

## PBIO/BIO 381 – Ecological Genomics Spring 2017, 4 Credits

### Instructors:

	Dr. Stephen Keller (Prof)	Dr. Melissa Pespeni (Prof)	Andrew Nguyen (GTA)
<u>Office:</u>	335 Jeffords Hall	326A Marsh Life Science	211 Marsh Life Science
<u>Phone:</u>	802-656-5121	802-656-0628	802-656-0451
<u>Email:</u>	<a href="mailto:srkeller@uvm.edu">srkeller@uvm.edu</a>	<a href="mailto:mpespeni@uvm.edu">mpespeni@uvm.edu</a>	<a href="mailto:andrew.nguyen@uvm.edu">andrew.nguyen@uvm.edu</a>
<u>Office Hours:</u>	Mondays 3-4pm, Tuesdays 12-1pm	Tuesdays and Wednesdays 2-3pm	TBA

**Meeting Location:** 326 Jeffords Hall

**Meeting Times:** 8:30-11:30 AM, Mondays and Wednesdays

**Overview:** Genetic variation provides the foundation for biodiversity, and is the raw material for species adapting to changes in their environment. Advances in genomics, especially the advent of next-generation sequencing, now permit the study of genetic variation and gene expression across the genome, and in virtually any species. This has opened up new frontiers for integrating ecology with genetics to address fundamental issues such as the genetic basis of local adaptation, the networks of genes involved in phenotypic plasticity and response to environmental stress, and the genetic consequences of speciation and hybridization. Additionally, assessment of species diversity from high-throughput sequencing of “environmental DNA” has empowered new approaches to community ecology, especially of microbial communities (“microbiomes”) that are often difficult or impossible to culture.

Taking advantage of this new field requires careful consideration of experimental design, as well as the bioinformatics skills needed for handling and making biological sense of the massive amounts of data. This course will cover major research topics in ecological genomics and highlight case studies and applications of genomics to contemporary ecological and environmental problems. It will also cover the application of the primary analytical tools used to generate and analyze genomic data.

### Course Objectives:

- To impart an in-depth understanding of the interplay between genomic and ecological variability, and how each shapes the other. This includes (i) emphasizing the ways in which genomic variation can “scale up” to affect populations, communities and ecosystems, as well as (ii) understanding how the diversity and function of the genome has been shaped by ecological and environmental processes.

- To think creatively about major questions in ecological genomics and refine students' skills in posing testable hypotheses, including learning to pair these hypotheses with the appropriate 'omics data.
- To appreciate the role of careful experimental design and statistical analysis in ecological genomics, including issues of power, false positive rates, and how to evaluate statistical biases vs. biological signals in ecological genomics data.
- To achieve a working knowledge and level of comfort with some of the bioinformatics routines necessary for ecological genomics studies, and gain hands-on experience in working with these analyses on real data.

### **Learning Outcomes:**

*Throughout the semester, students will learn to*

- Discuss exemplary and representative studies in the field of ecological genomics and identify the frontiers of the field.
- Develop an ecological genomics research project, motivate the study, and design the analysis.
- Work in the Unix computing environment using command line.
- Work on a remote server and transfer data among machines.
- Write basic scripts in bash and R.
- Create an electronic lab notebook to document their coursework through the semester and the work for their final projects.
- Create final presentations in groups and present the study and results.
- Write a final paper in the format of a scientific publication.

**Text:** There will be no course textbook; instead we will read heavily from the primary literature, emphasizing both classic studies and the latest papers applying ecological genomics approaches (see proposed reading list). Students seeking a reference text for review of basic concepts in population genetics, molecular evolution, and genomics should refer to the following texts: "Fundamentals of Molecular Evolution" (Graur and Li), "Evolutionary Analysis" (Freeman and Herron), and "Principles of Population Genetics" (Hartl and Clark). A good reference for computational biology is "Practical Computing for Biologists" (Haddock and Dunn). Copies of these books are available from the instructors for short term loan.

**Grading:** Students will be evaluated based on presentation and analysis of the class project (final presentation: 5%, and final paper: 25%), short homework assignments (4 assignments, totaling 40%), "info update" presentations (10%), contributions to class paper discussions (10%), and lab notebooks (10%).

