Spring 2023
BIOS 477/877
Bioinformatics and Molecular Evolution
Instructor: Etsuko Moriyama
(School of Biological Sciences)

TODAY’S TOPICS
➤ Course mechanics
➤ What is Bioinformatics?

Course Web page
http://bioinfolab.unl.edu/cmblab/Courses/BIOS477/current/index.html
➤ Course description
➤ Course schedule
  ➤ Check this often! Will be regularly updated.
  ➤ Readings (chapters from the optional textbook, etc)
  ➤ Lecture handouts (color PDF version)
  ➤ Homework assignment due dates
➤ Useful links (more links will be added)
➤ For non-biology students
  ➤ List of molecular biology/genetics keywords (study them!)
  ➤ Online review materials provided.
  ➤ No review session for molecular biology during the course

Course Home Page on Canvas
https://canvas.unl.edu
➤ Check Email address for your Canvas account!
➤ Announcements & Email for communication
➤ Syllabus
  ➤ Link to the course Web page
  ➤ Notes on grading, assignment submission, etc.
➤ Course Content
  ➤ Lecture slides (full size PDF file)
  ➤ Related articles (PDF)
  ➤ “Introduction to Bioinformatics”
    (Chapter 1 of Essential Bioinformatics by J. Xiong) [Lecture 1]
➤ Grades

Course Home Page on Canvas (continued)
https://canvas.unl.edu
➤ Assignments
  ➤ Download/submit homework assignments
➤ Course Survey (non-graded)
  ➤ Download from Assignments
  ➤ Submit by Jan 27 (Fri)
  ➤ This is to ensure that you have no problem using Assignments page. If you have any problem, we need to fix it ASAP!!

Textbook (optional)
Understanding Bioinformatics
Garland Science
ISBN 0815340249
Book website:
- eBook Rental is also available
Required in this course

➢ You should be familiar with basic molecular biology & genetics terminology
   ➢ Use the keyword list and online materials to prepare yourself
➢ You should be able to use/access internet, Web browser, and familiar with general computer tasks
   ➢ To work on homework assignments
➢ If you do not have or access to a computer … let me know!

Not required in this course

➢ Programming skill
➢ Unix/Linux knowledge

But they are very useful if you want to do lots of bioinformatics/genomics analyses in the future!!

Unix/Linux/Programming online tutorials available on the Course website (see Links page)

No computer lab

➢ It’s not necessary to bring your laptop to the class
   ➢ It may be useful to have it if you want to follow during demonstrations in the class.
➢ Homework assignments require you to do some web-based bioinformatics analysis

Assignments

➢ No in-class exam
➢ Multiple assignments (about every week)
   • manual computation of core methods
   • small sequence analysis
   • article review
     ➢ for core methods
     ➢ for bioinformatics research
➢ You need to commit your time
➢ Read Notes for assignment submission in the Syllabus page

REMEmBER THIS!

➢ Some assignments take time. Some require manual computation.
➢ Some assignments require you to use software on the Web.
   ➢ Web servers could become slow. They may crash.
   ➢ Canvas does not always cooperate. Submitting assignments may take time, especially just before deadlines!
   We experienced all of these in the past!
➢ Learning new programs requires time.
   ➢ Read manuals and on-line helps.
   Start working on your assignment as soon as possible.
   DO NOT wait until just before the deadline!

Grading

➢ Read Notes for grading in the Syllabus page
➢ If you are taking this course as BIOS477, your worst assignment score is excluded from calculating the final score (except for the final assignment)
   ➢ You can safely skip one assignment
➢ Late submission of an assignment will cost 5% (or more) deduction per day (indicated in each assignment)
➢ Graduate Student Only assignments:
   ➢ indicated as G1 and G2
   ➢ undergraduates can use them for bonus points
What you will learn from this course

➢ Core algorithms in bioinformatics
➢ How they work
   (If you want, you should be able to write your own program)
➢ How you can interpret the results for biological significance

GOAL: You should be able to learn new bioinformatics methods by yourself in the future

Bioinformatics fields evolve fast → requires continuous self-learning

Other Bioinformatics Courses at UNL

➢ BIOS 427/827 Practical Bioinformatics Laboratory
   ➔ Spring, Fri. Computer lab for molecular biology & bioinformatics tools
➢ BIOS 426/826 Computational Systems Biology
   ➔ Fall, online; Next-gen sequence analysis, protein interaction/regulatory network
➢ BIOS 337 Basic Application of Bioinformatics
   ➔ Fall; Basic hand-on bioinformatics
➢ AGRO 820: Bioinformatics Applications in Agriculture
➢ BIOC 439/839 Dynamics of Biochemical and Biological Networks
➢ BSEN 657: Advanced Mathematical Modeling in Biological Engineering
➢ FDS 406: Bioinformatics Skills for Food Genomics (including microbiome analysis)
➢ STAT 441: Statistical Methods for High Throughput Biological Data (next-gen data analysis)
➢ ECEN 450/850 Computer Science (Spring)
➢ BIOS 471/871 Introduction to Bioinformatics (Spring)
➢ BIOS 155/855: Computer Science (Fall)
➢ BIOS 471/871 Introduction to Bioinformatics (Spring)
➢ BIOS 155/855: Computer Science (Fall)

