Syllabus for the course

Foundations of Quantitative Reasoning (3 credits)

Spring 2024 (LIVING/LEARN D D107)

BIOL 6990C (CRN 15309)

Faculty information:
Dr. Joaquin C. B. Nunez
Henderson-Harris Fellow *in* Biology
Address:
Marsh Life Science (MLS),
Room 337B,
109 Carrigan Drive
Burlington, Vermont 05405

Email:
joaquin.nunez@uvm.edu
Course Website:
https://www.jcbnunez.org/questfqr
Brightspace Website:
BIOL6990C: Found. of Quantit. Reaso..
https://brightspace.uvm.edu/d2l/home/65684
Office Hours:
Mondays 3-5 PM (and by appointment)

GitHub repo: https://github.com/Jcbnunez/uvm-fqr

Course description

Foundations of Quantitative Reasoning (FQR) is a graduate-level class designed to provide graduate students with the knowledge and competencies needed to tackle complex problems in data analysis using first principles of evolutionary theory. As part of this process, students will work to develop a comprehensive analysis toolbox to conduct highly reproducible quantitative research in high-performance computation (HPC) environments. These topics will be pivotal to ensure success in the student's graduate careers in data-intensive fields. This course is open to all graduate students, and may also be taken by select, highly advanced, undergraduates with permission from the instructor.

BIOL 6990C (CRN 15309) is an advanced course that will focus on tackling a wide gamut of theoretical, analytical, and operational issues in quantitative research using **first principles** (i.e., "the foundations") of evolutionary theory. Accordingly, most analyses will be done using *simulated* data. This choice of simulated data is deliberate and key to the goals of the class since students will focus on solving generalizable issues in data analysis and hypothesis testing without the need to worry about the idiosyncratic issues of *real* datasets (i.e., the output of sequencing platforms like Illumina, singular, Oxford nanopore, PacBio, *etc*.). Yet, the course expects that, having mastered the Foundations of analyses using simulations, students will be well-equipped to extract the real signal from the noise inherent in their analyses.

Some thoughts to share about the class:

Is this class a "data science" class? This class uses "big data" to achieve its goals and uses principles of data science, as such, it may be thought as a "data science class". Yet, we are also interested in the biological interpretation of said "big data". Accordingly, our class will have a very heavy emphasis on understanding evolutionary theory to "derive meaning" from the data.

What are the expectations of the class? As a graduate class, the expectation is that all students taking the course are deeply invested in their own professional development, and are fully self-motivated.

***I have never done coding before, how should I prepare*?** Do not worry, the expectation of the class is that students have little coding experience as they enter the class. The class will provide you with all the tools needed to achieve competence in HPC environments and workflows. It is key, however, that you remain self-motivated and, should roadblocks arise, contact Dr. Nunez ASAP, to avoid falling behind in the course.

***I have ample experience coding, how can I get value out of this class*?** On occasion, students taking the class will have ample coding experience and may be puzzled as to "how to extract value from class time". I submit to these folks that they may get great value from the course using a *four*-prong approach:

1. *Polish* your knowledge. While you may have ample coding experience, this class will offer you the opportunity to polish your knowledge and expand your toolbox using software that you may not have used before, or tricks that other folks have developed over years of practice.
2. Become a *peer mentor*. You can use your knowledge to help others in the class. Knowing how to teach complex skills to others is a true sign of mastery over a topic. You can practice your mentoring skills by helping other folks master HPC skills. If this is something that interests you talk to Dr. Nunez about being placed in a group of beginners. Also, and most crucially, be kind to others in the class.
3. *Bring your own data*. It is possible to plug your own data into the class in order to advance research goals. Talk to Dr. Nunez about this early on during the class.
4. This is a great opportunity to read papers and *practice your presentation and proposal writing* skills.

It is my goal for this class to provide value to ***all*** students regardless of starting skill level.

Requirements

To enroll in this class you must:

- Be a graduate student
- Have an interest in the class

  or
- Permission form the instructor

For undergraduates: instructor permission. Interested students may email joaquin.nunez@uvm.edu

You will need a computer able to connect to UVM's internet as well as UVM's VPN. There are no technical requirements for this computer since all of the course will take place in the VACC (https://www.uvm.edu/vacc). If you do not own a computer, please contact the college or your department, they may be able to procure a loaner for you.

