Schedule

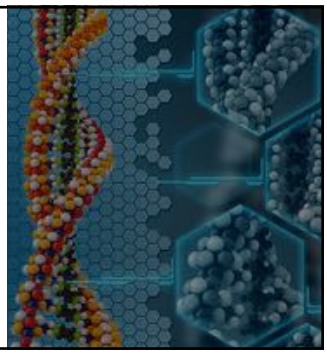
Tuesday, September 5th
- Connecting to servers
- Genomics datasets

Thursday, September 7th
- Substitutions and regular expressions
- Introduction to shell scripting

Tuesday, September 12th
- Processing data with awk and sed

Thursday, September 14th
- Conclusions and future directions
- Comprehensive final exam

CSU NSCI 580A4



Substitutions and regular expressions

Regular Expressions: character sequences that represent patterns.
- For example, the character sequence \n is the regular expression for a new line (see dokuwiki page and unix cheet sheet for a list of common regular expressions).

Substitutions: changing one pattern to another in a file, filename, etc.
- For example, substituting new lines (\n) for tabs (\t).

Substitutions and regular expressions

Common Linux tools that use regular expressions:

grep....search for patterns within a file and return lines containing the pattern.
usage.... $ grep "pattern" file

tr....deletes or replaces one set of characters in a file with another set of characters.
usage.... $ tr 'ACTG' 'TGAC' file

sed....for more complex substitutions than tr.
usage.... $ sed 's/"old_pattern"/"new_pattern"/g' file

awk....allows for more complex text editing and processing.
usage.... $ awk 'pattern { action }' file

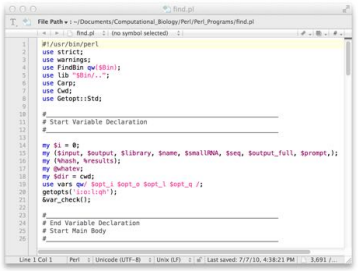
Introduction to scripting

Script: A series of commands or instructions to automate a task. The commands are written in a text file that is then executed by a program without being first compiled (converted into the binary machine code).

Scripting language: A computer programming language that supports scripts. The scripts are typically interpreted by the program and do not have to be compiled.

Writing scripts in text editors

Colorado State University



Common GUI text editors

Colorado State University

Mac Users: TextWrangler

PC Users: Notepad++

Linux users: gedit

Bioinformatics programming languages

Programming languages commonly used in bioinformatics:

- Python (Biopython)
- Perl (BioPerl)
- R (Bioconductor)
- C
- C++
- Java
- Ruby
- MATLAB (Software with GUI)