Daily Records

Caccone PostDoc

December, 2014

# 2014-12-19

## 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?)

### Problems:

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

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

# 2014-12-21

## Argot2 batch size reduction:

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

### 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)

### Problems:

* coding the split up logic

## Working on:

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

# 2014-12-22 (Monday)

## Argot2 batch size reduction:

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

### 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

### Next:

* Collect the resulting analysis files into single repository.

# 2014-12-23 (Tuesday)

## Meeting with Dan

* [markdown-docs/notes/meetings/dan-2014-12-23/dan-2014-12-23.md](file:///home/gus/Dropbox/repos/git/markdown-docs/notes/meetings/dan-2014-12-23/dan-2014-12-23.md)

## Argot2 batch size reduction:

* see 2014-12-22
* [2014-12-23\_create\_argot2\_functional\_annotation\_db\_GfusI1.1\_prerelease.ipynb](http://localhost:8888/jupiter/notebooks/YALE/ddrad58/2014-12-23_create_argot2_functional_annotation_db_GfusI1.1_prerelease.ipynb)

# 2014-12-24 (Wednesday)

## Argot batch size reduction

* see 2014-12-23

### 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.ipynb](http://localhost:8888/jupiter/notebooks/YALE/ddrad58/2014-12-24_store_argot2_functional_annotation_db_GfusI1.1_prerelease_as_HDF5.ipynb)

### File location

louise/data/genomes/glossina\_fuscipes/annotations/ \  
functional/GfusI1.1\_pre/argot2\_out/argot\_functional\_annotations\_ts150.h5

### PROJECT COMPLETED

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

## 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

## T: Functional Annotations of genes near SNPs of interest

* Project start.

# 2014-12-26 (Friday)

## 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