Instructors:  
Betsey Dyer  
bdyer@wheatoncollege.edu  
SC-227  508.286.3951  
Hours: by appt. or  
MWF 8:30-9:20

Mark LeBlanc  
mleblanc@wheatoncollege.edu  
SC-103  508.286.3970  
Hours: by appt. or  
TBA

Meeting Times: Mon/Wed/Fri  11:30-12:20  
Room A118, csLab, or genetics lab (room 121)

Text:  
Travels in DNA Land (a working manuscript)

Additional readings and use of web sites will be assigned in class.  
The details are provided in the syllabus and/or URLs are listed in  
the Web Resources section on the next page.

An amazing blend of science, computing, and mathematics emerges when  
considering the molecule "Deoxyribonucleic Acid" (DNA). DNA is the blueprint  
of life for all organisms on Earth and throughout evolutionary time. Its  
distinctive and beautiful physical nature, a double helix of four bases, maps  
onto its functionality as a bearer of information, generation after  
generation. Fully sequenced genomes including the human genome have become  
the starting point for attempts to answer a wide range of biological and  
quantitative questions. This course explores DNA from the following points of  
view: (1) computer science for deciphering the meanings of the sequences of  
bases (A, C, G, and T) and writing new software to answer original questions;  
(2) molecular biology by using various laboratory techniques for visualizing  
and analyzing DNA; (3) organismal and evolutionary biology for interpreting  
the varied consequences of having such an information-rich and yet mutable  
molecule, and (4) applied mathematics when considering counting problems,  
introductory statistical analyses, and the implications of having too much  
data. As time permits, historical and ethical aspects of DNA are discussed.  
Students are encouraged to enhance their broader understanding of DNA by  
taking the connection with Philosophy 111 (Ethics). This DNA-242 course  
satisfies the Quantitative Analysis (QA) general education requirement.

GOALS for the COURSE:
1. Encourage creativity in programming with the Perl language and using  
regular expressions to explore sequences of DNA.
2. Instill confidence with solving quantitative problems.
3. Gain exposure to a breadth of DNA-related topics.
4. Foster personal experiences with a diversity of aspects of DNA.
   <insert your goals here>
Web Resources:

Online documentation for Perl
http://www.perldoc.com/perl5.8.0/pod/perlintro.html
http://www.perldoc.com/perl5.8.0/pod/perlsyn.html

The Timeline of Perl and its Culture v3.0
http://history.perl.org/PerlTimeline.html

From the PBS show NOVA: “Cracking the Code of Life”
Watch the program in 16 online video segments
http://www.pbs.org/wgbh/nova/genome/program.html

From the Dolan DNA Learning Center & Cold Spring Harbor Laboratory:
“DNA From the Beginning”. DNA from the Beginning is organized around key concepts. http://www.dnaftb.org/dnaftb/

NCBI: National Center for Biotechnology Information

Wheaton College Genomics Group
http://genomics.wheatoncollege.edu/

SOME OF THE PROGRAMS YOU WILL WRITE in Perl
#1 -- String Play
#2 -- Chargaff’s Counts
#3 -- Chargaff Revisited on Entire Genomes
#4 -- Gene Finder
#5 -- eLmer - a motif finder
#6 -- Building a Concordance

Because the lab and class preparation is costly both in money ($) and time, you must attend all classes and labs. When homework is “due in class”, this means at the beginning of the class on that particular day, e.g., we will collect your homework at the start of class. Homeworks that arrive after the class and up to one day late will lose 10%. No homeworks are accepted after one-day late.

It is expected that you spend at least 2 hours on reading and practice problems for every hour of lecture. This computes to at least 6 hours of work outside of class per week. This should be done throughout the semester. Programs must be electronically submitted by 5am of the day after the due date (yes, you get 5 extra hours if you need it and you are willing to stay up 😊). A program beyond 5am and one day late will be docked 10%. No programs can be submitted after more than one day late. Of course, one day late extends for 24 hours after the 5am deadline. Please assume all deadlines are fixed. Obviously, see one of us if you know of a conflict beforehand.

