

BioSci 1540: Computational Biology

Nathan L. Brouwer

2021-08-13

Contents

1	Welcome	7
2	Course outline	9
2.1	UNIT1 1: Getting started	9
2.2	UNIT 2: Computational Ecology	9
2.3	UNIT 3: Phylogenetics & Bioinformatics	10
2.4	UNIT 4: Bioinformatics 2 - Sequence Alignment and BLAST searches	10
2.5	Unit 5: Bioinformatics 3 - Algorithms and statistics of BLAST .	10
3	Introduction	11
4	Meeting times	13
5	Instructor	15
6	Academic integrity	17
7	Assessments - Homework policies	19
8	Assessments Overview	21
8.1	Online Test-taking policies	21
8.2	Final Exam Policies	21
8.3	Assessments	21
9	Communication - University email policy	23
10	Course communication - Email and Canvas Messages	25
11	Course communication - Overview	27
11.1	Primary modes of communication	27
11.2	Secondary modes of communication	27
12	Catalog description	29

13 Course materials overview	31
14 Course materials - Software	33
14.1 Primary software	33
14.2 Code sharing	34
14.3 R packages	34
15 Bibliography	35
15.1 Bioinformatics & computational biology	35
15.2 Phylogenetics	35
15.3 Statistics	35
15.4 Reference Books	36
15.5 Reference articles, blogs etc	36
16 Disability Resource & Services	37
17 Extra credit policies	39
18 File names for homework	41
18.1 Course Goals	41
19 Grading scale	43
20 Meeting times	45
21 Mental health & wellness	47
21.1 Resources:	47
22 Office hours	49
22.1 INSTRUCTOR OFFICE HOURS	49
22.2 UTA OFFICE HOURS	49
23 Point breakdown	51
24 Semester schedule	53
25 Skills	55
26 TopHat	57
26.1 In-class response / Homework system	57
27 Updates to schedule & syllabus	59
28 Appendix: graduate school	61
29 Cheat sheets	63
30 Appendix: Getting involved in (computational) research	65

<i>CONTENTS</i>	5
31 Shortcuts in R	67

Chapter 1

Welcome

This is the syllabus.

```
install.packages("bookdown")  
# or the development version  
# devtools::install_github("rstudio/bookdown")
```


Chapter 2

Course outline

2.1 UNIT1 1: Getting started

2.1.1 Week 1: Hello Computational Biology!

Reading: NIH Working definition of bioinformatics and computational biology
Supplemental reading: Markowetz (2017). All biology is computational biology
Supplemental reading: Poisot et al (2019). Data based, synthesis drive: Setting the agenda for computational ecology

Supplemental reading: Wikipedia. Computational biology

2.1.2 Week 2: R Camp

Reading: Thieme (2018). R generation

2.1.3 Week 3: Stats Camp

Reading: Meillere et al (2015). Traffic noise exposure affects telomere length in nestling house sparrows

2.2 UNTI 2: Computational Ecology

Reading: Dennis et al (1991) Estimation of growth and extinction parameters for endangered species

See Abstract, Introduction and Grizzly bear case study (pg 131-133). Reference: Morris and Doak (2002). Quantitative conservation biology. Reference: Steven (2009). A primer of ecology in R.

2.3 UNIT 3: Phylogenetics & Bioinformatics

Higgs and Atwood 2005. Bioinformatics and Molecular Evolution Lycett et al 2007. Phylogenetic analyses of behavior support existence of culture among wild chimpanzees Lycett et al 2010. Are behavioral differences among wild chimpanzee communities genetic or cultural? An assessment using tool-use data and phylogenetic methods.

