



Department of Computer Science  
College of Engineering  
University of the Philippines  
Diliman, Quezon City



COURSE NO.	: CS 175
COURSE TITLE	: Topics in Computational Science - Introduction to Bioinformatics for Genomics and Transcriptomics
COURSE CREDIT	: 3 units
SEMESTER	: First Semester, AY 2017-2018
INSTRUCTOR	: Jan Michael C. Yap
EMAIL ADDRESS	: janmichaelyap@gmail.com

### COURSE DESCRIPTION

This course serves as an introductory course to the design and implementation of bioinformatics solutions to genomics and transcriptomics level analyses and tasks. In particular, the course will be focusing on methods for sequence and annotation analysis, phylogenetics, DNA barcoding, and gene expression analyses.

### CLASS SCHEDULE

(MRU) Mondays, 8:00 - 11:00 AM

### COURSE OUTLINE

1. Introduction
  - (a) Overview
  - (b) Programming Foundations
    - i. Introduction to Linux Command Line Interface and Scripting
    - ii. Quick Introduction to Programming Using R
  - (c) Biological Foundations
    - i. DNA, RNA, and the Genome
    - ii. Central Dogma of Molecular Biology
    - iii. Molecular Evolution
    - iv. Sequencing and Measuring Gene Expression
  - (d) Pairwise Alignment
  - (e) Multiple Sequence Alignment
  - (f) Sequence Database Searching: BLAST
  - (g) DNA Barcoding
  - (h) Phylogenetic Tree Reconstruction
    - i. Distance-Based Techniques
    - ii. Character-Based Techniques
  - (i) Sequence Assembly
  - (j) Annotations
  - (k) Gene Expression Analysis
    - i. Differential Gene Expression Analysis
    - ii. Coexpression Analysis
2. Analyses and Methods
  - (a) Sequence Alignment
  - (b) Pipelining

## GRADING SYSTEM

Requirements:

Problem Sets	60%
Mini-Project	40%

$92 \leq FG \leq 100$	1.00
$88 \leq FG < 92$	1.25
$84 \leq FG < 88$	1.50
$80 \leq FG < 84$	1.75
$76 \leq FG < 80$	2.00
$72 \leq FG < 76$	2.25
$68 \leq FG < 72$	2.50
$64 \leq FG < 68$	2.75
$59 < FG < 64$	3.00
$FG \leq 59$	5.00

## CLASS POLICIES

**CONSULTATIONS.** Schedule of the instructor's free times for the semester would be **Tuesdays to Fridays, 8:30am-10:30am, then 1:30pm-4:30pm**. Due to other commitments apart from teaching and to ensure proper queueing of multiple requests for consultations, students who wish to schedule a consultation must first send an e-mail to the e-mail address mentioned in the first part of the syllabus setting an appointment **at least 24 hours before** the requested date and time of consultation. Note however that a rescheduling of the consultation may be done subject to the availability of instructor.

**WEBSITE.** The link to the class website is <http://bit.ly/CS175BioinfoNiJan>. Announcements and certain materials will be posted there. Pertinent announcements will be posted **Sundays, 9:00pm at latest**. If no announcements are posted by that time, then classes are assumed to proceed as usual.

**DELIVERABLES.** Submission of deliverables should be done on or before the designated deadlines (to be announced in future sessions). Late submissions would incur a 20% deduction per day late, and can incur a maximum of 100% deduction. Note though that even if a deliverable will no longer merit any points, it is considered a requirement, and non-submission of the deliverable on or before the deadline of submission of grades would mean a grade of INC assuming FG is passing.

**INTELLECTUAL DISHONESTY.** Intellectual dishonesty of any form will be subjected to proper disciplinary action as per Section V.2 of the Code of Student Conduct (download link posted in our website). If found guilty, results of the requirement/s in question will be nullified, final grade will be recomputed, and a corresponding change of grade will be filed if warranted.

**ATTENDANCE.** A student whose absence exceeds three (3) days will be forcibly dropped from the course and may be given a grade of 5.0 if the excuse for the majority of the absences are not

excused/valid. Absence is defined as not being in class at least 30 minutes after the official start time of the class.

**DROPPING THE COURSE.** A student is considered dropped upon completion of the dropping procedure (see <https://crs.upd.edu.ph/downloads/OnlineDroppingFlowchart.pdf>). A student may officially drop the course **on or before 14 November 2017**.

