

BIOL 497/597 - Genomics & Bioinformatics

Syllabus & Timetable

Sven Buerki - Boise State University

2021-01-05

Contents

1	Instructor	2
2	Course delivery type	2
2.1	What technology do you need?	3
3	Class schedule	3
3.1	Meeting days & times	3
3.2	Zoom information	3
3.3	Class location	4
4	Course goal and description	4
5	Course format	6
6	Content of the course	6
7	Course learning outcomes	6
8	Pre-requisite	9
9	Sharing teaching material	9
10	Publications, textbooks & websites supporting this course	9
11	Bioinformatic tools	10
11.1	Installing the main software	10
11.2	Installing markup languages	10
11.3	Installing R packages	10
11.4	RStudio Cheat Sheets	11
11.5	R tutorials	11

12 Journal club	11
13 Grading information	12
13.1 Tests conducted during the class	12
13.2 Mini-reports (TOTAL: 100 points)	12
13.3 Individual oral publication presentation (TOTAL: 50 points)	13
13.4 Lab report (TOTAL: 150 points)	13
13.5 Lab presentations (TOTAL: 50 points)	14
14 Late work policy	14
15 Engagement expectations	14
16 What you can expect of me	14
17 This course was designed with you in mind	15
17.1 This class's community is inclusive	15
17.2 We use each other's preferred names and pronouns	15
17.3 This course is accessible to students with disabilities	16
17.4 For students responsible for children	16
18 Academic integrity	16
19 Student well-being	16
19.1 What do you need?	17
20 Timetable	18
21 References	20
22 Appendix 1	20

1 Instructor

- Name: Sven Buerki
- Office: Science building, office 114 (ground floor).
- Email: svenbuerki@boisestate.edu
- Office hours: By appointment.

2 Course delivery type

To accommodate physical distancing and other safety measures necessitated by the coronavirus pandemic, the university is making changes to its Spring 2021 course schedule. **This course**

will be mostly taught remotely; however we might have opportunities to meet in-person for computing laboratory sessions (see below). This means that courses meet virtually via technology (here Zoom) at scheduled class meeting times (see below).

2.1 What technology do you need?

- You will need reliable access to a computer with an internet connection and a webcam.
- You will need to install a suite of software on your computer ([Zoom](#), [R](#), [RStudio](#)). Please see [the computing tools of reproducible science](#) section below for more details.
- More information on Zoom provided by BSU is available [here](#).
- VPN access using [Cisco AnyConnect Secure Mobility Client](#). This software will facilitate remote access to Linux boxes for bioinformatic analyses. Students using Windows operating systems will have to download Putty or a similar software to remotely access Linux computer using `ssh` protocol. Putty can be downloaded at this URL <http://www.putty.org> (follow instructions to install it on your computer).
- If you have concerns about access to technology needed for your courses this fall, the [Help Desk](#) is the place to begin to discuss your needs. The instructor would also gladly help you with any technology needs.

3 Class schedule

3.1 Meeting days & times

- **Lecture sessions:** Wednesdays from 9:00-10:50 AM.
- **Laboratory sessions:** Fridays from 9:00-10:30 AM.

Please see section [20](#) for tentative timetable for this course.

3.2 Zoom information

Class will take place remotely using the [Zoom](#) platform. A fully-licensed Zoom account is available to all faculty, staff, and students at Boise State. Zoom is a great tool for conducting remote meetings that allows you to see and hear one another, chat, share your computer screen, and more. In our case, we will be taking advantage of the remote control feature allowing users to take control of another participant's screen in a meeting. This is especially useful to help working on and debugging R code. Please see the following Zoom webpages for more details on:

- Remote support session: <https://support.zoom.us/hc/en-us/articles/208072316-Remote-Support-Session>
- Requesting or giving remote control: <https://support.zoom.us/hc/en-us/articles/201362673-Requesting-or-giving-remote-control#heading=h.bdhzh38o7ljq>

Details to join the Zoom meetings and find URLs of class recordings are provided [here](#). Students will be granted access to this document prior to the start of the semester.

3.3 Class location

We might be able to schedule in-person laboratory classes to use the Linux computers. These sessions will take place in the Science building in the computer laboratory: **SCNC149**. Further information will be provided by the instructor.

4 Course goal and description

The goal of this course is to provide students with the theoretical and applied knowledge in genomics and bioinformatics to sequence, assemble and annotate genomes, especially for non-model organisms (see Figure 1).