**Class project:** You will work in small groups of 2-4 people to design and analyze our own ecological genomics study as a class project, based on previously sequenced (but unpublished!) data. This project will be a collaborative effort and will run throughout the semester. Project groups will come together early in the semester to pose an ecological genomics question to answer with the data, formulate an analysis strategy, and present their ideas to the class. Throughout the semester, groups will work through a comprehensive analysis of the data, and give an oral presentation of their findings to the class at the end of the semester. Each

student will also individually write-up the project results and submit a final paper describing the experiment, the analysis, and its results and interpretation.

The final paper will be in the style of the journal, *Molecular Ecology*. Formatting should follow that of an Original Article (see author guidelines on the journal's website: [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1365-294X/homepage/ForAuthors.html](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1365-294X/homepage/ForAuthors.html)). However, you should restrict the length of the manuscript to no more than 4500 words of text exclusive of references, tables, abstract and figure captions. You may include a **combined total** of 5 figures and tables.

**Short Write-ups:** Throughout the semester, we will gain proficiency and build our skills and confidence from working hands-on with different types of genomic data and analysis approaches. Data and concepts will be introduced in class, but a significant part of the analysis will then be expected to occur outside of class time. Students will turn in short write-ups of 1-2 pages single-spaced (max) plus tables/figures for each assignment. Short write-ups will be due 1 week after being assigned.

**Info-updates:** These will be short informational presentations, presented by someone other than the discussion leader for that day, that take place in the first ~20 minutes of class. The purpose of the info-update is to introduce a concept, or explain a technique or some terminology that we need to understand in order to discuss the paper effectively. When it's your turn to present an info-update, you should come with an organized plan for the material, but use the board to explain (no powerpoints). You may also choose to provide a handout with figures and/or tables. Each student must present an info-update during the semester. We will provide instructions in class on topics and how to sign up for a spot.

**Paper discussions:** Frequent readings will be assigned from the primary literature to accompany the topics covered that week and to provide an applied example of genomics to ecological and environmental problems. These will include both background review type articles, and data papers that illustrate the application of a particular approach. All students will be expected to have read the papers prior to class meeting and contribute actively to the class discussion. Everyone should bring a copy of the papers and be prepared to discuss the figures and findings.

#### **Lead Discussion:**

Each student must sign up to lead a discussion during the semester. As the discussion leader, you will need to:

- Give a **brief introduction** (maximum 5 minutes) placing this paper in context. □Note: this introduction should NOT be a summary of the findings of the paper. We are going to discuss the findings as a group. □
- You will also **start discussion** by picking someone to start off explaining Figure 1 (or results mentioned in the text, if they are not in a figure). Discussion will continue around the circle until we finish the figures/tables/results. □
- If someone is having a hard time with a figure, you can **ask questions** to draw out both their interpretation or to encourage others to join in. □For example, you could ask: "What does this experiment show?" Or "Why did they do this experiment?"□

- **Pose questions to the group** that stimulate thinking about the deeper issues in this paper or about connections to other papers we have read.

**Participate in Discussion:** All students are expected to participate actively in the discussions. This means joining in the discussion, providing explanations, asking questions, and making connections between something in the paper being discussed and other material in class or outside of class. **To get full credit for participation**, you must speak not only when you are called on to present a figure, but must also join in the discussion by offering comments or questions when other people are presenting or when we are discussing the themes and findings of the paper.

**Lab Notebooks:** A key to success in computational biology is being able to understand and repeat your analyses. This involves organizing your files, keeping good notes on your analysis pipeline, and documenting the history of your work and input/output files. Github enables researchers to achieve all of these things by saving changes to any/all files in your project, which in turn, allows you to track the different versions, known as version control. Furthermore, Github supports a relatively simple word processing language (markdown), and lets you share your work online. Therefore, all research activities on your computer can be shared with others.

We have created a model project layout on Github ([https://github.com/adnguyen/2017\\_Ecological\\_Genomics](https://github.com/adnguyen/2017_Ecological_Genomics)) for you to use as a template to create your own Github repository. You will use github throughout the semester to keep an online lab notebook of your work, as well as document your progress towards your group project. Your lab notebooks should be kept up to date regularly, and will be assessed periodically for completeness. More information will be given in class about using Github for your lab notebook.