Course goals and outcomes

By completing this course you will be able to:

1. ***Understand*** *a "common language" and work in "common languages"*
    a. Students will be able to understand the basic concepts of evolutionary theory. Moreover, these concepts will serve as a common language to collaborate across disciplines.
    b. Students will be able to design and deploy analytical workflows using the most widely used programming languages in our field (Unix, R, Python, and SLiM which is like R), in an HPC environment using reproducible principles.

2. ***Develop*** *testable hypotheses informed by "first principles" and simulation*
    a. students will be able to scale up the first principles of evolution in order to create testable hypotheses using simulation data to inform their work.

3. ***Integrate*** *hypotheses into a research project with defined goals.*
    a. Students will be able to craft small written documents that summarize the proposed work.
    b. Students will be able to give presentations on proposals and receive feedback from peers.

Textbook and Readings

This class has **no** textbook requirement. All readings will be from the primary literature.

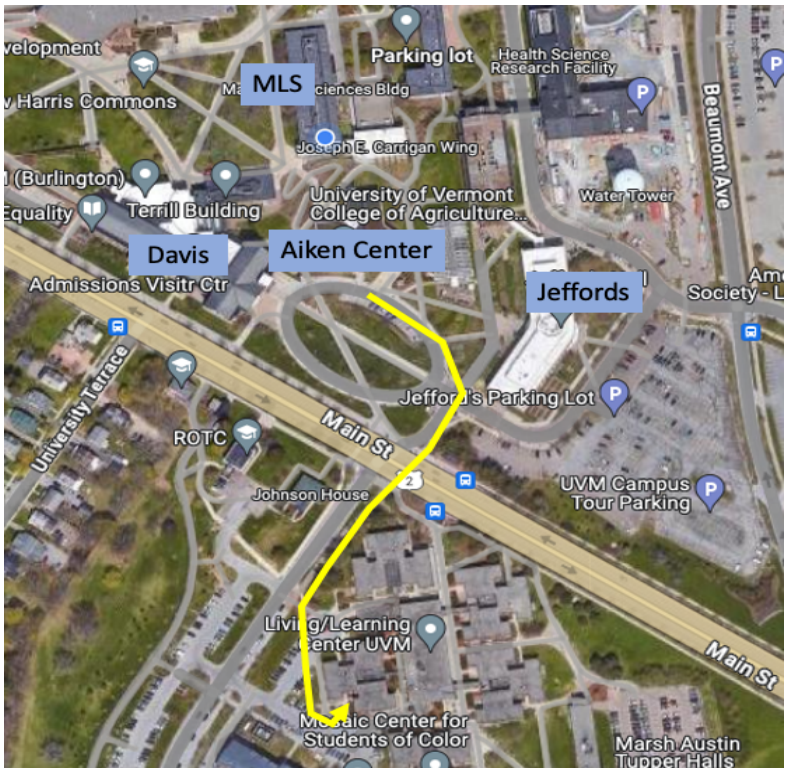
Class Meeting

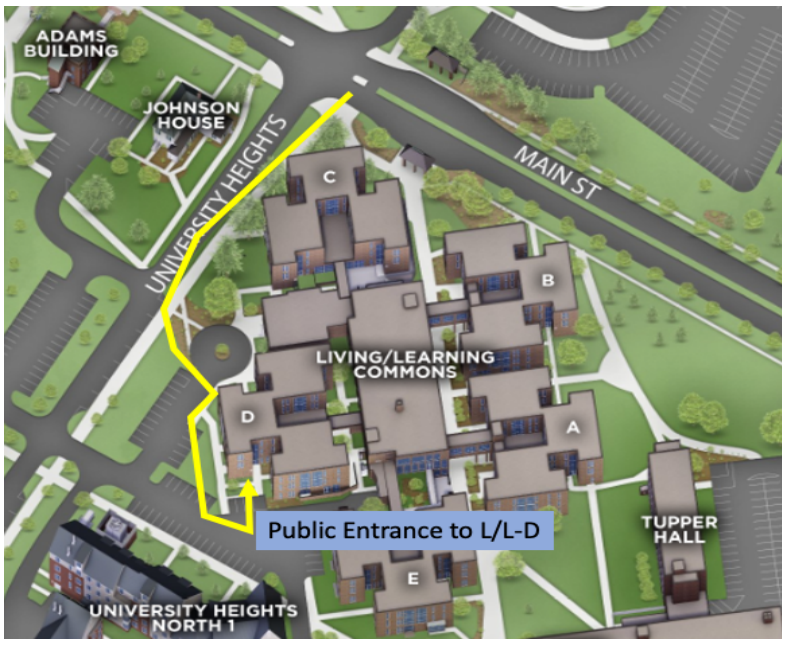
Tuesday, Thursday: 10:05 - 11:20
Location LIVING/LEARN D D107