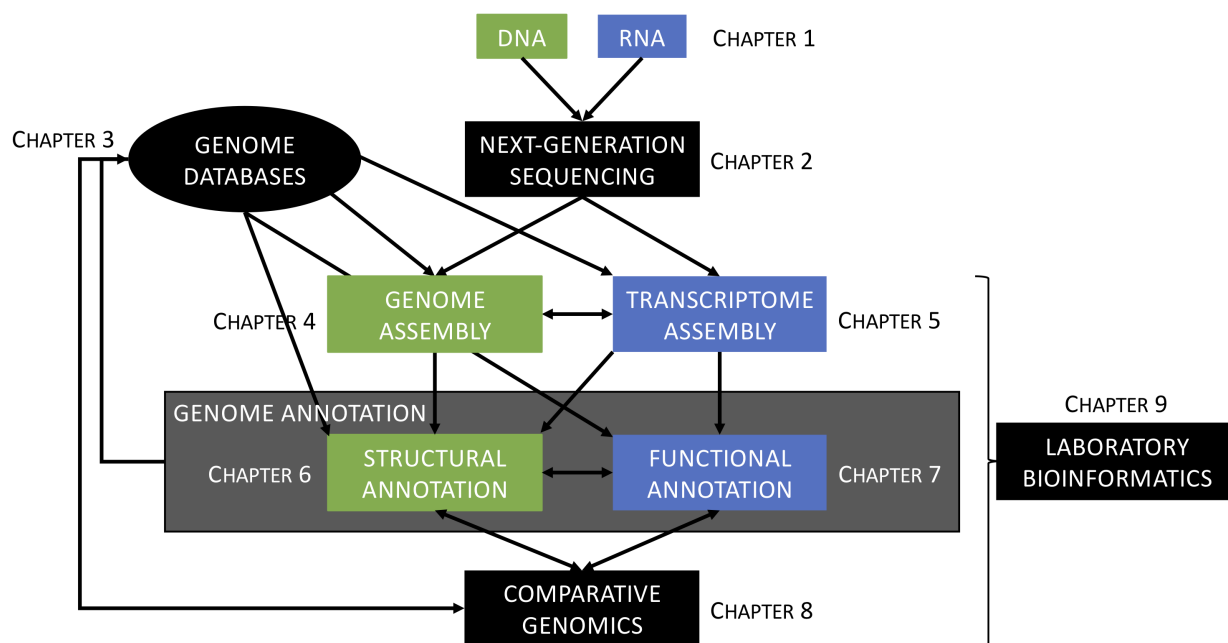


Figure 1: Overview of the structure of the course. See text for more details.

A suite of unix-based open source software will be used to perform genomic analyses (see Figure 2). In addition to learning how to execute programs, students will also be taught procedures to install them therefore gaining critical expertise for their career. Finally, students will produce scientific reports and present their results. Reports together with documents provided by the instructor will provide students with a framework allowing them to analyze their own genomic data.

