Instructor: Mark LeBlanc  SC-103  508.286.3970  mleblanc@wheatoncollege.edu
Office Hours: MW 11:30-12:30, 2-3

Meeting Times: Monday, Wednesday, Friday  10:30 - 11:20
Room 154, SciCenter or csLab—A118b

Text: Perl for Exploring DNA by LeBlanc and Dyer
(Oxford University Press 2007)

10 GOALS for this course

(0) You are at a cocktail party and the topic of genomes comes up. You are able to recall significant phrases, terms, and techniques and your understanding of the main ideas and concepts enables you to lead the conversation for a while. Your friends raise their eyebrows.

(1) You learn to identify and classify problems that are candidates for a computer to handle.

(2) You demonstrate the ability to think algorithmically, breaking what originally seems like an overly complicated problem into a series of smaller, manageable tasks.

(3) You learn to craft creative solutions by “writing software” (“to program”, “to script”).

(4) You appreciate when to include external modules of previously written software.

(5) You design experiments to first solve small computationally-intensive tasks (e.g. on one gene sequence) and then scale your solutions to very large sets of data (e.g. all genes in a genome).

(6) You apply problem solving strategies previously learned to novel situations.

(7) You learn to professionally document your software and produce quality summaries, graphs, and reports of your results.

(8) You begin to appreciate genome-wide microarray experiments and the accompanying software that is required to analyze results.

(9) You feel empowered to evaluate the ethical implications of your work and learn to appraise, critique, and defend your own as well as the work of others.

Connections: The Bigger Picture

This course is part of the connection “Genes in Context” with Philosophy 111 (Ethics). Throughout the semester, students will be exposed to the ethical aspects of living in a post-genomic world and the increasing use and challenges of sequenced genomes as applied to “personalized medicine”. A series of assignments and events with students from other classes will include:

• A showing of the movie GATTACA and follow-up discussion.
• Talks and discussions with a genetic counselor and professor of bioethics.
• Student-produced, one-minute YouTube “commercials” of companies currently promoting and selling medical profiles based on individual genomes. The commercials will be framed from one of two points of view: (i) from the point of view of the company (e.g., 23andME, deCODEme, Navigenics, etc) or (ii) from a consumer advocacy point of view.
Catalog Description
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 and hundreds of microbial genomes have become the starting point for attempts to answer a wide range of biological and quantitative questions.

This DNA-242 course satisfies the Quantitative Analysis (QA) requirement.

Exact pages to reading and use of web sites will be assigned in class. URLs are listed as assigned/needed in the onCourse (Moodle) website.

Web Resources:
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/

DNA interactive: http://www.dnai.org/a/index.html

NCBI: National Center for Biotechnology Information

A Perl FAQ
http://theoryx5.uwinnipeg.ca/CPAN/perl/pod/perlfaq1.html

Notes on grading, due dates, and submission procedures
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.

Programs are due on various dates (see detailed syllabus); however, since we know from experience that many students like to use the last night for testing, we will allow you to submit your programs until 5am the following day. For example, Program #1 is due THUR, Sept 17th, but you can submit it electronically until 5am FRI, Sept 18th -- Careful! The course website (onCourse) makes it appear as if the program is due on Friday, but remember, Friday at 5am! Programs submitted after this time will lose 20% each day it is late up to two days late. A program submitted more than two days late will not be accepted and will receive a grade of zero. In addition to electronically submitting your Perl programs (more instructions will be given in class), you must submit a hardcopy printout of your program in class after it is due. Note that Perl programs must always be printed in landscape mode and your source listing must be stapled to your sample output and documentation (README) file. (Your professor will say @%^$##* and will deduct points if you do not staple.)
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. Please assume all deadlines are fixed. Obviously, see your professor 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. All homeworks, Perl programs, and other hand-written work must be your own from beginning to end unless otherwise noted in the instructions (e.g., paired work).

