

Fall 2021

BiS232: Bio-data structures
Mon/Wed, 10:30 am - 12:00 pm

Instructor: Prof. Young-suk Lee

Email: youngl@kaist.ac.kr

Office hour: MW 3:30 pm - 4:30 pm (Rm#1113, E16)

TAs: TBD

Course Objectives

At the end of this course, students will be able to:

- Learn data structures and algorithms in bioengineering
- Apply algorithmic thinking to specific biological questions
- Handle and interpret pressing questions in big data biology

Course Description

Programming is now one of the basic tools used in most, if not all, areas of biomedical engineering. There are many excellent online and offline resources that introduce coding at the beginner level, but not many offer a foundation of programming with the mathematical rigor and excitement in the context of biomedical applications. In light of this need, the Department of Bio and Brain Engineering is offering an introductory course in data structures and algorithms that is specialized for those interested in computational biology and bioengineering. In this course, we will start from a specific biological question, reduce it into a concrete computational problem, and then design efficient data structures and algorithms that provide testable solutions. Example topics covered in this course include: algorithmic thinking, basic computational theory, some probability and statistics, dynamic programming, sequence alignment, evolutionary tree reconstruction, and parsimony problems.

Recommended Prerequisites

* Please contact the instructor for questions regarding prerequisites

- CS101 Introduction to Programming
- BS120 General Biology
- BS122 Diversity of Life

Textbook

- "Bioinformatic Algorithms" by Phillip Compeau and Pavel Pevzner

Grading

Programming assignments (30%), problem sets (10%), midterm exam (25%), final project (25%), and participation (10%).

Tentative Schedule

Week 1: Why “Bio-” in data structures?

- Course description
- Genome assembly

Week 2: Pseudocode and nontrivial algorithms

- Special tutorial on ssh, unix, vscode, and remote servers

Week 3: Sets, Propositional Logic, Predicates, and Quantifiers

- Problem set #1: Introduction to discrete structures
- Special tutorial on python programming #1 (random variables, control flow, function)

Week 4: Theorems and proofs

- **No class Monday 9/20 and Wednesday 9/22 (Chuseok Holidays)**
- Special tutorial on python programming #2 (list, objects, modules, file handling)

Week 5: What is a “clearly stated” computational problem?

- Introduction to coding style, academic integrity, and plagiarism
- Programming #1: Clump Finding Problem

Week 6: How to quantify “surprising”?

- **No class Monday 10/4 (National Foundation Day)**

Week 7: Why it’s important to be “approximately” correct in biology

- **No class Monday 10/11 (Hangeul Day)**

Week 8: Midterm

Week 9: Why need “computation” in molecular biology?

- **No class Wednesday 10/27 but will reschedule**

Week 10: Biological question in disguise

- Exercise #1 Space efficient change problem

Week 11: The role of (discrete) math in computational biology

- Programming #2: Multiple sequence alignment
- **Special Seminar in Computational Biology Wednesday 11/10**

Week 12: Where is computer science?

Week 13: Which animal gave us SARS-CoV-2?

- Programming #3: Space efficient sequence alignment

Week 14: More math and more algorithms

- Exercise #2 SARS-CoV-2 alpha variant

Week 15: Next generation computational biology