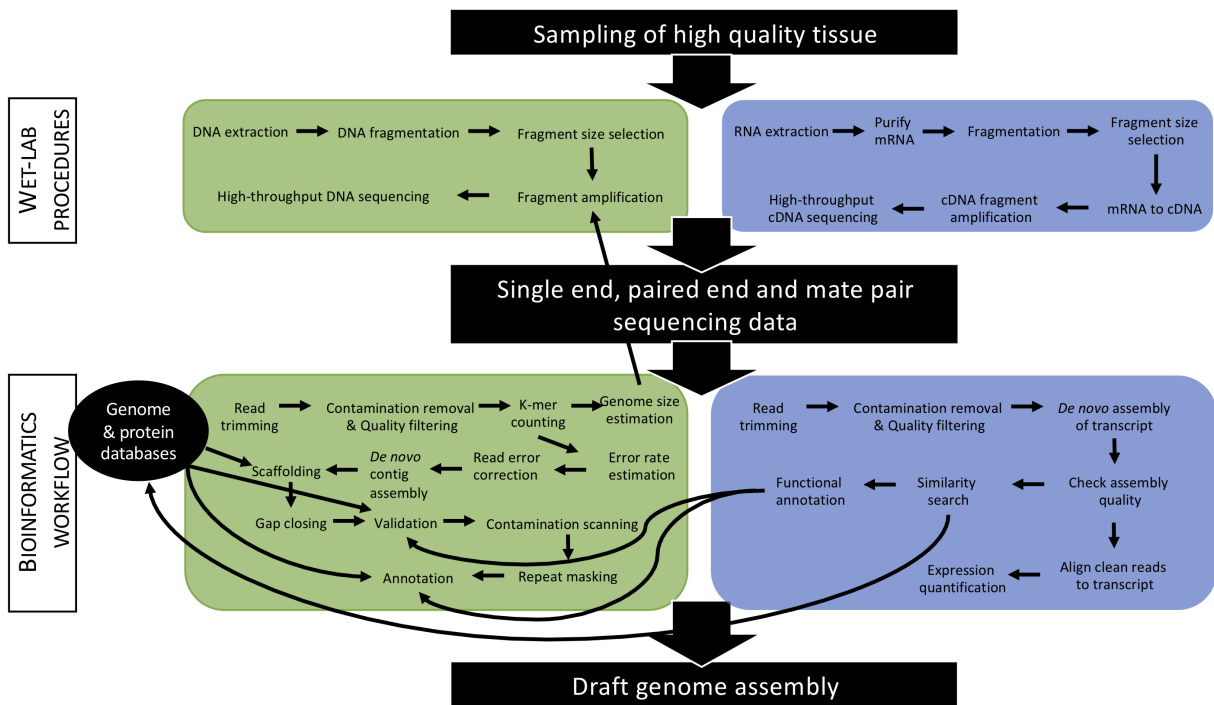


Figure 2: Overview of an example of an approach applied to produce a draft genome assembly. In this course, students will become accustomed with such approach and master some specific key steps.

5 Course format

This course provides a mixture of classic lectures together with more applied, hands-on bioinformatic tutorials (conducted both during lecture and lab sessions). Tutorials are designed to support students in mastering theoretical genomic concepts through their implementations in bioinformatic protocols. As mentioned above, the genomic field heavily relies on bioinformatic expertise, especially related to unix-based software.

We will be working on computers running the Ubuntu Linux operating system. This operating system is widely used in genomics. Using these computers will provide opportunities for students to become familiar with **bash/shell**, **Python** and **R**. Finally, lecture sessions will also serve as a platform to: i) work on graded mini-reports. ii) study genomic literature through a journal club (see below).

While reading scientific papers, we will be focusing on mastering methodological concepts allowing students to assess the quality and reproducibility of published genomic studies.

6 Content of the course

The course is subdivided into nine chapters segregated between lecture and laboratory sessions as follows (see Figure 1):

Lecture sessions:

- **Chapter 1:** DNA & RNA.
- **Chapter 2:** Next-generation sequencing.
- **Chapter 3:** Genome databases.
- **Chapter 4:** Genome assembly.
- **Chapter 5:** Transcriptome assembly.
- **Chapter 6:** Structural annotation.
- **Chapter 8:** Comparative genomics.

Laboratory sessions:

- **Chapter 9:** Applied bioinformatics.

7 Course learning outcomes

Please find below a non-exhaustive list of learning outcomes associated to the chapters presented above (Figure 1):

Chapter 1: DNA & RNA

- Understand the differences between “Genetics” and “Genomics.”
- Know the basic dogma that DNA is transcribed to RNA, which is translated to protein.

- Appreciate the diversity of genome organization in organelles, prokaryotes and eukaryotes.
- Appreciate that organelle genomes have forged eukaryotic chromosomes and that this has to be accounted for in designing genome assembly projects.
- Appreciate that eukaryotic genomes contain extensive repetitive regions of several different kinds.
- Appreciate the diversity of published genomes and their differences in the level of completeness.
- Understand the importance of computer science and bioinformatics in producing the raw sequence data and assembly genomes.
- Understand the challenges in sequencing a genome.

Chapter 2: Next-generation sequencing

- Learn the next-generation sequencing (NGS) jargon underpinning genome assembly and annotation.
- Understand the general principle of genome sequencing (wet-lab) and assembly (bioinformatics workflow) (see Figure 2).
- Be familiar and know the limitations of the following NGS platforms:
 - Illumina.
 - PacBio.
 - Oxford Nanopore.
- Be accustomed to handle NGS data output, especially FASTA and FASTQ files.
- Learn protocols to quantify the quality of the nucleobases generated by NGS platforms using Phred quality score.

Chapter 3: Genome databases

- Understand the importance of computer science and bioinformatics in:
 - Producing raw data.
 - Creating databases in molecular biology.
 - Archiving and careful curation of the data.
 - Distributing them via the Internet.
 - Creating information-retrieval tools to allow effective mining of the data for research application.
- Gain knowledge on major molecular biology databases, which are key to the assembly and annotation of genomes:
 - Nucleic acid sequences databases.
 - Protein sequences databases.
 - Gene ontology databases.
 - Metabolic pathways databases.
 - Specialized annotated genome portals.
- Learn protocols to query genomic information and remotely download files from GenBank.

Chapter 4: Genome assembly

- Understand key steps involved in producing a genome assembly and best strategies to

get there.