Guest Lectures
Jenny Lanni, Biology “The Central Dogma”, Sept. 4th
Teresa Celada, Philosophy “Designer Babies”, Oct. 2nd
Mike Kahn, Statistics “Randomness”, Oct. 30th

YOUR GRADE
5% ATTENDANCE/PARTICIPATION - based on attendance/participation in all sessions
10% QUIZZES (5) - see detailed schedule (usually at start of class)
15% HOMEWORKS
  Sept 4  Homework #1 – 1 point - installation of Perl environment
  Sept 11 Homework #2 – 2 points - RegEx (Part I)
  Sept 30 Homework #3 – 10 points - one-minute “personalized medicine” video
  Oct 02  Homework #4 – 2 points - chalking RegEx in sidewalk art

50% PROGRAMS (six)
  Sept 17  Program #1 (5%) - String Play
  Sept 24  Program #2 (5%) - Chargaffin Counts
  Oct 2    Program #3 (10%) - Gene Finder
  Oct 16   Program #4 (10%) - “eLmer” a motif finder
  Oct 30   Program #5 (10%) - Fuzzy Olfactory Gene Counter
  Nov 18   Program #6 (10%) - Building a Concordance

20% Final Project
  Nov 23  Written Proposal (Intro & Methods) (5%)
  Dec 7, 9, 11 Oral Presentation (5%)
  Dec 11  Final Paper, Software, and Results (10%)

Throughout the semester, you will have the opportunity to show Superior Effort, for example, a homework or programming assignment has additional steps. Remember, when determining your final, overall grade: an ‘A’ is superior, ‘B’ is above average, ‘C’ is average effort, etc.

As you can see, the final projects are a significant part of your final grade (20%). Your professor will determine the team pairs. Each pair of students will be given a certain number of points and it will be up to the pair to determine how those points should be allocated. For example, if the pair was given 150 points, then the pair could decide that Person A, who did more of the work, should get 80 and Person B should get 70. If they pair felt they each did the same amount of effort, then each would get 75 points. Each of the three parts (Proposal, Oral Presentation, Final Paper including your software and results) will receive a separate grade.

Note: Bonus Points will be awarded to students who discover errors (especially technical and factual errors) in the book. The number of points awarded will be determined by the instructor. Before you report an error, please check the Errata to see if someone else has already found this error (see onCourse link to the Errata).
Week 1

Sept 02, WED

print "hello DNA";

Reading:
- Text: Acknowledgements, Preface, Chapters 1 and 2.
- Reading on Central Dogma (read before class on Friday); see onCourse

Homework #1 (Due at start of class, Friday Sept 4): Install Perl summary.

Sept 04, FRI

Guest Lecture: Dr. Jenny Lanni, Biology – “The Central Dogma”

Reading:
- Text: Ch. 4 p47-61 (you need to read this before next WED’s lab)
- Bring your textbook to lab next WED.

Week 2

Sept 07, MON – Labor Day <no class>

Sept 09, WED (meet in csLab)

Quiz #1 – Collectively we give a brief Central Dogma lecture. (See Homework #1).

csLab: Playing with Regular Expressions (“Regex”):
- Regex: a tool for searching for patterns in DNA sequences—Be sure to have the book with you; we will be using Chapter 4 “String Play with Regular Expressions”.

Homework #2 (Due at start of class, Friday Sept 11): Finish Part I handout on Regex

Opportunity for Superior Effort: If you really get into regex’s, complete Part II … and even make yourself a regex bookmark.

Fun Readings:

Sept 11, FRI

Homework #2 (Playing with Regex – Part I) due at start of class.

