Weekly Progress Report

Camille Scott

2015-03-15

Lamprey

- New abundance data is finished – need to incorporate into notebooks
- Office computer had a minor. . . failure. Will need to rerun some postprocessing (yay rep. pipelines. . .)
- Did some more comprehensive background reading

  1. "Programmed loss of millions of base pairs from a vertebrate genome": One of Jeramiah Smith's papers from 2009; a companion to the genome paper which was published while the genome project was ongoing.
  2. "Genetic Consequences of Programmed Genome Rearrangement": Another of Jeramiah's papers, this time from 2012. Has a more detailed description of specific genes targeted by PGR, and should be a really good source for validation. Shows that most deleted sequence is non-coding (but not necessarily non-transcribed)
  3. "Cytogenetic evidences of genome rearrangement and differential epigenetic chromatin modification in the sea lamprey": from 2014, European lab. Additional gene PGR gene targets, data on GC content of deleted regions.
  4. "Small RNA Expression During Programmed Rearrangement of a Vertebrate Genome": Master's thesis from one of Jeremiah's students, 2014. Still need to read through it.

  Notably, these papers all add gene targets that we can validate our sequence with, and none of them use the comprehensive RNA-seq data that we have. So, replicating some of their methods along with the germline data we should be getting soon might be fruitful.
- Did some more fiddly work with the metadata and pipelines (switching to JSON)

khmer multiprocessing

- Still hitting my head against the wall with deadlock issues. Could use a new set of eyes on it at this point.

khmer generally

- Optimized median-filter implementation. Need to add better battery of tests (planning on tomorrow, this should be quick). Consider exploring improved exact median algorithms in the future (this is low priority for me and likely of marginal benefit)

Other

- Continued work with Greg on the Teaching Tech Together workshop

- Some more work on Avida stuff w/ Alita (pipeline was a good testbed for switching my metadata handling over to JSON)