The living-learning center L/L is located at the interception of Main Street and University heights. Its address is:

633 Main St, Burlington, VT 05405
Living/Learning Center UVM, Address
Please refer to the map below in order to see how to get to LIVING/LEARN D D107

Walking from the circle in font of the Davis center, Aiken and Jeffords, cross main street, and continue walking parallel to university heights road. **<u>The entrance is in front of university heights.</u>**





Point Ranges	Letter Grade Equivalence
>97	A+
93-96.9	A
90-92.9	A-
87-89.9	B+
83-87.9	B
80-82.9	B-
77-79.9	C+
73-76.9	C
70-72.9	C-

Grading Policy

Passing grades: As per UVM's grading guidelines, graduate students do not receive grades equivalent to "D". Accordingly, grades in this class will vary from 70% (C) to >90% (A).

Failing grades: Any graduate student at risk of obtaining a grade lower than a "C" must see Dr. Nunez **immediately**.

Special cases: Undergraduates taking this class will be subject to the same rules as graduate students. No "D" grades will be given in this class. Agreement to this rule is implicit upon accepting the instructor override to enroll in BIOL 6990C (CRN 15309).

Assignments

The grade for this class will be derived from the following activities:

Participation, paper discussion, and attendance: 5 pts.
All students are expected to attend class regularly and participate actively. Throughout the class, we will read and discuss, in depth, a number of papers. Individuals and or groups in the class are expected to be the discussion leader for at least one paper (sometimes two; depending on class enrollment). While everyone in the class is expected to read the paper, the discussion leader(s) is/are expected *to create 3 questions to be discussed by the class, as a whole.*

Bi-Weekly reflections: 15 pts.
Once every two weeks, I expect students to submit a succinct report (max 2 pages), summarizing their progress in the course. These reports should contain a small summary of what new skills were learned in the past two weeks, small examples of results obtained by work conducted, as well as a reflection on what challenges a student may be experiencing with code or concepts, as well as what previous challenges have now been successfully tackled.

Short Quizzes and knowledge spot-checks: 10 pts. These are weekly short quizzes on basic concepts covered in class. These quizzes are available in brigthspace.

Proposal: 30 pts.
As individuals, or groups (depending on the instructor instructions), students are expected to propose a miniproject to be presented at the end of the class. The rationale and methodology of this project must be presented as a written document to Dr. Nunez (also as a group/individual). The

structure of this proposal is expected to follow the same structure as an NSF GRFP research proposal (https://www.nsfgrfp.org/). The only difference is that our proposal won't have the broader impacts section and the page limit is max 3 pages. Please find all the pertinent instructions at the NSF PAPPG site. (Note: I expect graduate students to be familiar with this format as well as this funding instrument. Advanced undergraduates are encouraged to talk to Dr. Nunez about this assignment).

Presentation of the project: 25 pts
Students are expected to present their mini projects at the end of the class. Students will use simulation methods to generate preliminary data to substantiate the expected results of their projects. This presentation will be given in person using slides. A time limit will be given in class.

Proposal mock panel participation: 15 pts (panel)
Students will participate in the process of peer-reviewing each other's proposals. This will be done first individually and finally as a group. This will be done by emulating the review practices of an NSF panel.

