

BIOL 367-01 / CMSI 367-01 / HNRS 398-05

BIOLOGICAL DATABASES

<http://myweb.lmu.edu/dondi/fall2010/cmsi367>

<https://www.cs.lmu.edu/biodb/fall2010>

Fall 2010

Doolan 219

T 4:30–7:00pm

3 semester hours

Kam D. Dahlquist, PhD

kdahlquist@lmu.edu

Office Hours: MW 2–4pm, or by appt.

Seaver 218; (310) 338-7697

John David N. Dionisio, PhD

dondi@lmu.edu, AIM: dondi2LMU

TR 9am–12nn, or by appt.

Doolan 106; (310) 338-5782

Course Objectives and Learning Outcomes

This course is built upon L. Dee Fink's *taxonomy of significant learning*, as applied to biological databases. Long after the course concludes, our hope is that:

- You understand how biological information is encoded in the genome and can apply this knowledge to a variety of biological tasks and problems
- You understand the core concepts, structure, and functions of a database, ranging from individual files to a full relational database management system, and can perform useful tasks with such data
- You show discipline and proficiency in day-to-day science and engineering *best practices*, such as maintaining journals and notebooks, managing your files and code, and critically evaluating scientific and technical information
- You recognize and care about how the biological and technological issues presented in this course relate to and affect society, our daily lives, and ourselves
- You have some skills and tools for “leaving your comfort zone,” flourishing outside of it, and learning more about biology and computer science on your own
- You learn how to communicate and work effectively with colleagues from different disciplines

Prerequisites/Recommended Background

To take this course, you must have upper division standing in either the College of Science and Engineering or the Honors Program. Otherwise, there are no strict prerequisites; backgrounds in biology and computer science, as well as prior experience with database or information management applications, may be helpful but not necessary.

Classroom and Laboratory Environment

We are all responsible for maintaining a classroom and laboratory environment that is safe and conducive to learning. As such, we will observe the following:

1. You are responsible for your own learning and for being a good class citizen.
2. We will act with honesty and integrity at all times.
3. We will always treat individuals with respect.
4. Class will start promptly on time.
5. You are expected to come to class having done the assigned reading and preparatory work.
6. You are expected to bring the required materials to each class session.
7. Cell phones, pagers, and other communication or music devices will be turned off.

All students are governed by LMU Community Standards Publication, found here:

http://www.lmu.edu/studentlife/Judicial_Affairs/Standards_Publication.htm

Materials and Texts

- David P. Clark and Lonnie Dee Russell, *Molecular Biology Made Simple and Fun* (a.k.a. “the monkey book”). Cache River Press, 1997.
- Recommended but not required: Jeffrey Ullman and Jennifer Widom, *A First Course in Database Systems*, Third Edition. Prentice Hall, 2008.
- Assorted handouts, articles, and sample code to be distributed throughout the semester.
- Keck Lab account

Additional information is also available on the web; do not hesitate to look for further sources of information regarding the concepts, techniques, tools, and paradigms that we will discuss.

Course Work and Grading

Graded coursework consists of quizzes (10%), presentations (10%), an ongoing, online course notebook/journal (40%), and 1 biological database project (40%). Letter grades are determined as follows: $\geq 90\%$ gets an A– or better; $\geq 80\%$ gets a B– or better; $\geq 70\%$ gets a C– or better. The instructors may curve grades upward based on qualitative considerations such as degree of difficulty, effort, class participation, time constraints, and overall attitude throughout the course. Grades are never curved downward. Current grades will be made available at MyLMU Connect (the system formerly known as Blackboard).

Quizzes

Two quizzes will be administered to assess students' progress towards meeting the course learning outcomes. Dates for the quizzes will be preannounced; see the *Attendance & Make-Up Work* policy below.

Presentations

Periodically, students will give presentations in class, either individually or as groups. They will be graded based on content, organization, clarity of visuals, and presentation style.

Course Notebook/Journal/Homework

One of the most important skills you can develop as a scientist is keeping an excellent laboratory notebook. The computer science equivalent of the biology paper-based lab notebook is documentation, both in the form of comments in the code, commit logs, and user or developer manuals. For this course you will practice the documentation skills needed by both users and developers of bioinformatics tools by keeping an electronic course notebook or journal. The technology we will use is a MediaWiki web site, open only to the students and instructors of this course, that we will create and edit during the semester. You will create an individual user page and make weekly entries that the instructors will read and grade. You will use the wiki to complete some of the homework assignments as well. The following guidelines apply:

- Your weekly wiki entry is due every midnight on Monday. Note that wiki entry timestamps use Coordinated Universal Time (UTC).