- Be able to perform reads quality checks as follows:
 - Read quality assessments.
 - * Read trimming.
 - * Contaminant screening.
 - * k-mer counting.
 - * Genome size estimation.
 - * Read error correction.
- Understand the use of de Bruijn graphs for *de novo* genome assembly.
- Gain an overview of current genome assemblers and their specificity.
- Study the benefit of long reads in closing gaps and correcting draft genome assemblies.

Chapter 5: Transcriptome assembly

- Learn about the types of RNA molecules in cell.
- Study the link between the transcriptome and proteome via the genetic code.
- Understand key steps involved in producing a transcriptome assembly and best strategies to get there. We will mostly compare referenced-based vs. *de novo* strategies, but also look at the “combined strategy.”
- Gain an overview of current transcriptome assemblers and their specificity.

Chapter 6: Structural annotation

- Learn what we exactly mean by the term “genome annotation.”
- Understand key steps involved in genome annotation and best strategies to get there.
- Be familiar with the following key stages of the structural genome annotation process:
 - Repeat identification.
 - Evidence alignment.
 - *Ab initio* gene prediction.
 - Evidence-driven gene prediction.
 - Introduction to available structural annotation software.

Chapter 7: Functional annotation

- Be aware of the challenges to obtain accurate data on gene functions.
- Study and compare available pipelines to perform automated annotations of genomes.
- Review approaches to assess annotation quality.
- Be familiar with functional annotation software.
- Overview of the procedure to submit annotated genomes to public databases.

Chapter 8: Comparative genomics

- Learn to read and present a scientific paper reporting new genomic or transcriptomic data.
- Complement expertise by reading and discussing scientific publications during the mini journal club sessions (see below for more details).

Chapter 9: Applied Bioinformatics (corresponding to material taught during laboratory sessions) In this chapter, we will be using data produced by the sagebrush [GEM3](#) project to

learn protocols to:

- Clean Illumina reads.
- Assemble draft genome.
- Mine scaffolds in an Illumina draft genome containing target genes.
- Validate mining by locating Open Reading Frames (ORFs), reconstructing protein sequences and inferring gene phylogenetic trees.
- Annotate validated genes.

Please see section 20 for tentative timetable associated to these chapters.

8 Pre-requisite

BIOL 310 GENETICS (3-0-3)(F,S,SU). A study of the principles of genetics as they relate to living organisms. PREREQ: BIOL 191-192 or BIOL 191 and BIOL 320. PRE/COREQ: CHEM 301 or 307.

9 Sharing teaching material

Teaching material will be shared with students on GitHub and Google drive. Google drive will mostly be used to share sensitive material (e.g. IP addresses and passwords of lab computers, Zoom details and recordings), post publications and to allow students sharing their assignments.

The URL to the shared Google drive is (the instructor will grant you access to this folder prior to the start of the semester):

- <https://drive.google.com/drive/folders/1BFZFJeR8tW8N7QLfngqIVbjqEkDuIyWj?usp=sharing>

10 Publications, textbooks & websites supporting this course

The reading material at the basis of this class is composed of a mixture of publications and chapters from four textbooks (Dale et al., 2012; deSalle and Rosenfeld, 2013; Brown, 2017; Lesk, 2017). The instructor has copies of these textbooks and students are welcome to consult them at anytime.

In addition, there is also a plethora of website devoted to genomic and transcriptomic data. Please find a selection below:

- [Next-generation sequencing online course.](#)

- [Genomics and comparative genomics website](#).

11 Bioinformatic tools

Although we will be using software installed on the Linux computers, the instructor is advising students to install the following programs on their personal computers:

- The **R** statistical language that will allow you to gather data and analyze it.
- The **LaTeX** and **Markdown** markup languages that you can use to create documents (slideshows, articles, books, webpages) for presenting your findings.
- The *knitr* and *rmarkdown* **packages** for R and other tools, including **command-line shell programs** like GNU Make and Git version control, for dynamically tidying your data gathering, analysis, and presentation documents together so that they can be easily reproduced.
- **RStudio**, a program that brings all of these tools together in one place.

11.1 Installing the main software

As shown above, **R** and **RStudio** are at the core of this course and will have to be installed on your computers. This can be easily done by downloading the software from the following websites:

- **R**: <https://www.r-project.org>
- **RStudio**: <https://www.rstudio.com/products/rstudio/download/>

The download webpages for these software have comprehensive information on how to install them, so please refer to those pages for more information.

11.2 Installing markup languages

If you are planning to create LaTeX documents, you will need to install a Tex distribution. Please refer to this website for more details: <https://www.latex-project.org/get/>

If you want to create Markdown documents you can separately install the *rmarkdown* package in R (see [below](#) for more details).