Diversity and inclusion

This class is intended to be a space of support and an inclusive learning environment that respects student individuality across race, ethnicity, socioeconomic status, sexuality, gender identity, religion, ability, *etc*. I will strive to ensure that each student experiences a fair, safe, and constructive learning environment. If any student is uncomfortable with the material and/or something that has been said in class (by anyone, including other students), please contact me immediately. *To ensure respectful interactions within the class, please let me know your name, and if you like, also your preferred pronouns*. I would like for this class to be as inclusive as possible but I'm not perfect, so please let me know what I can do to improve the class on these matters.

Special accommodations

Accessibility: UVM is committed to the full inclusion of all students. Students with disabilities are welcome to contact the faculty early in the semester to arrange accommodations. Learn more about accessibility at UVM at:

https://www.uvm.edu/academicsuccess/student_accessibility_services

Attendance and religious holidays: Students have the right to practice the religion of their choice. Each semester students should submit in writing to their instructors by the end of the second full week of classes their documented religious holiday schedule for the semester. Faculty will treat these absences as excused and will provide reasonable accommodation to the student concerning missed instruction, assignments, and exams, including final exams. Any conflicts between student and

instructor may be presented for resolution to the course department chair or college dean's office. Learn more about UVM's policy here:

https://www.uvm.edu/registrar/religious-holidays

Illness: Please be advised that, in the unfortunate event of illness, Student Health Services will send me a notification on your behalf. This notification will specify whether the ***qualifying health reason*** requires flexibility for assignments and tests. Please come talk to me to make up for any points lost due to a qualifying health reason.

Honor Code

Students are required to complete their work independently, except where group projects are specifically encouraged. Violation of the Academic Honor Code is a serious issue, with serious consequences.

Please review the Academic Code at:
https://www.uvm.edu/sites/default/files/UVM-Policies/policies/acadintegrity.pdf

---

Email best practices and expectations

Most of my communications to you—outside of class—will happen via your UVM email address. And most likely, the reverse will be true as well. Please be advised that I mostly respond to emails during school hours. Also please note that I try my best to respond to emails promptly, yet, as the head of a research lab, sometimes life can get busy, so please allow 1-2 business days before sending a follow-up email. **When emailing me please include the course code in square brackets [BIOLXXX..] at the beginning of the email subject followed by your name**. Failure to do so may result in your email falling through the cracks of the never-ending stream of emails I often get.

If this is your first year in graduate school this article may be helpful: *How to email faculty*?
The article has some interesting tips:
https://medium.com/@lportwoodstacer/how-to-email-your-professor-without-being-annoying-af-cf64ae0e4087

Grading Rubrics:

Participation, paper discussion, and attendance: 5 pts.

Aspect	5 pts	4 pts	3 pts	2 pts	1 pts	weight
Participation	Contributions to class are constructive and informed by well-understood concepts in the literature.	Contributions to class are constructive. Concepts derive from the literature but are not fully understood	Contributions to class are constructive but do not reveal an understanding of the literature	Some attempts are made to contribute to class, but they reveal that papers were not read. No preparation for class discussion is evidenced.	No attempts are made to contribute to class.	80
Attendance	Attendance is timely, <9% abcenses are unjustified	10%-19% unjustified absences	20%-29% unjustified absences	30%-39% unjustified absences	>40% unjustified absences	20

Bi-Weekly reflections: 15 pts.

Aspect	15 pts	12 pts	9 pts	6 pts	3 pts	weight
Content	Assignments are completed and properly summarized. Appropriate figures and findings are presented	Assignments are completed and properly summarized. figures and findings are presented but there are not fully integrated into the narrative	Assigments are completed and properly summarized. Core figures or findings are missing	Assignments are not completed and improperly summarized. Core figures or findings are missing	Assignments are not completed.	70
Quality of writing	Writing is clear, concise and reflect mastery of the customs and standards of scientific writing.	Writting is clear, concise yet the customs and standards of scientific writing have not been mastered.	Writting clear but neededly verbose.	Writing is unclear.	Writting is unclear and disorganized.	30

Pre-proposal: 25 pts.

