Description: This course will take students from raw DNA sequencing data through quality assurance, through to data interpretation, statistical analysis, and presentation of the results as a mock scientific article. A background in microbiology, microbial ecology, or genetics would be beneficial. No programming or data analysis experience is required. Students who are performing research may bring their own sequencing data to process in class. Students will become familiar with command-line programs and basic computer programming techniques; understand bioinformatics methods such as quality trimming, assembling contigs, sequence alignment, using reference databases, and statistical comparisons; gain hands-on experience in bioinformatic analysis of DNA sequences using the R platform and its packages; primarily, DADA2, phyloseq, vegan, ggplot2; and be able to apply the knowledge gained in class to other sequence types and programs. Students may bring their own data, or some can be provided. AVS 454 and 554 cannot both be taken for credit.

Credit Hours: 2
Prerequisites: AVS 254 or BIO 319 or Bio 350 or BMB 280 or WLE 200 or SMS 300, and STS 232 or MAT 215; or graduate student standing
General Education requirements satisfied: Quantitative Literacy
Mode of Instruction: In-person course. Remote connection will automatically be provided each week for off-campus students, but local students may elect to attend remotely at any time.
Time: Synchronous, but lectures are recorded and made available for asynchronous students

Course Schedule Disclaimer (Disruption Clause): In the event of an extended disruption of normal classroom activities (due to COVID-19 or other long-term disruptions), the format for this course may be modified to enable its completion within its programmed time frame. In that event, you will be provided an addendum to the syllabus that will supersede this version.

Digital Services, Hardware, Software: Brightspace, Zoom

Instructional Material: Reading material is provided as electronic journal articles via Brightspace that reflects current literature in host-associated microbial ecology and data analysis. All software used is free online. Sequence data will be provided; however, students may elect to work on their own data. Accommodations to class format or material available as needed.

Class format: Short lectures followed by guided computer laboratory time. Various outputs from the analysis will be submitted online for assignments. This course requires access to a computer.

Student Learning Objectives:
After completion of the course, students will be able to:
• Use an understanding of bioinformatics methods, such as quality trimming, assembling contigs, sequence alignment, using reference databases, and statistical comparisons, to curate a data processing and analysis workflow. This may include bioinformatic analysis of DNA sequences, using the R platform and its packages, MEGA, NCBI genome assembly, MG-RAST, etc. (Quantitative Literacy)
• Demonstrate proficiency in taking raw DNA sequence data through quality control steps to interpretation, and summation of the workflow and results into mock scientific journal article manuscripts. (Quantitative Literacy)

• Demonstrate scientific writing skills, specific to manuscript preparation, including incorporating instructor and peer-review comments and revisions. Submit multiple drafts and progression the ideas with each draft.

• Demonstrate skills in peer-reviewing manuscripts, including reviewing, editing, and scientific critique.

Attendance policy: Students are expected to attend lectures, but it is understood that life often precludes this and that students may be performing field work or are located off-campus. Students may attend class virtually, through Zoom, which will be offered for each class. Students who will miss a significant number of classes, or who require additional accommodations, may contact me to make alternate arrangements.

➢ Pregnancy, lactation, and parenting: I am happy to make accommodations for students based on pregnancy, lactation, and parental needs, as well as work with the Office of Equal Opportunities. Maine state and UMaine policy allows students to breastfeed in any space, including in class. If a lactation space is required, please contact E.O. for arrangements.

  o Pregnant on Campus Initiative, pregnancy and parenting resources in Orono
    https://pregnantoncampus.studentsforlife.org/campus/umaine-orono/

➢ Food insecure? Need clothes? Check out the Black Bear Exchange’s Food Pantry: https://umaine.edu/volunteer/black-bear-exchange/ or Old Town Crossroads Ministry.

Class participation: Students are expected to participate in discussions in class. I strive to create inclusive discussions, but if students still find it challenging to participate please notify me and I will alter the discussion format as needed.

Late Assignments: Assignments will be accepted after deadlines, but you might not receive feedback. Assignments will not be accepted after the last day of the semester.