11.3 Installing R packages

We will be using a number of R packages especially designed to support reproducible research. Many of those packages are not included in the default R installation and will need to be installed separately. To install key packages used in class, copy the following code and paste it into your R console:

```
install.packages(c("brew", "countrycode", "devtools", "dplyr", "ggplot2", "googleVis",  
  "knitr", "rmarkdown", "tidyr", "xtable"))
```

Once you enter this code, you may be asked to select a CRAN “mirror” to download the packages from. Simply select the mirror closest to you.

Finally, it is highly likely that we will have to install additional packages. In this case, you can simply install it by using the same R function `install.packages()` or by using RStudio as follows: Select “Tools” -> “Install Packages ...” and then type the name of the package in the window (make sure to tick the “Install dependencies” box).

11.4 RStudio Cheat Sheets

RStudio provides a suite of cheat sheets that can be accessed by going to the “Help” menu and selecting “Cheatsheets.”

Five cheat sheets are especially relevant to chapters taught in this course:

- *RStudio IDE : Cheat sheet*
- *Data Manipulation with dplyr, tidyr*
- *Data Visualization with ggplot2*
- *R Markdown Cheat Sheet*
- *R Markdown Reference Guide*

11.5 R tutorials

Please find below two documents providing a comprehensive introduction to R:

- R for beginners (a tutorial by Emmanuel Paradis): https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf
- An introduction to R: <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>

12 Journal club

The aim of the journal club is to train students to read scientific publications related to genomic data. Journal clubs will also provide opportunities for students to learn how to present their scientific publications.

Students are expected to have read the scientific publications before coming to class. The instructor will start by briefly summarizing the methodology and main outcomes of the study and the rest of the class will be devoted to general discussion (please come with some questions and have a critical, but positive mid set).

Please find below the list of publications that we will discuss in class (see section 20 for information on dates):

- **Publication 1:** Li et al. (2009) published a draft genome of the giant panda and screened for genes responsible for the detoxification of cyanide produced by bamboos.
- **Publication 2:** Ming et al. (2015) published the draft genome of the pineapple and investigated the origin of CAM photosynthesis in this lineage.

If you are not familiar with approaches to read scientific publications, please come and seek advice. For undergraduate students, there is a class BIOL 306 that you could take next semester. Please find below the details of this class:

BIOL 306 COMMUNICATION IN THE BIOLOGICAL SCIENCES (3-0-3)(F,S)(CID) Development of written and oral communication skills necessary for future careers in the biological sciences. Skills include summarizing and evaluating scientific research, and communicating scientific information to targeted audiences. PREREQ: BIOL 304 and ENGL 101, 102.

13 Grading information

13.1 Tests conducted during the class

- 3 mini-reports (2x 25 points and 1x 50 points, TOTAL: 100 points).
- 1 individual oral publication presentation (50 points).
- 1 lab report (150 points).
- 1 lab presentation (50 points).

Exams are summing to a total of **350 points** and grading will follow the scale presented in Table 1. Please see section 20 for information on dates when exams are due.

13.2 Mini-reports (TOTAL: 100 points)

To further gain expertise in this field, students will be producing three mini-reports on the following topics (see Figure 1):

- **Mini-Report 1:** Sequencing technologies (25 points; this task is related to Chapter 2).
- **Mini-Report 2:** Molecular biology databases (25 points; this task is related to Chapter 3).
- **Mini-Report 3:** Use knowledge from the previous mini-reports to review plant genomes available on GenBank database (50 points). Here, we will be using R to mine and analyze data.

Time will be allocated in class to work on these mini-reports, but the instructor expects students to complete those on their own time.

Table 1: Grading scale applied in this course.

Percentage	Grade
100-98	A+
97.9-93	A
92.9-90	A-
89.9-88	B+
87.9-83	B
82.9-80	B-
79.9-78	C+
77.9-73	C
72.9-70	C-
69.9-68	D+
67.9-60	D
59.9-0	F

13.3 Individual oral publication presentation (TOTAL: 50 points)

Each student will have to present a scientific publication at the end of the semester related to material presented in Chapter 8. Time will be allocated to select a paper and work on presentations, but the instructor strongly advise students to start this endeavor sooner. **The presentation should not exceed 4 minutes (plus 2 minutes for questions) and contain a maximum of 5 slides.**

13.4 Lab report (TOTAL: 150 points)

Students will be working individually or in groups to produce a report associated to the following research question:

- What Aquaporin genes are encoded in the sagebrush genome (*Artemisia tridentata*) and what are their functions?

Lab reports will be formal documents written in the form of scientific manuscripts (Introduction, Material & Methods, Results, Discussion and References sections).

The instructor does not fix a minimum number of words or pages, but urges students to be succinct and avoid redundancy.

