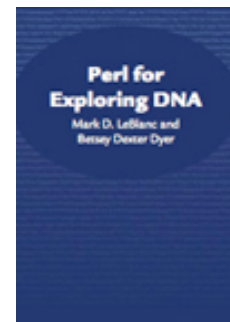


syllabus for
DNA
bio/comp 242

Instructors: **Betsey Dyer** **Mark LeBlanc**
bdyer@wheatoncollege.edu mleblanc@wheatoncollege.edu
SC-227 508.286.3951 SC-103 508.286.3970
Hours: by appointment/TBA Hours: M/T/W 2-3, R 10-11

Meeting Times: Tuesday, Thursday 11:00-12:20
Room A118, Science Center (on some days we will continue on
to either the csLab or the genetics lab, room 121)

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



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. Historical and ethical aspects of DNA also will be discussed. *Students are encouraged to enhance their broader understanding of DNA by taking the **connection "Genes in Context"** 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>
5. _____

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.dnafb.org/dnafb/>

NCBI: National Center for Biotechnology Information

NCBI creates public databases, conducts research in computational
biology, develops software tools for analyzing genome data, and
disseminates biomedical information. <http://www.ncbi.nlm.nih.gov/>

A Perl FAQ

<http://theoryx5.uwinnipeg.ca/CPAN/perl/pod/perlfaq1.html>

The **Timeline of Perl** and its Culture v3.0

<http://history.perl.org/PerlTimeline.html>

Wheaton College Genomics Group

<http://genomics.wheatoncollege.edu/>

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 Mondays (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 MON, Sept 17th, but you can submit it electronically until 5am Tue, Sept 18th. Programs submitted after this time will lose 20% each day it is late upto 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 the immediate Tuesday after it is due. Note that Perl programs must **always** be **printed in landscape mode** and your source listing must be **stapled**. (We will deduct points if you do not print in landscape and/or you do not staple your source code listing).

Final projects are a significant part of your final grade (20%). Your professors will determine the team pairs. Each pair of students in a group will receive the same grade for each part of the project (see next page).

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 one of your professors 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

Shawn McCafferty, Biology	“Sequencing DNA”, Sept. 20 th
Mike Kahn, Statistics	“Randomness”, Oct. 18 th
Teresa Celada, Philosophy	“Ethical Questions”, Oct. 11 th and “More Questions”, Nov. 27 th

Ethical Moments (sprinkled throughout the semester)

YOUR GRADE

5% ATTENDANCE/PARTICIPATION	based on attendance and participation in lab, lecture, and guest lectures	
10% QUIZZES (five)	on Perl syntax, dates TBA, (usually at the beginning of the class following the one where you get new project specs)	
15% HOMEWORKS (eight)		
Sept 4	Homework #1 – 2 points	- review of 1953 articles
Sept 6	Homework #2 – 2 points	- installation of Perl environment
Sept 11	Homework #3 – 2 points	- your own RegEx and DNA annotation by hand
Sept 26	Homework #4 – 2 point	- attend the Norman Johnson Lecture
Oct 5	Homework #5 – 2 points	- WIKI on dog gene regulation
Oct 16	Homework #6 – 2 points	- BLAST lab
Oct 19	Homework #7 – 1 point	- chalking RegEx in sidewalk art
Nov 1	Homework #8 – 2 points	- Spider web explanations, SCOPUS search (Homework #8 will be with student pairs)
50% PROGRAMS (six) These are due on Mondays		
Sept 17	Program #1 (5%)	- String Play
Sept 24	Program #2 (5%)	- Chargaffin Counts
Oct 1	Program #3 (10%)	- Gene Finder
Oct 15	Program #4 (10%)	- “eLmer” a motif finder
Oct 29	Program #5 (10%)	- Fuzzy Olfactory Gene Counter
Nov 12	Program #6 (10%)	- Building a Concordance
20% Final Project		
Nov 20	Proposal (Introduction & Methods)	(5%)
Dec 4 or 6	Oral Presentation (15 minutes)	(5%)
Dec 6	Final Paper, Software, and Results	(10%)

As you can see, the final projects are a significant part of your final grade (20%). Your professors will determine the team pairs. Each pair of students in a group will receive the same grade for each part of the project. Each of the three parts (Proposal, 15 minute 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 instructors.

Week 1**Aug 30th THURS**

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

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

- (1) Writing software to decipher information in DNA sequences
- (2) Working with and visualizing DNA in the wet and dry labs
- (3) History and ethical implications of our understanding of DNA

Why Programming?

