Analysis of Ribose-seq Libraries

Alli Gombolay, M.P.H
Storici Lab | School of Biology
Georgia Institute of Technology
Contact: alli.gombolay@gatech.edu

Last Updated: May 2016

Project Overview

Goal: To create a streamlined bioinformatics pipeline to analyze Ribose-seq libraries on any Linux platform

Background: The original scripts used to analyze Ribose-seq libraries required Platform Load Sharing Facility (LSF) workload management software to run properly. To enable the user to efficiently analyze their sequenced Ribose-seq libraries, I have created an updated, streamlined version of the Ribose-seq analysis pipeline that does not depend on LSF. This pipeline was designed to run on any Linux platform, and it consists of both Bash scripts and Python scripts. The pipeline includes a script to automatically download all required input files, the appropriate directory structure (as a GitHub repository), and the required software. Currently, the pipeline is tailored to analyze the DNA of either of four different common organisms, including *Saccharomyces cerevisiae* (nuclear and mitochondrial DNA), *Escherichia coli K12*, *Mus musculus* (mouse embroynic fibroblast DNA), and humans (HeLa cell DNA).

Significance: This pipeline allows any user to easily download and install all required software, download all required input files, and download the appropriate directory structure to analyze Ribose-seq libraries. Now, users can readily choose the species of interest and run the corresponding pipeline to map the distribution of rNMPs in their Ribose-seq libraries.

Usage

To display help menu:
```
wrapper -h
```

To run Ribose-seq analysis pipeline:
```
wrapper -g 'FASTQ.GZ Files of Interest' -r 'Bowtie Index'
```

References

- Ribose-seq *Nature Methods* Paper, 2015
- Jay Hesselberth's GitHub Page for Ribose-seq
- Georgia Tech 2015 News Article on Ribose-seq

Program Set-up

Part A: Software Set-up

- Script to automatically download required input files
- Script to automatically download and install software

Part B: Directory Set-up

1. Clone the Ribose-seq Analysis Pipeline Directory Structure:
   ```
   git clone https://github.com/agombolay/Ribose-seq-
   Project/tree/master/Ribose-seq-Directory
   ```