Please make sure that you present all evidence allowing an external reviewer to replicate your analyses. In order to achieve this goal and support reproducible science, students will have to provide associated data and code underpinning their research. The instructor will provide guidance on these protocols during lab sessions.

Please note that if students are working in group then each student should contribute equally to the project. If participation is not judge adequate within a group, then the instructor will

grade students individually.

13.5 Lab presentations (TOTAL: 50 points)

Students will use their lab reports to prepare 10 minutes presentations (+ 5 minutes for questions) taking place during the final weeks of the semester.

The instructor expects each student to take part to the presentation and to be involved in answering questions.

Presentations will have to be deposited in advance onto a shared Google drive folder or be sent to the instructor prior to the presentation.

14 Late work policy

The instructor expects students to deliver their assignments on time and set enough time aside to work on their projects (see above for more details). However, if you have any issues preventing completion of your work on time, please contact the instructor as soon as possible to find common solutions.

15 Engagement expectations

The remote teaching mode of this course makes it harder for the instructor to take attendance. The instructor is expecting students to attend classes (please join on time and for the full duration of the class) and actively engage by asking questions and giving feedback on teaching material and course content. This course was designed to help students implementing a reproducible approach to their thesis projects. If you are judging that additional content should be covered in class, please contact the instructor. The instructor will do his very best to obtain information or seek support from colleagues to cover the requested material. Finally, the instructor is aware that COVID could potentially impact yourselves and your families. In the case that you have any issues attending class, please contact the instructor by email (svenbuerki@boisestate.edu) asap and see below for more details.

16 What you can expect of me

The instructor will be prepared for class, on time and not leave early. He will also be respectful of you and your opinions. Overall, the instructor wants to foster a kind and respectful class environment where all students can express themselves and share their opinions. This means that meaningful and constructive dialogue is encouraged in this class and it requires a degree of mutual respect, willingness to listen, and tolerance of opposing points of view. Respect

for individual differences and alternative viewpoints will be maintained at all times in this class. One's words and use of language should be temperate and within acceptable bounds of civility and decency. Finally, the instructor will reply to emails and grade tests as soon as possible (and provide positive criticism) to allow students mastering the material presented in class.

17 This course was designed with you in mind

The instructor has developed this course to provide a welcoming environment and effective, equitable learning experience for all students. If you encounter barriers in this course, please bring them to my attention so that I may work to address them.

17.1 This class's community is inclusive

Students in this class represent a rich variety of backgrounds and perspectives. The Biological Sciences department is committed to providing an atmosphere for learning that respects diversity and creates inclusive environments in our courses. While working together to build this community, we ask all members to:

- Share their unique experiences, values, and beliefs, if comfortable doing so.
- Listen deeply to one another.
- Honor the uniqueness of their peers.
- Appreciate the opportunity we have to learn from each other in this community.
- Use this opportunity together to discuss ways in which we can create an inclusive environment in this course and across the campus community.
- Recognize opportunities to invite a community member to exhibit more inclusive, equitable speech or behavior—and then also invite them into further conversation. We also expect community members to respond with gratitude and to take a moment of reflection when they receive such an invitation, rather than react immediately from defensiveness.
- Keep confidential any discussions that the community has of a personal (or professional) nature, unless the speaker has given explicit permission to share what they have said.
- Respect the right of students to be addressed and referred to by the names and pronouns that correspond to their gender identities, including the use of non-binary pronouns.

17.2 We use each other's preferred names and pronouns

Please let the instructor know of your preferred or adopted name and gender pronoun(s), and he will make those changes to his own records and address you that way in all cases.

To change to a preferred name so that it displays on all BSU sites, including Blackboard and the course roster, contact the Registrar's Office at (208) 426-4249. Note that only a legal

name change can alter your name on BSU official and legal documents (e.g., your transcript).

17.3 This course is accessible to students with disabilities

The instructor recognizes that navigating your education and life can often be more difficult if you have disabilities. I want you to achieve at your highest capacity in this class. If you have a disability, the instructor needs to know if you encounter inequitable opportunities in this course related to: - Accessing and understanding course materials. - Engaging with course materials and other students in the course. - Demonstrating your skills and knowledge on assignments and exams.

If you have a documented disability, you may be eligible for accommodations in all of your courses. To learn more, make an appointment with the university's [Educational Access Center](#).

17.4 For students responsible for children

The instructor recognizes the unique challenges that can arise for students who are also parents or guardians of children. If you have any specific needs related to this topic, please contact the instructor asap.

18 Academic integrity

To create a welcoming, engaging, and effective learning environment, the instructor expects all of us to exhibit behavior that reflects Boise State's [Statement of Shared Values](#). The Shared Values emphasize academic excellence, caring, citizenship, fairness, respect, responsibility, and trustworthiness. In keeping with these values, the instructor expects students in this course to uphold the standards outlined in the Boise State University [Student Code of Conduct](#).