2.4 UNIT 4: Bioinformatics 2 - Sequence Alignment and BLAST searches

Hildebrand and Soriano (1999). Shroom, a PDZ domain-containing actin-binding protein, is required for neural tube morphogenesis in mice. Cell. (link ; pdf). See section "shrm Encodes a Novel PDZ Domain Protein" and Figure 4, pg 488-489. Hagens, Hilderand et al (2006). A new standard nomenclature for proteins related to Apx and Shroom (link)

2.5 Unit 5: Bioinformatics 3 - Algorithms and statistics of BLAST

Chapter 3

Introduction

This is the syllabus

Your first assignment in the class will be to read through the syllabus and complete a “Syllabus treasure hunt” assignment on TopHat. After you’ve complete that assignment feel free to contact me with any questions you have about course policies.

Chapter 4

Meeting times

Chapter 5

Instructor

Nathan L. Brouwer, PhD Office: Crawford Hall
A351 (the “Bridge”)
Email: nlb24 at pitt.edu

Chapter 6

Academic integrity

Dr. Brouwer, the Bio Sci department, and the University all take academic integrity very seriously. Cheating includes any form of plagiarism, including copying other students' work or using other resources without proper attribution.

1. If you are caught cheating on a graded assignment, you will receive a zero on the assignment.
2. If you are caught cheating on an exam, you will receive a zero on the exam, an F in the course, and an Academic Integrity Violation Report will be filed.

Below is the University's Policy on Academic Integrity:

“Students in this course are expected to comply with the University of Pittsburgh School of Arts & Sciences Academic Integrity Code located at www.as.pitt.edu/faculty/policy/integrity.html. Any student suspected of failing to meet the student obligations of the code during the semester will be required to participate in the procedures for adjudication, initiated at the instructor level. This may include, but is not limited to, confiscation of the assignment of any individual suspected of violating the code. A minimum sanction of a zero score for the assignment will be imposed. Violation of the Academic Integrity Code requires the instructor to submit an Academic Integrity Violation Report to the Dean.”

Chapter 7

Assessments - Homework policies

Additional homework will be assigned via TopHat or Canvas. No late assignments will be accepted. Once an assignment closes on TopHat it will no longer be accepted. There will be no makeup homework assignments given. If you miss an assignment it will receive a zero. Your lowest 2 assignment scores will be dropped. This calculation will not be done until after finals.

Chapter 8

Assessments Overview

The primary form of assessment in the course will be regular tests (5).

Other forms of assessment may include, 1. Test revisions (aka Test Reworks), which will count as a form of homework.

1. Quiz questions related to lecture videos, code checkpoints to assure that you can execute key functions, and software checkpoints to assure that you have the proper software involved. 1. Additional homework 1. Practice tests

8.1 Online Test-taking policies

Tests may be administered online.

8.2 Final Exam Policies

The final exam will ... 1, Take place during finals week on the day scheduled by the University 1. May be online or in person 1. Is required 1. Follow the general format of other tests.

8.3 Assessments

As noted previously, your lowest 2 test score will be dropped. Your next lowest test score can be replaced by your final exam, if your final exam score is higher. The final exam can therefore only benefit your score.

Chapter 9

Communication - University email policy

Each student is issued a University e-mail address (username at pitt.edu). This e-mail address may be used by the University for official communication. Students are expected to read e-mail sent to this account on a regular basis. Failure to read and react to University communications in a timely manner does not absolve the student from knowing and complying with the content of the communications.

The University provides an e-mail forwarding service that allows students to read their e-mail via other service providers. Students that choose to forward their e-mail from their pitt.edu address to another address do so at their own risk. If e-mail is lost as a result of forwarding, it does not absolve the student from responding to official communications sent to their University e-mail address.

To forward e-mail sent to your University account, go to <http://accounts.pitt.edu>, log into your account, click on Edit Forwarding Addresses, and follow the instructions on the page. Be sure to log out of your account when you have finished.

For the full E-mail Communication Policy, go to bc.pitt.edu/policies/policy/09/09-10-01.html.

Chapter 10

Course communication - Email and Canvas Messages

Announcements to the whole course will occur typically on Friday afternoons. Other communication related to course administration will also occur via Canvas message.