Aspect	25 pts	20 pts	15 pts	10 pts	5 pts	weight
Content	Intellectual merits, methods, and preliminary data are appropriately presented. The narrative is compelling and there is clarity why this proposal "should be funded"	At least two of the following sections, Intelectual merits, methods, and preliminary data, are appropriately presented, yet the third section needs further development. The narrative is compelling but may not make the case for funding.	At least one of the following sections, Intellectual merits, methods, and preliminary data, are appropriately presented, yet the other two sections needs further development. The narrative is compelling but may not make the case for funding.	Intellectual merits, methods, and preliminary data need further development. The narrative is compelling but may not make the case for funding.	Intelectual merits, methods, and preliminary data need further development. The narrative is not compelling. The proposal is not competitive.	80

Quality of writing	Writing is clear, and concise and reflects mastery of the customs and standards of scientific writing.	Writting is clear, and concise yet the customs and standards of scientific writing have not been mastered.	Writting clear but neededly verbose.	Writing is unclear.	Writting is clear unclear and disorganized.	20

Presentation of the project: 25 pts

Aspect	25 pts	20 pts	15 pts	10 pts	5 pts	weight
Content	Intelectual merits, methods, and preliminary data are appropriately presented. The narrative is compelling and there is clarity as to why this proposal "should be funded"	At least two of the following sections, Intelectual merits, methods, and preliminary data, are appropriately presented, yet the third section needs further development. The narrative is compelling but may not make the case for funding.	At least one of the following sections, Intellectual merits, methods, and preliminary data, are appropriately presented, yet the other two sections need further development. The narrative is compelling but may not make the case for funding.	Intelectual merits, methods, and preliminary data need further development. The narrative is compelling but may not make the case for funding.	Intellectual merits, methods, and preliminary data need further development. The narrative is not compelling. The proposal is not competitive.	80
Quality of presentation.	Audiovisuals are clear, and concise and reflect mastery of the material, including the first principles of evolutionary biology.	Audiovisuals are clear, yet the aspects of the material, including the first principles of evolutionary biology have not been mastered.	Audiovisuals are clear but neededly verbose.	Audiovisuals are unclear.	Audiovisuals are unclear and disorganized.	20

Proposal mock panel participation rubric and instructions; **15 pts (panel)**

About the panel exercise:
The purpose of this panel is two-fold. First, it seeks to emulate real federal science funding panels at the NSF and the NIH where proposals are discussed and scored based on different aspects of scientific merit review. Second, it seeks to be a space where students can practice the skills associated with **giving professional** feedback to peers. For folks seeking to pursue careers in academic science and technology, both these skills will be crucial for professional success. For folks seeking to pursue careers in non-profits or industry STEM, submitting federal grants will also be an important part of the profession. Finally, for folks seeking more data mining-based positions with little expected interfaced with federal agencies, being able to summarize complex scientific ideas in concise documents will be a key skill for professional advancement in any company.

How does the panel work?
Similar to an NSF panel, after proposals are submitted, they will be assigned to a peer reviewer (another student of the group). This peer reviewer must read the proposal and assess the different reasons why the proposal meets or fails to meet different aspects of scientific merit (see below). These reasons are then discussed with other members of the panel and with the instructor of record, who acts as the *de facto* program officer (PO; or SRO in the NIH system).

Will the pannel's assessment affect my project's grade?
Similar to the NSF, the job of the peer reviewers is to provide expert advice and review on the merits of the proposal. Yet, the decision to fund any given proposal rests solely on the discretion of the PO (the instructor in this case). As such, while the instructor will take into account the opinion of peers in the assessment, the final grade depends exclusively on the instructor's assessment.

What is the panel review grade?
The grade associated with the review process is given to the ***reviewer***. The instructor will assess the level of preparation and professionalism of the peer reviewers while conducting their review. The reviewers must prepare by reading the proposals in full and allow themselves sufficient time to assess the proposal's merits. Proposal review must be done in *good faith*. Any evidence of malicious or unprofessional review will result in substantial grade reductions.

