmscan.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.7.hmmscan.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.8.hmmscan.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.9.hmmscan.zip

```

Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.union.0.blastp.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.union.1.blastp.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.union.2.blastp.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.union.3.blastp.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.union.4.blastp.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.union.5.blastp.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.union.6.blastp.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.union.7.blastp.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.union.8.blastp.zip
Glossina-fuscipes-IAEA_PEPTIDES_GfusI1.union.9.blastp.zip

3.1.2 Next:

- Collect the resulting analysis files into single repository.
-

4 2014-12-23 (Tuesday)

4.1 Meeting with Dan

- [markdown-docs/notes/meetings/dan-2014-12-23/dan-2014-12-23.md](#)

4.2 Argot2 batch size reduction:

- see 2014-12-22
 - [2014-12-23_create_argot2_functional_annotation_db_GfusI1.1_prerelease.ipynb](#)
-

5 2014-12-24 (Wednesday)

5.1 Argot batch size reduction

- see 2014-12-23

5.1.1 Storing in Python-friendly format (pandas.HDFStore)

- had to install pytables
- test work well and data is stored
- documentation: [2014-12-24_store_argot2_functional_annotation_db_GfusI1.1_prerelease_as_HDF5.i](#)

5.1.2 File location

```
louise/data/genomes/glossina_fuscipes/annotations/ \
functional/GfusI1.1_pre/argot2_out/argot_functional_annotations_ts150.h5
```

5.1.3 PROJECT COMPLETED

“Argot batch size reduction” project is now considered completed.

5.2 Installing PyTables

First attempt failed due to cryptic or at *least* slightly misleading error about numpy and numexpr:

```
$ pip install git+https://github.com/PyTables/PyTables.git@v.3.1.1#egg=tables
```

Tried installing numexpr directly with:

```
$ pip install numexpr
```

Tried PyTables again:

```
$ pip install git+https://github.com/PyTables/PyTables.git@v.3.1.1#egg=tables
```

5.3 T: Functional Annotations of genes near SNPs of interest

- Project start.
-

6 2014-12-26 (Friday)

6.1 T: Functional Annotations of genes near SNPs of interest

- [] write code to create table of functional annotation info, given gene-names and annotation database.
 - [] draft in ipython notebook
 - [] copy to spartan
- [] write methods for the functional annotation paper section