

Application of Bioinformatics in Genetics Research

Instructors:

Dr. Matt Gitzendannert

Dr. Raad Gharaibeh

Dr. Lei Zhou (Course director)

Course web page: <http://zhoulab.net/GMS6014/home.html> for
classroom practices, homework, etc.

Application of Bioinformatics in Genetic Research

Time and location:

MWF : 12:00-1:00

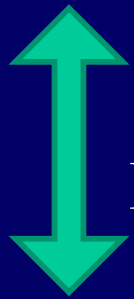
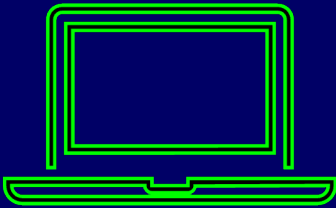
CGRC-291

Evaluation

- **50% classroom participation**
 - **Discussion.**
 - **Be ready to share your screen.**
- **50% homework**

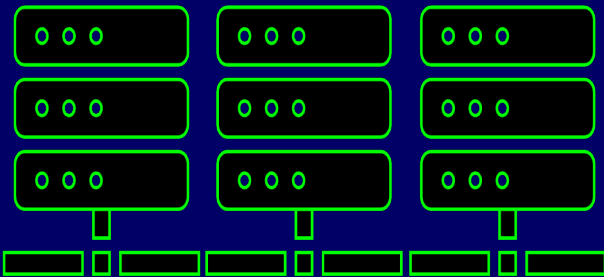
Required facility

- **Your own laptop**
 - **Browser(s)**
 - **text editor**
 - **Some programs**



FTP programs such as **FileZilla**

- **HiPerGator**
 - **All Linux/Unix programs**
 - **Large data set process.**



Practice

Practice: Download and install a text editor

Rule of thumb for doing your own bioinformatics:

- ❖ Make a folder for each program / project.
- ❖ **Do NOT** have space in folder and file name, consider using “_” to separate words.

History of bioinformatics – sequence analysis

- **Sequence comparison**
 - **Similarity search**
 - **Phylogenetic analysis**
- **Structure predication**
- **Gene prediction**
- **Genomics, omics, and systems biology**

Bioinformatics in the post genome era

The opportunity provided by genome sequence and genomic / proteomic technology is matched by the challenge to bioinformatics / computational biology

- **Information Representation.**
 - many new types of data, such as *Function, Location, Interaction, Regulatory pathway, Expression profile, etc. needs to be recorded*
- **Data Management**
 - Infrastructure for inputting, managing, access and retrieval of relevant information in a “sea of databases”. Cloud computing.
- **Systematics**

Bioinformatics in the post genome era

- Whole genome sequencing - SNP and whole genome wide association studies.
- Genomic/proteomic expression profiling (RNA and protein levels).
- Epigenomics, Comparative genomics, ...
- Regulatory pathway simulation - systems biology.

\$1,000 genome and ... \$500,000 analysis ?

Overwhelmed by data?

Objectives of GMS6014

- **Basic skills for retrieving and storing data, using web-based bioinformatics tools.**
- **Ability to install and run standalone local applications.**
- **Understanding the basis of bioinformatics applications using sequence similarity search as the example.**
- **An introduction to HTS analysis & HiPerGator**

Sequence Representation - nucleotide

N G R C W T G Y C Y

A G A C A T G C C C

C G T T T

G

T

For complete list of IUB/IUPAC nucleic acid codes, see
<http://www.ncbi.nlm.nih.gov/blast/fasta.shtml>

Sequence Representation - amino acids

Q:

What's the common property of these amino acids ?

1. D, E

2. I, L, V, M, F

Sequence Representation - amino acids

Example:

W D L L A Q I L C Y A L R I Y

W R F L A T V V L E T L R Q Y

W K F L A I T M C K V L K Q F

R C L L C N K L Y Y L W D L L A Q I L C Y A L R I Y

L N R L L A E L Y E V W R F L A T V V L E T L R Q Y

L R L L Q Q Q Q M V L W K F L A I T M C K V L K Q F

R C L L C N K L Y Y L L R K V

L N R L L A E L Y E V L C H I

L R L L Q Q Q Q M V L Q R Q Y

Coloring based on aa property.