- (i) when the data you have is not quite enough to answer your question
- (ii) learning to design algorithms provides an appreciation for the rigor and precision that is required to solve computationally-intensive problems
- (iii) an intensive experience with solving problems by writing your own programs will help you lead teams throughout your career
- (iv) scripting (often called programming) encourages you to be creative with your questions and adventurous like an explorer

Homework #1: in each of the three articles, highlight the most interesting sections and and come up with at least one question from each paper. The three questions are due in class both on the board and typed-up on Tues. Sept 4. **Come to class early and write one of your questions on the board!**

Readings:

Franklin, R. and Gosling, R.G. (1953). Molecular Configuration in Sodium Thymonucleate. *Nature*, 171, 740-741.

Watson, J.D. and Crick. F.H.C. (1953). A Structure for Deoxyribose Nucleic Acid. *Nature*, 171, 737-738.

Wilkins, M.H.F., Stokes, A.R., and Wilson, H.R. (1953). Molecular Structure of Deoxypentose Nucleic Acids. *Nature*, 171, 738-740.

Week 2**Sept 4th TUES**

Homework #1 due in class – you write a least one good question about the readings on the chalkboard *before* class begins.

A brief history of the discovery of DNA structure: 50+ years since the joint publications on the structure of DNA -- Franklin, Watson, Crick, and Wilkins in *Nature*, v171, 1953.

Homework #2: Read Chapters 1 and 2. **Install (1) Perl and (2) a Programming Environment on your own computer.** Write at least two paragraphs (and up to one page) about how the installations went. We are particularly interested in your assessment of the directions in the book; include two sections on (a) Your Difficulties and (b) Suggestions to the Authors. **(Due at the start of class on Thursday Sept 6th).**

Sept 6th THURS

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 book with you; we will be using Chapter 4 “String Play with 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 (e.g., red for any direct repeats, green for mirror repeats); (B) on a separate page, come up with a unique Question that requires a regex when searching individual words in an English dictionary. Follow the format in the book, that is, include (i) the Question, (ii) the RegEx one could use to answer the question, and (iii) some of the words that match with the RegEx (i.e., your answers when using an English dictionary). **Due in class, Tues. Sept 11th.**

Readings:

Agee, J. (1996). So many dynamos! And other palindromes. Farrar, Straus, Giroux Publishers, New York, NY.

Agee, J. (2000). Elvis Lives and other anagrams. Farrar, Straus, Giroux Publishers, New York, NY.

Week 3 ---- Chapter 3

Sept 11th TUES

Homework #3 is due.

Intro to Perl: **“hello DNA Land”**

Good practices when printing output, using variables, and built-in functions

Specifications “Specs” for Program #1 “String Play” and instructions about completing the program. Introduction to “genic” and “intergenic” regions and codons.

Sept 13th THURS

csLab – working with Perl’s string functions:

`index`, `substr`, `reverse`, `lc`, `uc`

and string operators

`tr` and concatenation (`.`)

Week 4 ---- Chapter 5 and 8

Program #1 due electronically by Monday 17th (with 5 extra hours allowed if you want to work all night and turn it in by 5 o'clock in the morning of the 18th.)

Sept 18th TUES

Good practices of using arithmetic operators, precedence, math functions, and formatted output with `printf`

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

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

Sept 20th THURS

Guest Lecture: “DNA Sequencer”
Professor Shawn McCafferty, Biology

Week 5 --- Chapters 6 and 7

Program #2 due electronically by Monday 24th (with 5 extra hours allowed if you want to work all night and turn it in by 5 o'clock in the morning on the 25th).

Sept 25th TUES

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.

An introduction to transcription and translation.

Hand out reading (essay due Oct. 5):
“Genetics and the Shape of Dogs.”

Sept 26th WED -- 5:30pm, Hindle Auditorium

The Norman Johnson Lecture by Andy van Dam of Brown University
“*When is the Pen Mightier than the Keyboard?*”

Sept 27th THURS

csLab – “Going Back for More” with `if-elsif-else`, `index`, subroutines

Week 6 ---- Chapters 6, 7, and 12

Program #3 due electronically by Monday 1ST (with 5 extra hours allowed if you want to work all night and turn it in by 5 o'clock in the morning on the 2nd.)