### **Computational resources and tools:**

1. **Terminal shell:** Throughout the semester, we'll be doing a lot of computational analysis of genomic data. The bulk of the heavy lifting in terms of processing will be done by a remote Linux server that we will all have access to, as the size and scope of the analysis typically exceeds what you would want to attempt on your personal laptop (nevermind that you might want to do something else with your machine sometime!). You'll need to be able to log-in to this server remotely, run jobs, and transfer results files as needed to your local machine for plotting, etc. If you have a laptop running Mac or Linux OS, then the best way to connect with the server is via secure shell (ssh) in your terminal. If you are a Window's user, you should download and install the free software, "PuTTY", that allows you to establish a secure connection via terminal to a remote Linux machine: <http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html>
2. **Remote file transfer:** For transferring files to and from a remote server, some people prefer to use command-line tools while others prefer a GUI interface for secure file transfer. Examples of the latter for MacOS users are Fetch (<http://fetchsoftworks.com>) or FileZilla (<https://filezilla-project.org>), and for Window's users are WinSCP (<https://winscp.net/eng/download.php>). File transfer can also be accomplished at the command-line, and we will cover how to do this.

3. R for stats and graphing: For many downstream statistical analyses, the statistical genetics community has been converging on R (<https://www.r-project.org>) as the tool of choice. Many R packages are being maintained and new ones being developed to facilitate genomic analyses. Also, even if a particular analysis is run using a program outside of R, R is an excellent tool for turning your results into publication-quality figures. You can use R directly at the command-line, or some users prefer R-Studio (<https://www.rstudio.com>), which has a nice front-end to the base R engine running under the hood.
4. Basic text editing: Lastly, you'll want a good, simple text editor for writing scripts, looking at data, and occasionally doing some data manipulation. Sometimes we will work directly within command-line Unix on the remote server to view/edit text (using Vim, for example). Other times, it's nice to have a simple text editor for working with files on your local machine. Examples of other good stand-alone free text editors are TextWrangler (<http://www.barebones.com/products/textwrangler/download.html>) for MacOS, and Notepad++ (<https://notepad-plus-plus.org/download/v6.8.1.html>) for Windows.
5. Github: We'll be using Gitub for online notebooks and version control. See information under "Lab notebooks"

### **Student Responsibilities and Rights:**

*Academic Integrity*: All students are expected to abide by UVM's Code of Academic Integrity (<http://www.uvm.edu/policies/student/acadintegrity.pdf>). Plagiarism, fabrication, or other forms of academic dishonesty will not be tolerated.

*Accommodation*: Students with a documented disability will be accommodated. Students should discuss their needs with the instructors, and should contact ACCESS, the office of Disability Services on campus (A170 Living/Learning Center - 802-656-7753 - [access@uvm.edu](mailto:access@uvm.edu)).

*Religious Holidays*: Students have the right to practice the religion of their choice. If you need to miss class to observe a religious holiday, please submit the dates of your absence to the instructors in writing by the end of the second full week of classes. You will be permitted to make up work within a mutually agreed-upon time.

Course schedule (Tentative; subject to change)

Week	Topic	Hands-on learning session	Topic	Assigned reading <i>Discussion papers in italics; [optional info-update papers in underline]</i>
Jan 18	Course Intro and structure	Discussion of class project; intro to lab notebooks	Intro	(none)
Jan 23	The next-generation sequencing revolution and the rise of non-model species	Tour of UVM genomics core	Intro	<i>Ellegren 2014</i> ; <u>van Dijt et al. 2014</u> ; <u>Stapley et al. 2010</u> ; <u>Glenn et al. 2011</u>
Jan 25	Thinking critically about genomic discovery of ecological adaptation	Small group discussion of project ideas	Intro	<i>Rockman 2012</i> ; <i>Lee et al. 2014</i>
Jan 30	Group presentations of project ideas		Intro	(none)
Feb 1	Sequencing strategies applied to biological questions: WGS, RNA-Seq; RAD/GBS, Amplicons	Intro to basic UNIX command line; connecting to remote machines; working with text files	Intro	<b>WGS</b> : <u>Ekblom and Wolf 2014</u> ; <b>RNA-Seq</b> : <u>Wolf 2013</u> ; <b>RAD/GBS</b> : <u>Andrews et al. 2016</u> ; <b>Amplicon-Seq</b> : <u>Bybee et al. 2011</u>
	<b>Short write-up #1: project proposal (due 2/6)</b>			
Feb 6	Transcriptomics 1:	Working with fastq files; Clean & trim reads	Transcriptomics	<i>Dunning et al. 2014</i> ; <u>DeWit, Pespeni, et al. 2012</u>
Feb 8	Transcriptomics 2:	Assemble transcriptome; Evaluate assembly; begin read mapping to reference transcriptome assembly	Transcriptomics	<i>De Panis et al. 2016</i> ; <u>Alvarez et al. 2015</u>
Feb 13	Transcriptomics 3:	finish read mapping -- merge sam files into combined assembly	Transcriptomics	<i>Johnston et al. 2016</i> ; <u>Todd et al. 2016</u>
Feb 15	Transcriptomics 4:	tabulate expression counts; intro to DESeq2 for measuring DE	Transcriptomics	<i>Zhao et al. 2016</i> ; <u>De Wit, Pespeni and Palumbi 2015</u>