15 pts	12 pts	9 pts	6 pts	3 pts
Reviewer has read and studied the proposal in it entirety and has provided high quality and good faith review	Reviwer has read and studied the proposal in it entirety and has provided average quality and good faith review	Reviwer has read and studied the proposal in it entirety and has provided low quality and good faith review	Reviwer has not read and studied the proposal in it entirety and has provided good faith review	Reviewer has provided a bad faith review

NSF MERIT REVIEW PRINCIPLES AND CRITERIA

Modified from https://www.nsf.gov/pubs/policydocs/pappg22_1/pappg_3.jsp#IIIA

The National Science Foundation strives to invest in a robust and diverse portfolio of projects that creates new knowledge and enables breakthroughs in understanding across all areas of science and engineering research and education. To identify which projects to support, NSF relies on a merit review process that incorporates consideration of both the technical aspects of a proposed project and its potential to contribute more broadly to advancing NSF's mission "to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense; and for other purposes." NSF makes every effort to conduct a fair, competitive, transparent merit review process for the selection of projects.

1. Merit Review Principles: These principles are to be given due diligence by PIs and organizations when preparing proposals and managing projects, by reviewers when reading and evaluating proposals, and by NSF program staff when determining whether or not to recommend proposals for funding and while overseeing awards. Given that NSF is the primary Federal agency charged with nurturing and supporting excellence in basic research and education, the following three principles apply:

- All NSF projects should be of the highest quality and have the potential to advance, if not transform, the frontiers of knowledge.
- NSF projects, in the aggregate, should contribute more broadly to achieving societal goals. These *broader impacts* may be accomplished through the research itself, through activities that are directly related to specific research projects, or through activities that are supported by but are complementary to, the project. The project activities may be based on previously established and/or innovative methods and approaches, but in either case, must be well justified.
- Meaningful assessment and evaluation of NSF-funded projects should be based on appropriate metrics, keeping in mind the likely correlation between the effect of broader impacts and the resources provided to implement projects. If the size of the activity is limited, evaluation of that activity in isolation is not likely to be meaningful. Thus, assessing the effectiveness of these activities may best be done at a higher, more aggregated, level than the individual project.

Concerning the third principle, even if the assessment of Broader impact outcomes for particular projects is done at an aggregated level, PIs are expected to be accountable for carrying out the activities described in the funded project. Thus, individual projects should include clearly stated goals, specific descriptions of the activities that the PI intends to do, and a plan in place to document the outputs of those activities.

These three merit review principles provide the basis for the merit review criteria, as well as a context within which the users of the criteria can better understand their intent.

2. Merit Review Criteria: All NSF proposals are evaluated using two National Science Board-approved merit review criteria. In some instances, however, NSF will employ additional criteria as required to highlight the specific objectives of certain programs and activities.

The two merit review criteria are listed below. Both criteria are to be given full consideration during the review and decision-making processes; each criterion is necessary but neither, by itself, is sufficient. Therefore, proposers must fully address both criteria. (*Chapter II.C.2.d(i)* contains additional information for use by proposers in the development of the Project Description section of the proposal.) Reviewers are strongly encouraged to review the criteria, including *Chapter II.C.2.d(i)*, before reviewing a proposal.

When evaluating NSF proposals, reviewers will be asked to consider what the proposers want to do, why they want to do it, how they plan to do it, how they will know if they succeed, and what benefits could accrue if the project is successful. These issues apply both to the technical aspects of the proposal and how the project may make broader contributions. To that end, reviewers will be asked to evaluate all proposals against:

1. What is the potential for the proposed activity to Advance knowledge and understanding within its field or across different fields (Intellectual Merit); and

2. To what extent do the proposed activities suggest and explore creative, original, or potentially transformative concepts?

3. Is the plan for the proposed activities well-reasoned, well-organized, and based on a sound rationale? Does the plan incorporate a mechanism to assess success?

4. How well qualified is the individual, team, or organization to conduct the proposed activities?

5. Are there adequate resources available to the PI (either at the home organization or through collaborations) to carry out the proposed activities?