Oct 2nd TUES

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

Specifications “Specs” for Program #4 “eLmer a Motif Finder” and instructions about completing the program

Introduction to gene regulation

Oct 4th THURS

csLab – Regular expressions revisited (in Perl code)

Oct 5th FRI – Wiki homework due

Week 7**Oct 9th TUES** “Fall Break”**Oct 11th THURS**

Guest Lecture: “Professor Teresa Celada, Philosophy

Week 8 ---- Chapters 11 and 13

Program #4 due electronically by Monday 15ST (with 5 extra hours allowed if you want to work all night and turn it in by 5 o'clock in the morning on the 16th.)

Oct 16th TUES

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)

Oct 18th THURS

Guest Lecture: Professor Mike Kahn, Statistics

Olfaction and seven trans-membrane proteins (Part II)

Oct 18th THURS continued

Sidewalk Chalk Art on the Dimple (Homework #7): Stake out a section of sidewalk somewhere around the dimple. Write in LARGE PRINT (1) a query, (2) a RegEx, and (3) a solution in large chalk on the sidewalk.

Oct 19th FRI (rain date Saturday morning, Oct 20)

Sidewalk Chalk Art due somewhere on the walkways around main Quad by 8am.

Week 9 ---- Chapter 9

Oct 23rd TUES

Olfaction and seven trans-membrane proteins (Part III)

Introduction to arrays and Perl's `split` and `join` functions

Oct 25th THURS

csLab – “Going Back for More” – Practice with arrays

Hand out reading (read in preparation for Tue, Oct 30th class):

“Taken For a Spin: Scientists look to spiders for the goods on silk.”

Week 10 ---- Chapters 9 and 10

Program #5 due electronically by Monday 29th (with 5 extra hours allowed if you want to work all night and turn it in by 5 o'clock in the morning on the 30th.)

Oct 30th TUES

Introduction to Spider Web proteins (read associated article before this class)

Homework #8 (in lab): Using BLAST and the literature review tool, SCOPUS.

Nov 1st THURS

Homework #8 (Spider Webs) is due.

Sorting with arrays and hash tables

Specifications “Specs” for Program #6 “Building a Concordance” and instructions about completing the program

csLab – Practice with hash tables

Week 11**Nov 6th TUES**

More hash tables and help on Program #6

Nov 8th THURS

Final Project Specification for Individual Projects

What might *you* do?

How will you present your results?

How do you write the Methods section of a paper?

What is expected of you?

Week 12

Program #6 due electronically by Monday Nov. 12th (with 5 extra hours allowed if you want to work all night and turn it in by 5 o'clock in the morning on the 13th.)

Nov 13th TUES

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

Nov 15th THURS

Samples of written and oral presentations—an analysis of examples

(If time permits: family tree and/or 3-D visualization of protein structures)

Week 13**Nov 20th TUES**

Final Project Proposal: type-written co-authored Title, Introduction, and Methods due by the beginning of class

Open lab time for individual and group help on final projects

Nov 22th THURS

Thanksgiving Break

Week 14

Nov 27th TUES

Guest Lecture: Professor Teresa Celada, Philosophy

Nov 29th THURS

Lab time for final presentations

Week 15

Dec 4th TUES

Oral Presentations on Projects

Dec 6th THURS

Oral Presentations on Projects

Program and paper are due at the **BEGINNING of CLASS**

Course Evaluations and Farewells

REFERENCES

- Agee, J. (1996). So many dynamos! And other palindromes. Farrar, Straus, Giroux Publishers, New York, NY.
- Agee, J. (2000). Elvis Lives and other anagrams. Farrar, Straus, Giroux Publishers, New York, NY.
- Cunningham, A. (2007). "Taken For a Spin: Scientists look to spiders for the goods on silk." *Science News*, v171, April 14, 2007, 231-234.
- Dwyer, R.A. (2003). Genomic Perl – From Bioinformatics Basics to Working Code (Ch. 1 The Central Dogma, including transcription and translation), Cambridge University Press.
- Franklin, R. and Gosling, R.G. (1953). Molecular Configuration in Sodium Thymonucleate. *Nature*, 171, 740-741.
- LeBlanc, M. and Dyer, B. (2007). Perl for Exploring DNA. Oxford University Press.
- Ostrander, E. (2007). Genetics and the Shape of Dogs. *American Scientist*, Sept-Oct 2007, v95, 406-413.
- Watson, J.D. and Crick, F.H.C. (1953). A Structure for Deoxyribose Nucleic Acid. *Nature*, 171, 737-738.
- Wilkins, M.H.F., Stokes, A.R., and Wilson, H.R. (1953). Molecular Structure of Deoxypentose Nucleic Acids. *Nature*, 171, 738-740.