Honor Code Revisited: It goes without saying that all submitted work will be the student's own, in keeping with the Wheaton Honor Code. For labs, you may get help from fellow classmates, but remember that all submitted work must be your own. For homework, the programs and hand-written work must be your own from beginning to end.
YOUR GRADE

10% COMPUTER AND WET LABS
based on attendance in lab and
at guest lectures, hand-ins, check-offs

10% HOMEWORKS
Sept  2  Homework #1 –
Sept  7  Homework #2 –
Sept  9  Homework #3 –
Oct 21  Homework #4 –
various dates
- review of 1953 articles
- installation of Perl environment
- your own RegEx and DNA annotation by hand
- chalking RegEx in sidewalk art
- miscellaneous calculations in class

60% PROGRAMS (six)
Sept 16  Program #1 (10%)  - String Play
Sept 23  Program #2 (10%)  - Chargaffin Counts
Sept 30  Program #3 (10%)  - Chargaffin Counts Revisited on Genomes
Oct  12  Program #4 (10%)  - Gene Finder
Oct  28  Program #5 (10%)  - “eLmer” a motif finder
Nov 11  Program #6 (10%)  - Building a Concordance

20% Final Individual Project
Mon., Nov 21  Proposal   (5%)
week of Dec 5  Presentation (5%)
week of Dec 5  Software     (5%)
week of Dec 5  Results      (5%)

Note: Bonus Points will be awarded to students who discover errors (especially technical and factual errors) in the manuscript. The number of points awarded will be determined by the instructors.
**Week 1** (short week starting August 31)

31st WED (A118)

```
print "hello class";
```

Pre-evaluations

The THREE THREADS of this class (and keeping track of them as they intertwine)

1. Using computers to decipher information in DNA sequences
2. Working with and visualizing DNA in the wet and dry labs
3. History of our understanding of DNA

Why Perl?
(i) when the data you have is not quite enough to answer your question
(ii) learning to script in Perl is doable and fun
(iii) scripting (often called programming) encourages you to be creative with your questions

**Homework #1:** in each of the three articles, highlight the most interesting sections and explain why in the margins (due in class and on the board, Fri. Sept 2). **Come to class early and write on the board!**

**Readings:**

2nd FRI (A118)

**Homework #1 due in class** – you write on the chalkboard before class begins.


Homework #2: Read Chapters 1 and 2 of the manuscript. **Install (1) Perl and (2) a Programming Environment on your own computer.** Write at least one paragraph (and up to one page) about how the installations went. We are particularly interested in your assessment of the directions in the manuscript; include two sections on (a) your difficulties and (b) suggestions. (Due at the start of class on Wednesday Sept 7th).
Week 2

5th MONDAY is a holiday

7th WED (csLab)

Homework #2 (an account of your successful installation) due in class.

Introduction to Regular Expressions (RegEx): a tool for searching for patterns in DNA sequences—Be sure to have the manuscript with you; we will be using Chapter 3 “Regular Expressions”.

Homework #3 (two parts, A and B): (A) using color (pencils, hi-liter) annotate one page of DNA sequence, e.g., use colors to highlight certain patterns that you observe; (B) on a separate page, come up with a unique Question that requires a regex; include (i) the Question, (ii) the RegEx, and (iii) some of the words that match with the RegEx (i.e., your answers when using an English dictionary). Due in class, Fri. Sept 9.

Readings:

9th FRI (csLab)
More RegEx

Homework #3 Part (A) and (B) is due: (A) Write your RegEx on the whiteboard so that others may try it. (B) Tape your hand-annotation on the top of the board.

Week 3

12th MON (A118)
Playing with Strings of DNA using your first Perl programs.
Be sure to have Chapter 4 “Strings” with you in class.

“hello DNA Land”

Working with built-in string functions and built-in string operators
Working with numerical data and arithmetic operations

Program #1 “String Play” assigned, specification handed-out:
Due Friday, Sept 16
Week 3 continued …

14th WED (csLab)

Arithmetic operations in Perl and hands-on practice

Introduction to SmartFTP to transfer files around the Wheaton network and beyond

16th FRI (A118)

Mini Lecture: Researchers isolate and analyze the chemical parts (including DNA) of organisms. The function of DNA was unknown.

LAB PROJECT: Isolate DNA and calculate the length of total DNA in a bacterial cell, in a human cell, in a human body.

Program #1 “String Play” is due today (actually, it is due by 5am, Sept. 17th)

Week 4 (week of Sept 19)

19th MON (A118)

Introduction to Subroutines in Perl
Have the “Subroutine” chapter with you in class

Program #2 “A Chargaffian Sequence Counter” specification handed-out:
Due Friday, Sept 23

21st WED (csLab)

Hands-on practice with writing your own functions and subroutines in Perl

23th FRI (A118)

Mini-Lecture: Researchers use the “latest” staining techniques to visualize chromosomes (literally “colored bodies”)

LAB PROJECT: Staining chromosomes with a traditional method.

