# Introductory Bioinformatics - BISC 499
# Syllabus - 2011 Fall Semester

## 1. Basic Information

| | |
|---|---|
| *Course:* | Introductory Bioinformatics, BISC 499, 2 credits |
| *Textbook*: | "Python for Dummies" |
| *Place and time:* | Monday: 9:00 am - 10:50 am<br>Ray R. Irani Building, room 301 |
| *Faculty:* | Dr. Matthew D. Dean<br>Assistant Professor, Molecular and Computational Biology |
| *Office:* | 304A Ray I. Irani Building. |
| *Telephone:* | 213-740-5513 |
| *Email:* | matthew.dean@usc.edu |
| *Office Hours:* | Thursday: 8:00 am - 9:00 am or by appointment |
| *Prerequisites:* | none |
| *Class web page:* | https://blackboard.usc.edu (follow link to BISC 499 web space) |

## 2. Classroom policy

Students must bring laptops to class. Whether you use Windows, Mac, or other (Linux, Unix, etc.) does not matter, but laptops are critical because lectures include hands-on programming. If you do not have access to a laptop, one will be provided.

Any other electronic communication devices (phones, blackberries, and similar) must be turned off, and no instant messenger/chat type programs are allowed in class.

## 3. Course goals and learning objectives

The main goal of Introductory Bioinformatics is to teach students how to write computer programs to analyze biological data. The class is divided into two main parts. In the first part, we will learn how to use Python, an object-oriented computer language that is an ideal combination of power and simplicity. Our philosophy in this class is to learn Python in a hands-on way, through tutorials and weekly homeworks that challenge the student to break down problems into manageable units. In the second half of the course, students will apply their Python skills to address a bioinformatic question of their own construction. Students, especially graduate students, are encouraged to bring their own data sets to analyze and to ask a question that is specific to their thesis. Students without their own data will be given important bioinformatic questions by the instructor.

**In this class, bioinformatics refers to any computational approaches that are incorporated into the analysis of biological data. The ability to write code is a critical aspect of success, regardless of field of interest or type of data.**

The only pre-requisite for this course is scientific curiousity. Students are not expected to know

anything about bioinformatics. This class is not meant to teach advanced algorithmic design or statistics (such classes already exist in our department), though there are many themes that overlap with those fields. The emphasis in this course is on practical implementation, not on computational aesthetics.

Specifically, we will learn:

- How to manipulate large datasets using read, write, and comparative functions.
- How to create customized statistical tests using simulations. Real biological data often violate critical assumptions of standard statistical tests. The more sophisticated biologist knows how to account for complexities of the data through permutation and randomization.
- How to construct a pipeline of different programs that automates genomic analysis.
- Most importantly, this class will provide students the means to break down a scientific hypothesis into its fundamental elements, a necessary prerequisite to coding for the answers.

## 4. Course plan and weekly readings

To maximize the benefit of attending class, you must read the selected pages listed below before coming to class.

| Week | Topic | Textbook pages | thenewboston.com tutorials |
|---|---|---|---|
| 1 | Installing Python; Basic usage | Ch. 1, 2 | 1, 5 |
| 2 | Basic Elements & Syntax | Ch. 3, 4, 5 | 2-3, 6-7 |
| 3 | Strings | Ch. 6 | 17-18 |
| 4 | Lists and Tuples | Ch. 8 | 9-13, 16 |
| 5 | Dictionaries | Ch. 9 | 19 |
| 6 | Loops, comparisons | Ch. 10, 15 | 20-26 |
| 7 | Definitions & functions | Ch. 11 | 27-31 |
| 8 | Classes | Ch. 13, 14 | 8, 33-34, 41-43 |
| 9 | Midterm | | |
| 10 | Randomization & permutation | assignments | |
| 11 | Graphing & stats (R and matplotlib) | assignments | |
| 12 | Navigating public genomic databases | assignments | |
| 13 | Genomic analysis | assignments | |
| 14 | Final presentations | | |
| 15 | Final presentations | | |
| 16 | Final presentations | | |

Weeks 1-9: These initial weeks will be spent learning Python from the ground up, in a hands-on

way. After 9 weeks, students will be fluent in Python. We will then apply our newly gained knowledge to address a specific scientific question. Students (especially graduate students) are encouraged to bring their own data and their specific question to class for this purpose. Otherwise, genomics level problems will be assigned to them.