Classroom policy: Supporting inclusion and community in science is an active process that involves both invitation, and support to ensure that the scientific community is and remains an equitable and inclusive place. Students are expected to conduct themselves in a professional and courteous manner, and to abide by University policies.

Campus Policies: “The University of Maine is an EEO/AA employer, and does not discriminate on the grounds of race, color, religion, sex, sexual orientation, transgender status, gender expression, national origin, citizenship status, age, disability, genetic information or veteran’s status in employment, education, and all other programs and activities.” Follow the links for more information.

  Academic Honesty Statement*
  Students Accessibility Services Statement*
  Course Schedule Disclaimer*
  Observance of Religious Holidays/Events*
  Sexual Discrimination Reporting (Long)*
  Sexual Discrimination Reporting (Short)*
  ** I am a “mandatory reporter”*. If you disclose something to me, I am obligated to disclose to the relevant campus Title IX office. This includes information revealed in class assignments.
Grading (out of 100 points): A = 93–100; A− = 90–92; B+ = 87–89; B = 83–86; B− = 80–82; C+ = 77–79; C = 73–76; C− = 70–72; D+ = 67–69; D = 63–66; D− = 60–62; F = 0–59

Grading:

| Mini scientific manuscript: 60% (3 drafts at 10%, 20%, and 30%) | Lab work output will be used to write one scientific manuscript using amplicon sequencing data, which will be submitted with successive revisions 3 times during the semester. We will generate the Methods and Results section in lab, and students will be responsible for generating the Introduction and Discussion sections independently. Students may work collaboratively with up to three students per group, but the manuscript length, depth of information, and quality of the writing should reflect the number of students in the group. Graduate students are expected to present a higher quality of writing, > 15 citations, more nuanced statistical analysis or graphical representation, and more in-depth discussion sections. Specific instructions are provided on Brightspace, and guidelines may be found in the “writing manuscripts” reading. For each successive submission, students will incorporate revisions from instructor and (when available) peer review to progress the complexity of the scientific content and the maturity of the writing style. At the end of the semester, students may opt to use their analysis and manuscript and pursue submission in a scientific journal along with the research team who provided their dataset. Not all datasets may be applicable, and the final decision will rest with the student and with the Principal Investigator who owns the data. Submission for journal review is completely elective and is not considered in the grading of this class in any way. |
| Peer Review Undergrad: 1 at 20% Graduates: 2 at 10% each | Review another student’s manuscript for the amplicon manuscript submission, per instructions in the “Peer Reviewing” PowerPoint. Graduate students will perform two peer reviews. |
| Assignments: 20% (4 at 5% each) | As instructed on Brightspace and in the Lecture schedule |

Note on authorship: If you are pursuing a manuscript for publication in this class, the work you generate is your intellectual property. I do not expect to be an author on your manuscript, or to have ownership over any materials you generate. I would like me/the class to be mentioned in the Acknowledgments section. I will help you facilitate authorship roles with the full research team (i.e. the people that generated these data).

**Prior to the first class**
Download and install R (the program candy): [https://www.r-project.org/](https://www.r-project.org/) and Rstudio (the fancy wrapper): [https://rstudio.com/products/rstudio/download/](https://rstudio.com/products/rstudio/download/)
- **Suggested Reading (for new R users):** [http://www.r-tutor.com/r-introduction](http://www.r-tutor.com/r-introduction)
- **Suggested Reading (for new R users):** “Basic Info on R” ppt, Ishaq, on Brightspace
- Suggested additional software for viewing/editing files: Sublime 3 text editor: [https://www.sublimetext.com/3](https://www.sublimetext.com/3)
**After every class**

Update/clean up your code, annotate with notes, add to your methods or results section of your manuscript by describing what you did that day.