## REFERENCES AND RESOURCES

- Jones NC, Pevzner PA (2004). **An Introduction to Bioinformatics Algorithms**. MIT Press, Cambridge, USA.
- Sequence Alignment
  - Al-Karadaghi S (n.d.). **Sequence alignment and substitution matches**. <http://www.proteinstructures.com/Sequence/Sequence/sequence-alignment.html>
  - Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990). **Basic Local Alignment Search Tool**. *J Mol Biol* 215(3): 403-410.
  - Autenrieth F, Isralewitz B, Luthey-Schulten Z, Sethi A, Pogorelov T (2005). **Bioinformatics and Sequence Alignment**. <http://www.ks.uiuc.edu/Training/Tutorials/science/bioinformatics-tutorial/bioinformatics.pdf>
  - Mullan LJ (2002). **Multiple sequence alignment the gateway to further analysis**. *Briefings in Bioinformatics* 3(3): 303-305.
- BLAST: Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990). **Basic Local Alignment Search Tool**. *J Mol Biol* 215(3): 403-410.
- DNA Barcoding
  - Cold Spring Harbor Laboratory (n.d.). **DNA Barcoding 101**. <http://www.dnabarcoding101.org/>
  - Digital World Biology (n.d.). **DNA Barcoding**. <http://digitalworldbiology.com/dwb/tutorial/dna-barcoding>
- Phylogenetic Tree Reconstruction
  - Baldauf SL (2003). **Phylogeny for the faint of heart: a tutorial**. *TRENDS in Genetics* 19(6): 345-351.
  - Husmeier D (n.d.). **A Brief Tutorial in Phylogenetics** (slides). [http://www.bioss.ac.uk/~dirk/talks/tutorial\\_phylogenetics.pdf](http://www.bioss.ac.uk/~dirk/talks/tutorial_phylogenetics.pdf)
- Sequence Assembly
  - Miller J, Koren S, Sutton G (2010). **Assembly algorithms for next-generation sequencing data**. *Genomics*, 95(6): 315-327.
  - Salberg SL, et al. (2012). **GAGE: A critical evaluation of genome assemblies and assembly algorithms**. *Genome Res.* 22(3): 557-567.
  - Schatz M (2012). **Whole Genome Assembly and Alignment** (slides). <http://schatzlab.cshl.edu/teaching/2012/CSHL.Sequencing/Whole%20Genome%20Assembly%20and%20Alignment.pdf>

- Zimin A, et al. (2014). **Sequencing and Assembly of the 22-Gb Loblolly Pine Genome**. *Genetics* 196: 875-890.
- Annotation
  - Gene Ontology Consortium. <http://geneontology.org/>
  - The Sequence Ontology Project. <http://www.sequenceontology.org/>
  - Rhee SY, Wood V, Dolinski K, Draghici S (2008). **Use and misuse of the gene ontology annotations**. *Nature Reviews Genetics* 9 (7): 50915.
  - Schuurman N, Leszczynski A (2008). **Ontologies for Bioinformatics**. *Bioinformatics and Biology Insights* 2: 187-200.
- Gene Expression Analysis
  - von Heydebreck A (2003) **Statistical tests for differential gene expression (Slides)**. [http://lectures.molgen.mpg.de/Microarray\\_WS0304/anja\\_09\\_12\\_03.pdf](http://lectures.molgen.mpg.de/Microarray_WS0304/anja_09_12_03.pdf)
  - Langfelder P, Horvath S (2008) **WGCNA: an R package for weighted correlation network analysis**. *BMC Bioinformatics* 2008, 9:559.
- Pipelining
  - **Question: How To Organize A Pipeline of Small Scripts Together?**  
<https://www.biostars.org/p/79/>
  - **Question: Which Bioinformatic Friendly Pipeline Building Framework?**  
<https://www.biostars.org/p/91301/>
  - Cingolani P, Sladek R, Blanchette M (2014). **BigDataScript: a scripting language for data pipelines**. *Bioinformatics* 2014. doi: 10.1093/bioinformatics/btu595.
  - Napolitano F, Mariani-Costantini R, and Tagliaferri R (2013). **Bioinformatic Pipelines in Python with Leaf**. *BMC Bioinformatics* 14:201.