Feb 20	President's Day -- no class			
Feb 22	Transcriptomics 5:	Measure differential gene expression (DESeq2)	Transcriptomics	<i>Dixon et al. 2015 (plus SI); Hrdlickova et al. 2017</i>
Feb 27	Transcriptomics 6:	Co-expression analyses; Annotation	Transcriptomics	<i>Schunter et al. 2016; Mayfield et al. 2016</i>
Mar 1	Catch-up day	TBD	Transcriptomics	<i>no discussion/info update</i>
	<b>Short write-up #2: gene expression (due 3/8)</b>			
Mar 6	Population genomics 1:	intro to SNP and genotype calling; filtering strategies; working with vcf files	Population genomics	<i>Gayral et al. 2013; Nielsen et al. 2011</i>
Mar 8	Population genomics 2:	nucleotide diversity; site-frequency spectrum	Population genomics	<i>Romiguier et al. 2014; Lanfear et al. 2014</i>
Mar 13/15	UVM Spring Recess -- no class			
Mar 20	Population genomics 3:	Analysis of population structure: PCA; genotypic clustering	Population genomics	<i>Gompert et al. 2014; Fumagalli et al. 2013; Padhukasahasram 2014</i>
Mar 22	Population genomics 4:	Analysis of population demographic history	Population genomics	<i>McCoy et al. 2014; Sousa and Hey 2013</i>
	<b>Short write-up #3: population structure (due 3/29)</b>			
Mar 27	Population genomics 5:	Selection scans and outlier analysis part I. selective sweeps	Population genomics	<i>Laurent et al. 2016; Messer and Petrov 2013</i>
Mar 29	Population genomics 6:	Selection scans and outlier analysis part II. local adaptation	Population genomics	<i>Kubota et al. 2015; Hoban et al. 2015</i>
Apr 3	Catch-up day	TBD	Population genomics	<i>no discussion/info update</i>
	<b>Short write-up #4: testing for selection (due 4/10)</b>			
Apr 5	Annotation & enrichment analysis	Gene ontology (GO) term enrichment testing	Enrichment	<i>Berens et al. 2015 Primmer et al. 2013</i>
Apr 10	Metagenomics 1:	Community sequencing and barcoding strategies	Metagenomics	<i>Ofek-Lalzar et al. 2014; McFall-Ngai 2013</i>
Apr 12	Metagenomics 2:	Clustering OTUs and determining taxonomy	Metagenomics	<i>Bolnick et al. 2014; Goodrich et al. 2016</i>
Apr 17	Metagenomics 3:	Estimating diversity of microbiomes	Metagenomics	<i>Easson &amp; Thacker 2015; Hong et al. 2014</i>

Apr 19	Metagenomics 4:	Community structure and hypothesis testing	Metagenomics	<i>Roder et al. 2015; Zinger et al. 2014</i>
Apr 24	Class Projects -- open lab	Open lab		(none)
Apr 26	Class Projects -- open lab	Open lab		(none)
May 1	<b>Presentation of class projects</b>			
May 3	<b>Final papers due</b>			

### Other Online Resources:

Github: for keeping your online notebooks and group project work

<https://github.com>

SeqAnswers -- a helpful community resource for bioinformatics:

<http://seqanswers.com>

Several good online Linux tutorials and “cheat-sheets”:

<http://linuxcommand.org/index.php>

<http://ss64.com/bash/>

Useful blogs with code snippets:

The Molecular Ecologist: <http://www.molecular ecologist.com>