**Lecture schedule (D1x 2H):**

<table>
<thead>
<tr>
<th>Wk</th>
<th>Lec.</th>
<th>Topic and Notes</th>
</tr>
</thead>
</table>
| 1 | 1   | Introduction to R and sequencing data  
Lecture: “Intro” to the course. “Data files and quality”, intro to batch/workflow files and keeping good notes.  
Lab: Installing R and packages, keeping good notes and workflow files. Intro to sequencing files and the information provided within (i.e. quality data). Assessing data quality and quality filtering, and whether to use contigs or single read.  
- **Reading**: DNA technology and rRNA ppt, Ishaq  
- **Video (25 min)**: DNA sequencing technology ppt, Ishaq  
- **HW**: Continue personalizing your copy of the workflow, including file and folder names. Make sure you have the data files and metadata for your project on your machine. Complete the “filter and trim” step in DADA2 by next class.  
- **Assignment (5%)**: Plagiarism quiz on Brightspace, ~ 40 min of time, due by next lab |
| 1/27 | | |
| 2 | 2   | Picking either sequence variants or OTUs  
Lecture: “Picking sequences out of your data”. Overview of alignment, genetic distance, clustering and picking OTUs, or the alternative method; using sequence variants.  
Lab: dereplication, learning error rates, and picking SVs in DADA2  
- **Reading**: Callahan_2016_DADA2  
- **Reading**: Genetic distance ppt, Ishaq |
| 2/3 | | |
| 3 | 3   | Taxonomy, Chimeras and how to slay them  
Lecture: “Taxonomy, Chimeras and how to slay them”. Sequence identification using reference database files, and using those reference databases to identify and remove chimeric sequences  
Lab: remove chimeras with DADA2 and assign taxonomy (with Silva)  
- **Reading (option 1)**: Balvociute_2017_comparing_taxonomic_databases  
- **Reading (option 2)**: Smith_2020_database_choice_rumen_microbiome  
- **HW**: Complete dereplication, learn error rates, pick SVs, and remove chimeras. Complete assign taxonomy (with species is optional) by next lab |
| 3 | 4   | Removing contaminants  
Lecture: “Removing contaminants”. Revisiting data quality discussion and the wet-lab and dry-lab contaminants you are likely to find in sequencing data, and setting up code to use sequenced negative controls to clean up your data.  
- **Lab**: Introduction to phyloseq, workflow verification steps, and taking an initial look at the data. **If you have negative controls or DNA quantification data**: remove contaminating sequences from data using Dr. Ishaq’s code or decontam. If you don’t have negative controls, this time will be used to catch up, revise, or write. Either way, you can remove unwanted taxa by name. |
| 2/10 | | |
| 5 | 5   | Experimental design and models  
Lecture: experimental designs, and building your statistical model  
Lab: Write out research questions, make them specific. Free time to catch up on analysis, re-do, troubleshoot. |
<table>
<thead>
<tr>
<th>Date</th>
<th>Page</th>
<th>Topic</th>
<th>Reading</th>
<th>Assignment (5%)</th>
<th>Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/17</td>
<td>6</td>
<td>Rarefaction, and alpha diversity</td>
<td>Ramette_2007_multivariate_microbial_ecology</td>
<td>“Types of scientific writing” Available on Brightspace, ~ 30 min of time, due by next lab.</td>
<td>4/24</td>
</tr>
</tbody>
</table>

2/17

Date: 2/17

Topic: Rarefaction, and alpha diversity

Lecture: “Alpha diversity”, and how to measure it

Lab: Subsampling, and alpha diversity, including graphics generation (line/bar/violin, and correlograms) and stats.

- **Reading:** Ch8 species composition and distance

5/24

Date: 2/24

Topic: Comparing changes in taxonomy

Lecture and lab: “Delineation of taxonomic change”; general guidelines for displaying taxonomy, as well as DESeq2, forests, LEFSe.

- **Reading:** Rajendhran_2011_16S_phylogeny_diversity

- **Assignment (5%):** Complete the “DIY taxonomic reference database” exercise on creating reference databases using MEGA, due by next class. Instructions on Brightspace.