You can message me on Canvas or email me at nlb24 at pitt.edu.

Please “Comp Bio” as the 1st thing in your email subject line with an informative bit of information as the “...”. Eg “Comp Bio 2: problem accessing TopHat”.

If you use another email service as your primary email (eg GMail) please set your Pitt email to forward there.

For info on forwarding your Pitt email to your personal account follow this link: <https://bit.ly/2Riz7dx>

I try to answer all emails received on weekdays before 5 pm within 24-36 hrs. Emails received after 5 pm will be answered at the earliest the following morning. Emails received on the weekend will be answered Monday.

Please consult this syllabus before asking questions about course policies and the schedule, and refer to relevant information such as URLs, subject headings or dates. Screenshots are super helpful. If the entire answer to your question can be found in the syllabus I will likely respond by saying “This is in the syllabus, Cheers, Dr. B.”*.

Questions relevant to the whole class may be reposted (with identifying details removed).

Chapter 11

Course communication - Overview

11.1 Primary modes of communication

All important details will be sent via Canvas message and mentioned/discussed in class. It is your responsibility to regularly check Canvas messages to stay abreast of the course.

11.2 Secondary modes of communication

11.2.1 Github “issues”

In this class we will be making extensive use of the code management site GitHub. Most communication about homework and other assignments will occur via GitHub's Issues feature. Moreover, to foster discussion we will use GitHub issues to create forums around assignments and you will be required to submit comments.

11.2.2 Microsoft Share Point

Files containing code for this course will be distributed using Microsoft Sharepoint. Sharepoint has messaging features which I may experiment with but will not rely on.

Chapter 12

Catalog description

“This course is designed to give students a broad understanding of how computational approaches can be used to solve problems in biology. Both the biological and computational underpinnings of the methods will be addressed.”

Chapter 13

Course materials overview

Thanks to an Open Educational Resources (OER) grant from The Office of the Provost, there are no books to buy for this course - all course materials are free and will be provided digitally through Canvas.

Reading will take many forms: files containing computer code (scripts), articles and blog posts, as well as textbook. However, we'll be taking a very novel approach to the textbook for this course: **you will actually be building it yourself!**

Details will be provided in-class, but the gist regarding the textbook is: 1. Most weeks you will be provided 1 or 2 chapters of the book in the form of computer files ("scripts"). 1. For homework you will do exercises embedded in the chapters. 1. When the exercises are successfully completed you will be able to convert the files to Web Pages and post them to your own Portofolio of book chapters and other completed activites.

Chapter 14

Course materials - Software

Installation and account creation for all software will be covered in detail during class or via assignments and accompanying videos. If you want to get prepared early for the course you can explore these on your own.

14.1 Primary software

14.1.1 Programming environment: R

We will use the open-source language R for almost everything in this course.

NOTE: If you have used R before that's awesome! **Please update your installation when the class begins** - R goes through regular updates and works best with the newest version.

If you are having trouble installing R an excellent resource is: <https://learningstatisticswithr.com/book/introR.html>

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/.

14.1.2 Integrated development environment

We'll use RStudio as the front-end for working with R.

NOTE: If you've used RStudio before please update it when the course begins.

RStudio: www.rstudio.com/products/rstudio/download/#download

14.1.3 Cloud-based R

As a back-up to running R on your desktop I'll also introduce you to RStudio Cloud. There are also several excellent primers which we'll use.

RStudio cloud: <https://rstudio.cloud/>

14.2 Code sharing

14.2.1 Microsoft SharePoint

R code and other files will be distributed to you via a personal folder on SharePoint, accessed via your OneDrive account.

You will need to “sync” your OneDrive account to your computer desktops to allow me to help you with your work. No code files will be shared via email or Canvas during the course.

NOTE: If you do not have access to a computer you can sync your OneDrive account please contact me so we can workout a system.

14.2.2 GitHub

Most code for this course will be available on GitHub and you will need to create an account for some activities.