Good practices when printing output, using variables, and built-in functions;
- Introduction to “genic” and “intergenic” regions and codons.
Intro to Perl: “hello DNA Land”, print, variables, built-in functions
- length, lc, uc, reverse, index, substr, tr, s

Reading:
- Text: Ch. 3 p17-46 and some reading on “algorithms” p203-211

a1 Specification is ready: (see onCourse for Starter Kit)
- a1 algorithm is due MON, Sept 14
- a1 program is due Thursday, Sept 17
Week 3

Sept 14th MON
   a1 algorithm is due in class

      csLab – playing with Perl … strings, strINGS, STRINGS …
      index, substr, reverse, lc, uc
      and string operators
      tr and concatenation (.)

Sept 16th WED
   a1 submission is due tomorrow, THUR Sept 17 (really, Friday by 5am)

      Assign one-minute videos of “personalized medicine” companies
      e.g., deCODEme, 23andMe, Knome, Illumina, Pathway Genomics, etc.
      (Due: Wed, Sept 30 in class)

   more Perl

Sept 18th FRI
   a2 Specification is ready: (see onCourse for Starter Kit)
      a2 algorithm is due MON, Sept 21
      a2 program is due Thursday, Sept 24

   Using Perl to do calculations: arithmetic expressions

   Specifications “Specs” for Program #2 “Chargaffian Counts” and instructions about
   completing the program.

   Reading:
      Text: Ch. 5 (all) and Ch. 8 Reading from Files p135-142

Week 4

Sept 21st MON
   a2 algorithm is due in class

      csLab: good practices of using arithmetic operators, precedence, math functions, and
      formatted output with printf

      Using a subroutine written by someone else (e.g., readInDNA )

   Quiz #2 next time on Wed, Sept 23rd at start of class
Sept 23rd WED  
Quiz #2 at start of class  
a2 due on Thursday, Sept 24  

conditional control (if-elsif-elseif-else)  

Reading:  
Text: Ch. 6 p93-98, 103-106 and Ch. 7 (all)

Sept 25th FRI  
a3 Specification is ready: (see onCourse for Starter Kit)  
a3 algorithm is due MON, Sept 28  
a3 program is due Friday, Oct 2  

one-minute videos due next Wed, Sept 30  

Transcription – Translation … and a3 specification

Week 5

Sept 28th MON  
a3 algorithm is due in class  

Good practices of using if-else; using if and index together;  
Using subroutines revisited  
Specifications “Specs” for Program #3 “Gene Finder” and instructions about completing the program.

Sept 30th WED  
one-minute videos due today  

Regex Chalk: due Friday morning around the dimple (rain date, Sat morning)  

csLab – practice with if and index  

Reading (in preparation for Friday’s guest lecture): TBA

Oct 2nd FRI  
Regex Chalk: Due in morning before classes  
a3 due today  

Guest Lecture: Dr. Teresa Celada, Philosophy -- “Designer Babies”
Week 6

Oct 5\textsuperscript{th} MON
\textbf{a4 Specification} is ready: (see onCourse for Starter Kit)
\textbf{a4 algorithm} is due FRI, Oct 9\textsuperscript{th}
a4 program is due Friday, Oct 16\textsuperscript{th}

while loops, greediness in regular expressions, commenting those tough regex

Introduction to gene regulation

\textbf{Reading:}
Text: Ch. 6 p98 – 107 (if-elsif-else) and while loops, p107-115

Oct 7\textsuperscript{th} WED
Quiz #3

csLab – p114 #1,2,3,6

Oct 9\textsuperscript{th} FRI –
\textbf{a4 algorithm} is due

More looping, more looping, more looping, more looping ….