- Each submission will be worth a preannounced number of points, commensurate with the tasks for that week; failure to submit on time will result in a grade of zero.
- The instructors will read and comment on how to improve your journal entries/answers/homework. In some cases, the work may be subjected to peer review.
- Depending on the type of assignment for that week, you may be given an opportunity to improve your work and earn additional points based on the aforementioned feedback.
- 10% of the course notebook/journal grade will be based on the overall quality of your body of work, evaluated at the end of the semester. You are encouraged to make improvements to previous journal entries as the semester progresses.

Generally, your journal entries will consist of:

- Exercises related to the current course material
- Workflow and other documentation for hands-on exercises and projects
- Reflection on your learning — you will answer questions such as:
 - ▶ What have I learned?
 - ▶ What isn't clear to me (yet)?

Journal entries will be graded based on completeness, accuracy (as appropriate to the assignment), organization, and clarity.

Biological Database Project

In the biological database project, you will apply what you have learned in the first two units of the course to perform an original bioinformatics research project. Students will work in interdisciplinary teams to create a Gene Database for the GenMAPP program for a new species not currently supported by the program. You will then use GenMAPP to analyze publicly available microarray data for that species. This project is worth 40% of your final grade for the course.

Project deliverables will include a group project report (written and oral), related files, data sets, and documentation, and individual statements of work, project assessments, and self-reflection. The overall project grade will be based on completeness, accuracy, organization, writing, and polish, as appropriate to each deliverable.

Extra Credit: Seminars

Students may accumulate up to 2.5% of their final grade in extra credit by attending Biology Department seminars and completing the seminar sheets. Each seminar attended is worth 0.5% with up to 5 seminars (2.5%) total. Students arriving late to the seminar will not be granted credit — so don't be late! Certain non-Biology Department seminars may be approved in advance for extra credit at the instructors' discretion. To receive credit for these non-Biology Department seminars, you must turn in a one-page summary of the seminar within one week of the date of the seminar or they will not count as extra credit.

Attendance & Make-Up Work

Attendance at all class sessions is expected, but not absolutely required. Each week will include group activities, in-class assignments, and presentations that will be difficult to make up. Thus, if you must miss one or more class sessions, it is your responsibility to keep up with the course work. The instructors should be notified as soon as possible, electronically or by phone, of the reasons for all absences. We will make arrangements to discuss make-up work. At the discretion of the instructors, excessive absences may result in a grade of incomplete (I).

Note that the last day to add or drop a class without a grade of W is September 3. The withdrawal or credit/no-credit status deadline is November 5.

University Policy on Academic Honesty

Loyola Marymount University expects high standards of honesty and integrity from all members of its community. All students are expected to follow the LMU honor code. The *LMU Undergraduate Bulletin 2010-2011*, p. 59, lists examples of academic dishonesty. An online version is available at:

http://www.lmu.edu/about/services/registrar/Bulletin/Academic_Degree_Requirements_and_Policies.htm

You are required to sign the Honor Code Agreement for this course.

Americans with Disabilities Act

Students with special needs as addressed by the Americans with Disabilities Act who need reasonable modifications, special assistance, or accommodations in this course should promptly direct their request to the Disability Support Services

(DSS) Office. Any student who currently has a documented disability (physical, learning, or psychological) needing academic accommodations should contact DSS (Daum Hall, Room 224, x84535) as early in the semester as possible. All discussions will remain confidential. Please visit <http://www.lmu.edu/dss> for additional information.

Course Assessment

We ask that students in this class participate in the Research on the Integrated Science Curriculum (RISC) Survey administered by David Lopatto at Grinnell College, which will allow the instructors to evaluate the effectiveness of this course. Students are asked to complete the student pre-course and post-course surveys found at:

<http://www.grinnell.edu/academic/psychology/faculty/dl/risc>

Course Topics and Important Dates

University deadlines and holidays are shown in italics; full schedule details appear on the wiki.

| | |
|--------------------------|--|
| September | Central dogma, the genetic code, gene and genome structure, public biological databases; command line interaction; text file formats, conversion, searching, regular expressions |
| <i>September 3</i> | <i>Last day to add or drop a class without a grade of W</i> |
| October | DNA microarray data analysis, biological pathways and ontologies; relational databases, XML |
| November/December | Project workshop: progress reports, group sessions |
| <i>November 5</i> | <i>Withdraw/credit/no-credit deadline</i> |
| <i>November 24–26</i> | <i>Thanksgiving; no class</i> |
| <i>December 7</i> | <i>Final project presentations</i> |
| <i>December 14</i> | <i>Final project deliverables due</i> |

Version Control

If necessary, this syllabus and its contents are subject to revision; students are responsible for any changes or modifications announced in class. The most current version of this information resides on the course wiki:

<https://www.cs.lmu.edu/biodb/fall2010>