GitHub: <https://github.com/>

14.3 R packages

Many R packages (aka libraries) will be used in this course. Below are several key ones.

14.3.1 swirl

swirl is an R package that sets up interactive tutorials in your R console. It works both on desktop computers and RStudio cloud and is an excellent tool for learning and reviewing R.

14.3.2 Bioconductor

Bioconductor is a clearing house and repository for high-grade bioinformatics and computational biology software using R.

Chapter 15

Bibliography

Below are resources that you may wish to use for reference during the course or in the future for review.

15.1 Bioinformatics & computational biology

Pevsner 2015. Bioinformatics and functional genomics. 3rd edition.
Available **FREE** via the Pitt library, though you can only download a few chapters at a time. I do **NOT** recommend buying a hard copy because I am currently only planning on using the first 1/3 of the book. If you want a hard copy and buy it used be sure to get the 3rd edition.

15.2 Phylogenetics

Wiley, EO and BS Liberman. 2011. Phylogenetics: Theory and Practice of Phylogenetic Systematics. 2nd ed.

Available **FREE** via the Pitt library, though you can only download a few chapters at a time. I do **NOT** recommend buying a hard copy because I am currently only planning on using a few chapters. If you want a hard copy and buy it used be sure to get the 2nd edition.

Focal chapters in Wiley and Liberman (2011): * Chapter 4: Tree graphs * Chapter 5: Characters & homology * Chapter 6: Parsimony & parsimony analysis * Chapter 7: Parametric phylogenetics

15.3 Statistics

Navarro, D. 2019. Learning Statistics with R PDF: <https://learningstatisticswithr.com/lsr-0.6.pdf>

HTML: <https://learningstatisticswithr.com/book/>

Focal chapters in Navarro: * “Getting started with R” * “Additional R concepts”
* “Descriptive statistics” * “Pragmatic matters” * “Drawing graphs” * “Basic programming” * “Introduction to probability”

15.4 Reference Books

Cadotte & Davies. Phylogenies in ecology: A guide to concepts and methods. Free from library. Covers the use of phylogenies in community ecology and trait evolution.

Nei and Kumar. Molecular evolution and phylogenetics. Free from library. Contains more mathematical treatments of maximum likelihood etc.

15.5 Reference articles, blogs etc

Altman. 2009. Bioinformatics & computational biology = Same? No. <https://rbaltman.wordpress.com>

Gregory 2008. Understanding evolutionary trees. *Evo Edu Outreach* 1:121-137.

Koonin. 2001. Computational genomics. *Current Biology* 11:R155-R158.

Loman and Watson. 2013. So you want to be a computational biologist? *Nature Biotech.*

Markowetz, F. 2017. All biology is computational biology. *PLoS Biology*.

Nascimento, FF et al. 2017. A biologist’s guide to Bayesian phylogenetic analysis. *Nature Ecology & Evolution*. See GitHub repo and blog posts.

Nussinov et al. 2015. From “What is?” to “What isn’t?” computational biology. *PLoS Computational Biology*.

Visser et al 2015. Speeding up ecological and evolutionary computations in R: Essential of high performanc computing for biologists. *PLoS Computational Biology*.

Chapter 16

Disability Resource & Services

216 William Pitt Union
(412) 648-7890
(412) 383-7355 (TTY)

If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and Disabilities Resources and Services. DRS will verify your disability and determine reasonable accommodations.

Chapter 17

Extra credit policies

No extra credit will be offered in this course.

“Buffer questions” on the mini-tests, test revision, and the final are NOT extra credit.

Chapter 18

File names for homework

All files you create which get submitted for homework need to be in lower case and **strictly** adhere to this format:

“pittemail_lastname_assignmenttitle”

For example:

“nlb24_brouwer_assign_1.rmd”

18.1 Course Goals

The goals of Computational Biology is to introduce you to key concepts related to how computers are used to solve biological problems, and build skills essential to your success working on projects related to computational biology and bioinformatics.