Week 7

Oct 12\textsuperscript{th} MON and TUES 13\textsuperscript{th} “Fall Break”

Oct 14\textsuperscript{th} WED
Introduction to modules and BioPerl

Oct 16\textsuperscript{th} FRI
\textbf{a4 due today}

More BioPerl
Week 8

Oct 19\textsuperscript{th} MON
\begin{itemize}
\item \textbf{a5 Specification} is ready: (see onCourse for Starter Kit)
\item \textbf{a5 algorithm} is due FRI, Oct 23rd
\item a5 program is due Friday, Oct 30\textsuperscript{th}
\end{itemize}

\textbf{csLab} -- BLAST Lab (Homework #6; due at the end of lab)

Specifications “Specs” for Program #5 “Fuzzy Olfactory Gene Finder” and instructions about completing the program

Olfaction and seven trans-membrane proteins (Part I)

\textbf{Reading:}
\begin{itemize}
\item Text: 13 (all) – Randomness – “So, like totally random, dude”
\end{itemize}

Oct 21\textsuperscript{st} WED
\textbf{Quiz #4}

Using your BLAST output as input …

Olfaction and seven trans-membrane proteins (Part II)

Oct 23\textsuperscript{rd} FRI
\begin{itemize}
\item \textbf{a5 algorithm} is due
\end{itemize}

\textbf{csLab}: more subroutines and Perl’s rand

Week 9

Oct 26\textsuperscript{th} MON
Tweaking regex

\textbf{Reading:}
\begin{itemize}
\item Text: 12 (all) – Regex Revisited
\end{itemize}

Oct 28\textsuperscript{th} WED
Introduction to arrays and Perl’s \texttt{split} and \texttt{join} functions

\textbf{Reading:}
\begin{itemize}
\item Text: 9 – Arrays, p151-168
\end{itemize}

Heads/Tails homework – due at start of guest lecture on Friday, 30\textsuperscript{th}
Oct 30th FRI

a5 due today

Guest Lecture: Dr. Mike Kahn -- “Randomness”

Week 10

Nov 02nd MON
    csLab: Intro to microarray analysis – MagicTool software

Nov 04th WED
    csLab: Continuation of microarray analysis – MagicTool software

Nov 5th THUR
    Superior Effort: Join us in csLab to help the Genetics class use Magic Tool, 2-5pm.

Nov 6th FRI
    a6 Specification is ready: (see onCourse for Starter Kit)
      a6 algorithm is due WED, Nov 11th
      a6 program is due Wed, Nov 18th

    Sorting with arrays and hash tables
    csLab – Practice with arrays and hash tables

    Reading:
      Text: Ch. 9 – Arrays, p168-177 and Ch. 10 Hash Tables, p179-192

Week 11

Nov 9th MON
    More hash tables and help on Program #6
    Lots(!) of Reading:
      Counting motifs p192-202
      Breaking Sequence into Motifs p214-216

Nov 11th WED
    a6 algorithm is due

    csLab – Entering a valid motif $MIN <= \text{size} <= \MAX$

Nov 13th FRI

    csLab: Work on a6
Week 12

Nov 16\textsuperscript{th} MON
Final Project Specification for Individual Projects
  What might \textit{you} do?
  How will you present your results?
  How do you write the Methods section of a paper?
  What is expected of you?

Proposals due MON, Nov 23\textsuperscript{rd}

Nov 18\textsuperscript{th} WED
\texttt{a6} due today

Class time to search for appropriate datasets for your final project: Go to NCBI, select sequences, store them for future use.

Nov 20\textsuperscript{th} FRI
Quiz #5

Samples of written and oral presentations—an analysis of examples

Week 13

Nov 23\textsuperscript{rd} MON
Final Project Proposal due: type-written
Title, Introduction, and Methods due by the beginning of class

csLab: Open lab time for individual and group help on final projects

Nov 25\textsuperscript{th} WED – Nov 27\textsuperscript{th} FRI
Thanksgiving Break

Week 14

Nov 30\textsuperscript{th} MON
[open TBA]

Dec 2\textsuperscript{nd} WED
[open TBA]

Dec 4\textsuperscript{th} FRI
Lab time for final presentations
Week 15

Dec 7th MON
Oral Presentations on Projects

Dec 9th WED
Oral Presentations on Projects

Dec 11th FRI
Oral Presentations on Projects

Experimental program and paper are due at the BEGINNING of CLASS

Course Evaluations and Farewells