19 Student well-being

If you are struggling for any reason (COVID, relationship, family, or life's stresses) and believe these may impact your performance in the course, the instructor is encouraging you to contact the Dean of Students at (208) 426-1527 or email deanofstudents@boisestate.edu for support. Additionally, if you are comfortable doing so, please reach out to me and I will provide any resources or accommodations that I can. If you notice a significant change in your mood, sleep, feelings of hopelessness or a lack of self worth, consider connecting immediately with Counseling Services (1529 Belmont Street, Norco Building) at (208) 426-1459 or email healthservices@boisestate.edu.

19.1 What do you need?

The university has many resources designed to support you as a learner and human being. Among these are:

- [Albertsons Library](#) provides a treasure trove of physical and electronic resources.
 - As you enter the library, straight ahead you'll find the Reference Desk, where librarians can help you find the information and resources you need.
 - The Circulation Desk lets students borrow various technologies.
 - The MakerLab on the second floor offers tools for student use, and there are friendly staff in the MakerLab to help you learn how.
- The [Writing Center](#) offers individual consultations tailored to your needs, including making sense of writing assignment instructions, brainstorming, crafting a thesis, organizing an essay, revisions, citations, and more.
- [Counseling Services](#) helps you tap into your strengths and find resources to deal more effectively with concerns that impact your pursuit of personal and academic goals. It emphasizes prevention and early detection and provides a broad spectrum of short-term counseling, consultative, evaluative, teaching, and training functions. Counseling staff consists of licensed counselors, psychologists, and closely supervised trainees/post-graduate interns.
- The [Gender Equity Center](#) provides specialized, brief, no-cost, confidential support for campus members affected by:
 - Identity-related harassment.
 - Sexual harassment, sexual assault, unhealthy relationships, and stalking.
 - Life experiences related to gender, sexual orientation, and allyship.
 - Life transitions and crisis.
 - Personal, financial, or academic roadblocks.
- [Multicultural Student Services](#) works with all students interested in equity and social justice, students from different cultures, countries and ethnicities. The staff understands cultural differences and perspectives, and knows what it's like living in a dominant culture. The training and the programs Multicultural Student Services provides can help you navigate through political correctness and into respectful dialogue. This center also provides:
 - Support and guidance for planning cultural festivals and events.
 - In-center tutoring and studying lounge.
 - Volunteer opportunities.
 - Mentoring and Training.
 - Programs and events.
- Food assistance: If you are hungry and cannot afford to purchase food, the campus has some resources to help you. You can visit the [campus food pantry](#) or [get free meals in the campus dining hall](#).

20 Timetable

A tentative timetable for the Spring semester 2021 is presented in Table 2.

Table 2: Tentative timetable for the Genomics and Bioinformatics course during the Spring semester 2021.

WeekID	Date	Weekday	Topic	Deadline
1	2021-01-13	Wednesday	Syllabus	
1	2021-01-15	Friday	Chapter 1	
2	2021-01-20	Wednesday	Chapter 2	
2	2021-01-22	Friday	Chapter 2	
3	2021-01-27	Wednesday	Mini-Report 1	
3	2021-01-29	Friday	Journal Club	
4	2021-02-03	Wednesday	Chapter 3	
4	2021-02-05	Friday	Sagebrush Genome Project	Turn-in Mini-Report 1
5	2021-02-10	Wednesday	Mini-Report 2	
5	2021-02-12	Friday	Intro. to bioinformatic pipeline used to mine genome	
6	2021-02-17	Wednesday	Tutorial ssh protocol & unix commands	
6	2021-02-19	Friday	Mining genome	Turn-in Mini-Report 2
7	2021-02-24	Wednesday	Chapter 4 (QCs)	
7	2021-02-26	Friday	Mining genome	
8	2021-03-03	Wednesday	Chapter 4 (QCs)	
8	2021-03-05	Friday	Mining genome	
9	2021-03-10	Wednesday	Mini-Report 3	
9	2021-03-12	Friday	Mining genome	Submit preliminary lab reports for comment
10	2021-03-17	Wednesday	Chapter 4 (De novo analysis)	
10	2021-03-19	Friday	Mining genome	
11	2021-03-24	Wednesday	Journal Club	
11	2021-03-26	Friday	Mining genome	Turn-in Mini-Report 3
12	2021-03-31	Wednesday	Chapter 4 (De novo analysis)	Submit selected publications for approval
12	2021-04-02	Friday	Chapter 5 & 6	
13	2021-04-07	Wednesday	Chapter 7	
13	2021-04-09	Friday	Mining genome	
15	2021-04-21	Wednesday	Present publications (on comparative genomics)	
15	2021-04-23	Friday	Mining genome - Presentations	
16	2021-04-28	Wednesday	Present publications (on comparative genomics)	
16	2021-04-30	Friday	Mining genome - Presentations	Turn-in lab reports