Class Topics and Schedule

Week	Date	Day	Lecture Content	Computational Content
1	1/16/24	Tue	Introduction to the course. Logistics of the course. **Topic 1**: *Overview of evolutionary processes.*	**Prac 1:** Tour of the VACC
	1/18/24	Thu	**Topic 1 (cont):** *VACC visit → introduction to HPC computing and VACC idiosyncrasies* **Shelly Johnson visit**	**Prac 2:** *Data Challenge:* Data mining a genome and its annotation. *Quiz 1*
2	1/23/24	Tue	**Topic 2:** *Refresher of genetics*	(start with) **Prac 2 (cont.)** Open floor for questions
	1/25/24	Thu	**Prac 3:** *Refresher of basic statistics* using R, *tidyverse*, and GGplot *Quiz 2* ***Reflexion 1 Due***	
3	1/30/24	Tue	**Topic 3:** *Null models of evolution, evolutionary sampling, and power analyses (cont.)*	**Prac 4:** Scale up scripts using arrays
	2/1/24	Thu	Github and Replicability **Joe Gunn visit**	**Guest Workshop:** R/Github integration *Quiz 3*
4	2/6/24	Tue	**Topic 3:** *Natural Selection*	**Prac 5:** Basic modeling selection and fitness in R
	2/8/24	Thu	**Topic 4:** *Neutral evolutionary forces and genetic drift* ***Reflexion 2 Due***	**Prac 6:** Modeling allele frequencies in R *Quiz 4*
5	2/13/24	Tue	**Topic 5:** Linkage disequilibrium *and how it interacts with other selection and drift*	**Prac 7:** Reproducing published code: https://github.com/fcbnunez/Coille-Seasonality-2016-2019/blob/main/CODE/8.0.Linkage/Genetics_resubmission1/1.0.calculate_linkage_w_inversion.r
	2/15/24	Thu	Discussion of Albecker *et al.* https://doi.org/10.1098/rspb.2021.2122	**Prac 8:** Reproducing published code: https://github.com/DrK-Lo/LifeCyclePower/tree/master *Quiz 5*
6	2/20/24	Tue	Discussion of Hoban *et al.* https://doi.org/10.1086/688018	**Prac 9:** Using *msPrime* to generate scalable neutral simulations
	2/22/24	Thu	***Reflexion 3 Due*** **Topic 6:** Summarizing the site frequency spectrum to infer evolutionary histories	**Prac 10:** Methods to quantify genetic variation in genomic data *Quiz 6*
7	2/27/24	Tue	Discussion of Vitti *et al.* https://doi.org/10.1146/annurev-genet-111212-133526	**Prac 11:** Introduction to SLiM (**Wright-Fisher models; Demography**)
	2/29/24	Thu	**Prac 11 (cont'):** Introduction to SLiM; (**Wright-Fisher models; Demography**) *Quiz 7*	
8	**3/5/24**	Tue	**NO CLASS**	
	3/7/24	Thu	Discussion of Lotterhos *et al.* https://doi.org/10.1371/journal.pbio.3000070 **and** Discussion of Melzner et al. https://doi.org/10.1371/journal.pbio.3001641 ***Reflexion 4 Due***	**Brainstorming session for proposals** *Quiz 8*
9	**3/12/24**	Tue	**NO CLASS**	

	3/14/24	Thu		
10	3/19/24	Tue	**Prac 12: Selection in Wright-Fisher models**	
	3/21/24	Thu	**Prac 13: Selection ia *Non*-Wright-Fisher models and QTN** *Quiz 9* ***Reflexion 5 Due***	
11	3/26/24	Tue	Discussion of Lou *et al.* https://doi.org/10.1534/g3.120.401287	**Prac 14:** Power analyses in (simulated) experimental evolution
	3/28/24	Thu	Discussion of Matz *et al.* https://doi.org/10.1111/gcb.15060 ***Reflexion 6 Due***	Visit from senior QuEST trainee; A. R. McCraken. Implementing simulation into a thesis. *Quiz 10*
12	4/2/24	Tue	**Topic 7:** Explanation of ***project proposal*** structures Research Time + Group/individual visit with Dr. Nunez	
	4/4/24	Thu	Research Time + Consult with Dr. Nunez	
13	4/9/24	Tue	Research Time + Consultwith Dr. Nunez	
	4/11/24	Thu	Research Time + Consult with Dr. Nunez	
14	4/16/24	Tue	Research Time + Consult with Dr. Nunez	
	4/18/24	Thu	**Topic 8:** Review of Scientific Review Criteria **(*project proposals* *due at the end of the day*)**	
15	4/23/24	Tue	**Final Presentations** **(*read peer proposals for evaluation and assessment*)**	
	4/25/24	Thu		
16	4/30/24	Tue	**Proposal Mock Panel**	
	5/2/24	Thu		