Program #2 “A Chargaffian Sequence Counter” is due today
Week 5

26th MON (A118)
Handling entire genomes: Opening, storing, and reading from large files

Program #3 “Chargaff Revisited on Entire Genomes”; specification handed-out
Due: Friday, Sept 30th

Have the “Files” chapter with you in class

28th WED (A118)
Lecture: Darwin notices the natural variations and wonders about heredity, the missing component for the origin of new species.

30th FRI (A118)
Film Festival: on replication, transcription, and translation

Program #3 due today

Week 6

3rd MON (A118)
Making decisions: conditional control if ... elsif ... elsif ... else
Make sure you bring the if-else chapter with you to class this week

Program #4 “GeneFinder”; specification handed-out
Due: Wed., Oct 12th

5th WED (A118)
Doing it over and over: repetitional control (often called “loops”)
while, for, foreach
Make sure you bring the while chapter with you to class this week

7th FRI (csLab)
Putting it all together: hands-on practice with control structures
Week 7  
(Short week of Oct 10)

10th MON  (Fall Break)

12th WED  (csLab)

   BLAST Lab – using “the” Bioinformatics tool of the day
   Program #4 due today

14th FRI  (meet outside Room/Office 231)

   Sequencing lab for demo of new DNA sequencer
   Guest Professor Shawn McCafferty

Week 8  
(Week of Oct 17)

17th MON  (A118)

   Using Regular Expressions in Perl to seek patterns in words and in DNA

   **Program #5** “eLmer – a motif finder” specification handed-out:
   Due Friday, Oct. 28th

   **Homework #4** (due Fri., Oct. 21st at the start of class) (rain date Oct 24) Stake out a section of sidewalk somewhere around the dimple. Write (1) a query, (2) a RegEx, and (3) a solution in large chalk on the sidewalk.

19th WED  (ICUC Lab)

   Chromosomes in the Imaging Lab (ICUC) with Professor Bob Morris.

21st FRI  (meet in Science Center Lobby)

   “Gallery opening” -- We’ll walk around campus and view your Homework #4
Week 9

24th MON: (A118)

Arrays in Perl

29th WED: (csLab)

Hands-on practice with arrays in Perl

28th FRI: (A118)

Program #5 due today

Mini Lecture: Running gels becomes the essential lab technique for studying DNA.

LAB: Run a DNA gel. And understand what we just did! Graph and interpret results.

Demo: Look at sequencing gels and read sequences from them.

Week 10

31st MON (A118)

Program #6 “Building a Concordance” specification handed-out:
Due Friday, Nov. 11th

Introduction to Hash Tables in Perl

Nov 2nd WED (csLab)

Hands-on practice with Hash Tables in Perl

4th FRI (A118)

Final Project Specification for Individual Projects
What might you do?
How will you present your results?
What is expected of you?
Week 11  (week of Nov 7)

- 7th MON (A118)

  Like Totally Random
  Introduction to probability

9th WED: (csLab)

  Generating random files of DNA with Perl

11th FRI: (A118)

  Program #6 due today
  Mini-lecture: Family trees may be constructed from GenBank sequences
  LAB: Try making a family tree.

Week 12  (week of Nov. 14)

14th MON (A118)

  Catch-up on topics ….

16th WED (A118)

  Lecture: Evolution and Randomness

18th FRI (A118)

  Project Summary Presentations on Monday, Nov. 21
  Sample presentation from Dyer and LeBlanc
---

### Week 13

(very short week of Nov 21)

**21st MON (A118)**  
Student presentations of Project Summaries  
Students present in class to get feedback and suggestions  
(worth 5% of final project grade)

**WED and FRI (Nov 23-27)**  
Thanksgiving Break

---

### Week 14

(week of Nov 28)

**28th MON (A118)**  
Lecture: Super computers are used to decipher and reconstruct 3D protein and DNA structures

**30th WED (csLab)**  
LAB: Look at some of the 3D shapes at GenBank.

**2nd FRI (csLab)**  
Opportunity to whiteboard and program in the lab on final projects

---

### Week 15

(week of Dec 05)

**5th MON: (csLab)**  
Final Presentations: Software Demos and Poster Session Day I

**7th WED: (csLab)**  
Final Presentations: Software Demos and Poster Session Day II

**9th FRI: (csLab)**  
Final Presentations: Software Demos and Poster Session Day III  
Course Evaluations

---

End of the semester (no Final Exam)
REFERENCES