21 References

- BROWN, T.A. 2017. Genomes 4. Garland Science.
- DALE, J.W., M. VON SCHANTZ, and N. PLANT. 2012. From genes to genomes: Concepts and applications of DNA technology, 3rd edition. Wiley.
- DESALLE, R., and J. ROSENFELD. 2013. Phylogenomics: A primer 1st edition. Garland Science.
- LESK, A. 2017. Introduction to genomics, third edition. Oxford University Press.
- LI, R., W. FAN, G. TIAN, H. ZHU, L. HE, J. CAI, Q. HUANG, et al. 2009. The sequence and de novo assembly of the giant panda genome. *Nature* 463: 311 EP–. Available at: <http://dx.doi.org/10.1038/nature08696>.
- MING, R., R. VANBUREN, C.M. WAI, H. TANG, M.C. SCHATZ, J.E. BOWERS, E. LYONS, et al. 2015. The pineapple genome and the evolution of CAM photosynthesis. *Nature Genetics* 47: 1435.

22 Appendix 1

Citations of all R packages used to generate this report.

- [1] J. Allaire, Y. Xie, J. McPherson, et al. *rmarkdown: Dynamic Documents for R*. R package version 2.6. 2020. <URL: <https://github.com/rstudio/rmarkdown>>.
- [2] C. Boettiger. *knitcitations: Citations for Knitr Markdown Files*. R package version 1.0.10. 2019. <URL: <https://github.com/cboettig/knitcitations>>.
- [3] M. C. Koochafkan. *kfigr: Integrated Code Chunk Anchoring and Referencing for R Markdown Documents*. R package version 1.2. 2015. <URL: <https://github.com/mkoochafkan/kfigr>>.
- [4] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2019. <URL: <https://www.R-project.org/>>.
- [5] H. Wickham and J. Bryan. *usethis: Automate Package and Project Setup*. R package version 1.6.3. 2020. <URL: <https://CRAN.R-project.org/package=usethis>>.
- [6] H. Wickham, R. François, L. Henry, et al. *dplyr: A Grammar of Data Manipulation*. R package version 1.0.2. 2020. <URL: <https://CRAN.R-project.org/package=dplyr>>.
- [7] H. Wickham, J. Hester, and W. Chang. *devtools: Tools to Make Developing R Packages Easier*. R package version 2.3.2. 2020. <URL: <https://CRAN.R-project.org/package=devtools>>.
- [8] Y. Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*. ISBN 978-1138700109. Boca Raton, Florida: Chapman and Hall/CRC, 2016. <URL: <https://github.com/rstudio/bookdown>>.

- [9] Y. Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.21. 2020. <URL: <https://github.com/rstudio/bookdown>>.
- [10] Y. Xie. *Dynamic Documents with R and knitr*. 2nd. ISBN 978-1498716963. Boca Raton, Florida: Chapman and Hall/CRC, 2015. <URL: <https://yihui.org/knitr/>>.
- [11] Y. Xie. *formatR: Format R Code Automatically*. R package version 1.7. 2019. <URL: <https://github.com/yihui/formatR>>.
- [12] Y. Xie. “knitr: A Comprehensive Tool for Reproducible Research in R.” In: *Implementing Reproducible Computational Research*. Ed. by V. Stodden, F. Leisch and R. D. Peng. ISBN 978-1466561595. Chapman and Hall/CRC, 2014. <URL: <http://www.crcpress.com/product/isbn/9781466561595>>.
- [13] Y. Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.30. 2020. <URL: <https://yihui.org/knitr/>>.
- [14] Y. Xie, J. Allaire, and G. Golemund. *R Markdown: The Definitive Guide*. ISBN 9781138359338. Boca Raton, Florida: Chapman and Hall/CRC, 2018. <URL: <https://bookdown.org/yihui/rmarkdown>>.
- [15] Y. Xie, C. Dervieux, and E. Riederer. *R Markdown Cookbook*. ISBN 9780367563837. Boca Raton, Florida: Chapman and Hall/CRC, 2020. <URL: <https://bookdown.org/yihui/rmarkdown-cookbook>>.
- [16] H. Zhu. *kableExtra: Construct Complex Table with kable and Pipe Syntax*. R package version 1.2.1. 2020. <URL: <https://CRAN.R-project.org/package=kableExtra>>.