Chapter 19

Grading scale

Note: Students planning to major in Biological Sciences must pass this course with a C (not C- !) or better. Rounding is not done until final grades are computed and is done by computer to 1 decimal place.

Final letter grades are assigned after rounding and is done automatically by Canvas including the decimal value. For example, a score of 91.99% rounds to 92.0% and is an A, but a score of 91.94% rounds to 91.9% and is an A-.

Note: Students planning to major in Computational Biology must pass this course with a C (not C- !) or better.

Chapter 20

Meeting times

The course will meeting in person on Tuesdays and Thursday. There is no recitation for this course.

TopHat will be used during lectures but any questions can be answered outside of class.

Location: 1502 Posvar Hall
Daays: Tuesday & Thursday
Times: 4:00PM - 5:15AM

Chapter 21

Metnal health & wellness

School is hard - please take care of yourself as best you can given the many demands on your time. Diminished mental health, including significant stress, mood changes, excessive anxiety, or problems with sleeping can interfere with your academic performance. You have a support network here at Pitt to help you through challenging times. Acknowledging that you need help, and getting that help, is smart and courageous.

- If you are in an EMERGENCY situation, call 911 or Pitt Police at 412-624-2121.
- If your symptoms are due to FINANCIAL strain, please visit pitt.libguides.com/assistanceresources to see all available University resources.
- If your symptoms are due to strained RELATIONSHIPS, families, or personal crises, please visit the University Counseling Center at www.studentaffairs.pitt.edu/cc/ for free confidential services.
- If your symptoms are strictly related to your COURSE WORK and performance in this course, please contact me.

21.1 Resources:

University Counseling Center: 412-648-7930

Sexual Assault Response: 412-648-7856

RE:SOLVE crisis network: 888-796-8226

Pitt Police: 412-624-2121

Pitt Pantry food bank: <https://www.studentaffairs.pitt.edu/pittserves/the-pitt-pantry/>

Chapter 22

Office hours

22.1 INSTRUCTOR OFFICE HOURS

22.2 UTA OFFICE HOURS

Chapter 23

Point breakdown

Chapter 24

Semester schedule

Chapter 25

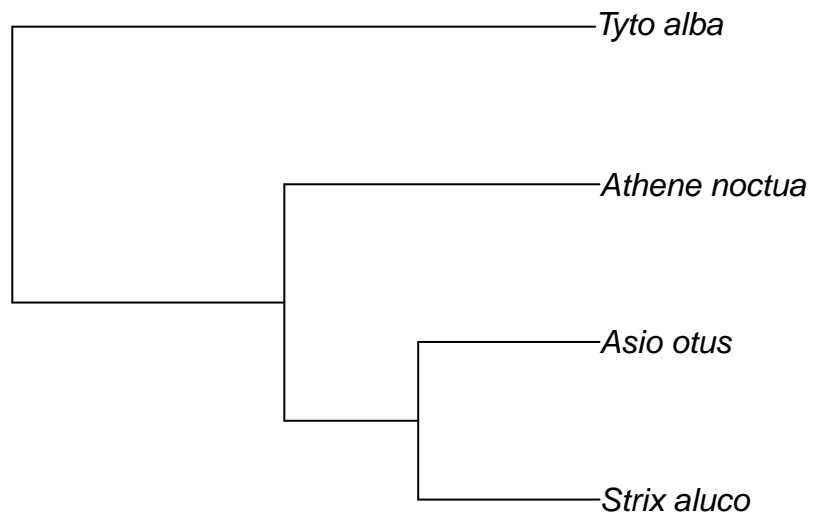
Skills

Key Skills: 1. Understanding R code
2. Data visualization
3. Graph interpretation
4. ...