List of more courses available on the “Links” page
Contact each instructor for more information

Useful tricks (more in the Syllabus page)

➢ Save your sequences in the plain TEXT format
   ➔ The default Word file (.doc or .docx) is not in the text format!
   ➔ Use “Save as” and choose the text format.
   ➔ Use Text editors (BBEdit, EditPlus, etc).
➢ Read the manuals and follow the input format!
➢ Control your computer. Don’t let your computer control you! Don’t just double click on files.
➢ DO NOT blindly believe what you find on internet (e.g., Google, Wikipedia)! Lots of incorrect information.
   Always double-check with the original information source (journal articles).

What is Bioinformatics?

➢ Too many definitions!!
   ~155,00,000 hits by google search using “bioinformatics” and “definition” as keywords
➢ Luscombe et al. (2001) (Xiong, Ch. 1, page 4)
   “An interdisciplinary field involving biology, computer science, mathematics, and statistics to analyze biological sequence data, genome content, and arrangement, and to predict the function and structure of macromolecules”
➢ National Cancer Institute (NCI’s Dictionary of Cancer Terms)
   “A field of science that uses computers, databases, math, and statistics to collect, store, organize, and analyze large amounts of biological, medical, and health information. Information may come from many sources, including genetic and molecular research studies, patient statistics, tissue specimens, clinical trials, and scientific journals. Also called computational biology.”

Any questions?
Who coined the term “Bioinformatics”?

> Paulien Hogeweg (in early 1970s)

A Dutch theoretical biologist, who started using the term “Bioinformatics” to refer to “the study of informatic process in biotic systems”.


Dr. Hogeweg’s lab website:
https://www-bnf.bio.uu.nl/ph/

---

What is the information in Bioinformatics?

**DNA -------> RNA**

Transcription (gene expression)

**Gene prediction** (expression pattern)

**DNA -------> RNA**

Transcription (gene expression)

**DNA --> RNA --> Protein**

Translation

**Similarity search**

Domain search

Transmembrane prediction

Protein classification (Functional annotation)

**DNA --> RNA --> Protein --> Function**

---

**Amino acid sequence** (primary structure)

**Secondary structure prediction** (α-helix, β-sheet)

**Tertiary structure prediction**

**DNA**

**RNA**

**Protein**

**Genomic (DNA) sequences**

**Need to be determined and assembled**
How do we analyze biological sequences?

**Single sequence analysis**

- Gene prediction (DNA)
- Gene expression (DNA)
- Primary structure prediction (from DNA to protein)
- Secondary structure prediction (protein)

**Multiple sequence analysis**

- Pairwise alignment
- Multiple alignment
- Distance estimation
- Phylogenetic reconstruction
- Conserved domain
- Protein function

More Topics in Bioinformatics

- Large-scale fragment/contig assembly
- Next-gen sequencing data analysis
- RNA-seq (gene expression) analysis
- RNA secondary structure prediction
- RNA tertiary structure prediction (ab initio)
- Tertiary structure prediction (ab initio)

These topics are not covered in this course...

Limitations of Bioinformatics

(Read Xiong Ch1 page 7)

- Bioinformatics and experimental biology are complementary.
- Raw data are produced by experimental science.
  - The quality of bioinformatics predictions depends on the quality of data.
- Bioinformatics predictions are not formal proofs.
  - They need to be tested experimentally.
- Bioinformatics provides interpretation of experimental data.
  - Facilitates generating a hypothesis for further experimental research.

History of Molecular Biology and Genomics

Dec/2022: 241,015,745 sequence entries in Genbank
Why Molecular Evolution?

Molecules as Documents of Evolutionary History
Zuckerkandl & Pauling (1965)

Molecular evolution mechanisms underlying evolution at the molecular levels
- Use to reconstruct how molecules evolved in the past
- Can be used to predict how sequences evolve in the future

Comparative analysis Bioinformatics

Why Molecular Evolution?

Molecular Evolution:
- Mechanisms of sequence evolution
  - Molecular Evolutionary Methods:
    - Sequence distance calculation
    - Phylogenetic reconstruction
    - Positive selection detection
    - Population genetics analysis
  - Computer Science, Informatics, Mathematics, Statistics:
    - Information theory
    - Database
    - Algorithms
    - Mathematical/statistical modeling
  - Bioinformatics Methods:
    - Pairwise alignment
    - Multiple alignment
    - Similarity search
    - Domain/protein family classification
    - Protein structure prediction
    - Gene prediction

Sequence relationships
- Protein functions
- Genome annotation
- Gene/protein network analysis
- Protein interaction
- Metabolic pathways