Weeks 10-14: During weeks 10-14 students will learn to navigate public databases and download genomic data that is valuable to their biological questions of interest. Students will work on their own bioinformatics projects during this time, and will have weekly meetings with the instructor to maintain progress and clear any programming hurdles that arise.

Weeks 14-15: The last three weeks of the course will be dedicated to student presentations, where students go from hypothesis, to data analysis, to conclusions using computational approaches.

## 6. Professor

Dr. Matthew D. Dean
matthew.dean@usc.edu
213-740-5513
304A Ray R. Irani Building
1050 Childs Way
University of Southern California
Los Angeles, CA 90089

Dr. Dean maintains an active research program focused on learning the ecological and evolutionary correlates of sexual selection. Bioinformatics represents an integral part of these endeavors.

*Unofficial Teaching Assistant:*
Brad Main
bmain@usc.edu
316 Ray R. Irani Building
1050 Childs Way
University of Southern California
Los Angeles, CA 90089

## 7. Required material

- Textbook: "Python for Dummies".
- Additional online tutorials at www.newboston.com
- laptop computer (if you do not have one, we can provide one for you)

*optional:*
- Learning Python, 4th edition. Mark Lutz.

## 8. Assessment

Grades are based on four scores: 1) midterm grade, which tests fluency with Python, 2) weekly homework assignments where students solve bioinformatic challenges by writing code, 3) the code written for their bioinformatics project, which must work and must be properly annotated, and 4) final presentations, a 20 minute talk including slides where the student describes their own bioinformatic project, the code that solves their particular problem, and draws valid biological conclusions:

| Assessment Procedure | Percent |
|---|---|
| Midterm | 25% |
| Weekly homeworks | 25% |
| Final project code | 25% |
| Final project presentation | 25% |

*8.1. Criteria for grading:* The midterm will be a closed book test that consists of both written questions and answers as well as computer programming problems. Bioinformatics code will be graded according to proper annotation of code and ability to solve the problem of interest. The final presentation will be graded according to clarity of scientific hypothesis, appropriateness of data to address that hypothesis, ability of the student to effectively communicate their bioinformatic strategy, and on the substance of their conclusions.

Students who are not able to meet deadlines due to medical or other emergency must notify the instructor or unofficial TA immediately.

*8.2. Course grade:* The course is not curved. Letter grades will follow a straight scale: 90% and above leading to A, 80-90% leading to B, etc. Pluses and minuses are assigned by dividing each range in corresponding halves (A, A-) or thirds (B+, B, B-, C+, ...).

**9. Policy against Cheating**

All USC students are responsible for reading and following the Student Conduct Code, which appears in the Scampus and at http://www.usc.edu/dept/publications/SCAMPUS/goverance. This policy does not apply to discussion or exchange of ideas. On the contrary, such interactions represent an important way to clear programming hurdles.

**10. Disability Policy Statement:**

Any Student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. The phone number for DSP is (213) 740-0776.

**11. Resources**

*11.1. Web page:* A class website will be setup on Backboard containing information about the course: syllabus, laboratory handouts, grades, miscellaneous information about weekly class activities, and an email directory of all people in the class. Use it as much as you find it useful. The web page can be accessed through the main stem https://Blackboard.usc.edu.

*11.2 Office Hours:* Office hours will be held weekly. Time and location for my office hours are at the beginning of the syllabus. Those of the unofficial teaching assistant will be decided with you in class. Both of us are available by email to help you as much as you need.

During weeks 10-13, every student will meet once a week so that progress on their projects can be assessed and any obstacles encountered solved.