

# BMTRY 789-02 Spring 2012 Homework 18

Adrian Michael Nida

2012-04-03

A prominent researcher in the field of Molecular Pharmacology is trying to determine which amino acids are most prevalent in the rat population he is studying. He has created an R script `mineAminos.R` that finds all tri-nucleotide patterns within the 'genome' he is studying. While his script does run and return valid values, the amount of time it requires is quite staggering. He has asked you to assist him by taking his script and running it on the division's cluster resource. Specifically you must:

1. Login to the cluster's head node, `ccrc.mdc.musc.edu`, and make a directory specifically for this project.
  2. Download the corresponding items from the class website using the UNIX command `wget`:
    - `genome.txt`
    - `mineAminos.R`
    - `mineAminos.batch.R`
    - `create.batchfile.R`
  3. source `mineAminos.R` from R and save the tab delimited results and time required for future comparisons.
  4. use `create.batchfile.R` to run `mineAminos.batch.R` with 3, 6, and 9 'slaves'. For each run, build a 'program' (in a language of your choice) that will compare the percent differences of each amino count and the time required. Time required should be defined as the max processing time of each run.
- Bonus: Create an Rmpi version of the `mineAminos` script. Run it and compare the percent differences of each amino count and the time required.
- Bonus: Explain any differences in amino counts that you discovered.