Key to Computational Biology will be developing skills as a critical practitioner and consumer of science. The course will heavily emphasize the application of skills related to scientific programming, including understanding and modifying R code, plotting data, organizing computational workflows, annotating computational projects. A central set of skill we be learnign to intepret the types of graphs used in computational biology, including alignments, phylogenies, networks, heatmaps, and ordination plots. The course will also introduce basic skills related to statistics, probablity, and mathematical modeling. We will also emphasize critical reading and interpretation of the primary literature.

```
library(ape)

cat("(((Strix_aluco:4.2,Asio_otus:4.2):3.1,",
"Athene_noctua:7.3):6.3,Tyto_alba:13.5);",
file = "ex.tre", sep = "\n")
tree.owls <- read.tree("ex.tre")
plot(tree.owls)
```



```
unlink("ex.tre") # delete the file "ex.tre"
```


Chapter 26

TopHat

26.1 In-class response / Homework system

We will use TopHat in-class and for homework. Please bring a charged TopHat compatible device to all lectures and recitations. See the **Course Policies** section for information related to if you are not able to answer TopHat questions during class due to technical difficulties.

TopHat

Join Code: **209015**

www.tophat.com

When logging into TopHat it will ask you for your university. Typing in “University of Pittsburgh” brings up 3 options: **select the first one** that just says “University of Pittsburgh.”

Chapter 27

Updates to schedule & syllabus

I reserve the right to update the syllabus, schedule, point allocation and all other components of the course as necessary. If changes occur after the first day of class, they will be clearly communicated in class and via email, and a revised syllabus and schedule distributed. After reading the syllabus and completing the “Syllabus treasure hunt” assignment (which will be posted on the first day of class) feel free to contact me with any questions about course policies.

Chapter 28

Appendix: graduate school

If you are considering graduate school you may wish to check out these resource:

So you want to go to graduate school? O'Connel, Tim. "A conversation about grad school" <https://eatmorecookies.wordpress.com/2018/07/17/a-conversation-about-grad-school/>

Carson, W.P. 1999. A primer on how to apply and get admitted to graduate school in ecology and evolutionary biology. *Bulletin of the Ecological Society of America* 80(4):246-250. (PDF here: <https://sites.google.com/site/walterpagecarson/publications>)

Stearns, S. Some modest advices for graduate students *Bulletin of the Ecological Society of America*, 1987 - JSTOR <https://stearnslab.yale.edu/some-modest-advice-graduate-students>

Thoughts on Stearns and Huey Fox, J. 2015. Re-reading Stearns and Hueys some modest advice to gradate students. <https://dynamicecology.wordpress.com/2015/09/08/rereading-stearns-and-hueys-some-modest-advice-to-graduate-students/>

On the publishing process: a tweet by ? <https://twitter.com/dsquintana/status/1154820610750595072?s=20>

Chapter 29

Cheat sheets

RStudio produces a number of cheatsheets; the most useful ones for beginners are below.

Basic R code <http://github.com/rstudio/cheatsheets/raw/master/base-r.pdf>

RStudio <https://github.com/rstudio/cheatsheets/raw/master/rstudio-ide.pdf>

RMarkdown <https://github.com/rstudio/cheatsheets/raw/master/rmarkdown-2.0.pdf> <https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>

Translations of some of these cheat sheets can be found at the bottom of the RStudio page.

Chapter 30

Appendix: Getting involved in (computational) research

Approaching a PI - A guide for undergraduates <https://thefemalescientist.com/guide/madeleine-hann/2053/approaching-a-pi-a-guide-for-undergraduates-in-stem/>

Chapter 31

Shortcuts in R

Programmers are often obsessed with using keyboard shortcuts. You'll see me use some, and you might want to learn them.

The back of this cheatsheet is a 1-page summary of shortcuts <https://github.com/rstudio/cheatsheets/raw/master/rstudio-ide.pdf>

Another summary sheet of Shortcuts in RStudio is here: <https://support.rstudio.com/hc/en-us/articles/200711853-Keybaord-Shortcuts>

A great blog post on shortcuts is here <https://appsilon.com/r-studio-shortcuts-and-tips/>