6/3

Date: 3/3

Topic: Beta diversity

Lecture and Lab: “Beta diversity”. Community-level similarity and ordinations, stats to go along with them.

- **Due:** Reference fasta and taxonomy file made in MEGA
- **Reading:** Lozupone_2008_measuring_species_diversity

7/3

Date: 3/10

Topic: Beta diversity Part II

Lecture and Lab: “Beta diversity component analysis”. more community-level analysis. RDA, CCA, db-RDA, WTF.

- **Reading:** Writing manuscripts ppt, Ishaq
- **Suggested reading:** How do you figure ppt, Ishaq

8/4

Date: 3/17

Topic: Tree building

Guest intro from Dr. Erin Grey, Assistant Professor of Aquatic Genetics and part of Maine-eDNA group, to talk about core and microorganism sequencing, genomics, etc. May also feature Dr. Andrew Rominger, Assistant Professor of Ecological Bioinformatics.

Lab: Trees as needed. Free time for additional analysis/writing

- **Manuscript (10%):** Draft 1 of amplicon analysis manuscript, ~1500 words not including references, > 5 citations. Specific directions on Brightspace. Due next lab 3/31.

9/11

Date: 3/24

No class

Whole-genome sequencing

10/11

Date: 3/31

Topic: Whole-genome sequencing

Lecture: “Intro whole-genome” and relevant tech.

Lab: quality trimming and contig assembly: de novo vs. scaffold based. Identification of SNPs. We will be working collaboratively on some Vibrio genomes.

- **Due:** First draft of amplicon analysis manuscript.
- **Peer review (20%):** peer review will be manually sent to you, due next lab 4/7
- **Reading:** Ekblom_2014_whole_genome_sequencing

11/12

Date: 4/7

Topic: Gene identification and genome annotation

Lecture and lab: “Genome identification and annotation”. We will be working...
| 4/7      | collaboratively on some Vibrio genomes.  
|          | • **Reading:** Zhulin_2015_databases_review  
|          | • **Due:** peer review  
| 12 13    | Whatever else needs to be done on whole genome  
| 4/14     | We will be working collaboratively on some Vibrio genomes.  
|          | • **Manuscript (20%):** Second draft of amplicon analysis manuscript, ~2000 words. Should include revisions, and more citations. Due next lab 4/21.  
| 13 14    | Finish up whole genome if needed or start Intro to metagenomics  
| 4/21     | • **Due:** second draft of amplicon manuscript  
|          | • **Reading:** Keegan_2016_Protocol_MG-RAST_Metagenomics  
|          | • **Reading:** Escobar-Zepeda_2018_taxonomy_metagenomics  
|          | • **Assignment (5%):** Walkthrough of MG-RAST, on Brightspace  

**Metagenomics**

| 14 15    | Intro to metagenomics, metatranscriptomics, and assembly  
| 4/28     | Lecture and lab: we will play around with some metatranscriptomics data on the ACG.  
|          | • **Reading:** Tringe_2005_metagenomics_terrestrial_marine  
|          | • **Reading:** Hu_2015_metagenomics_bio_heap_leaching  
|          | • **Due:** Walkthrough of MG-RAST, on Brightspace  
|          | • **Manuscript (30%):** Final draft of amplicon analysis manuscript. Should include revisions, and more citations. Due during finals week, 5/5.  
| “Final”  | Final draft of amplicon analysis manuscript due.  
| 5/5      |  

**Suggested readings on sequencing technology, bioinformatics for managing sequencing bias:**  
• Fuller_2009_Challenges_sequencing_by_synthesis  
• Goodwin_2016_10yrs_nexgen_seq_tech  
• Kozich_2013_developing_Illumina_pipeline  
• Dudley_2009_developing_bioinformatics_skills  
• Schloss_2011_reducing_sequencing_artifacts_16S  

**Suggested readings on 16S, whole genome, and metagenomics:**  
• Martinez-Porchas_2017_how_conserved_is_16S  
• Kim_2011_comparing_16S_variable_regions  
• Marston_2013_NGS_viral_RNA_genomes  
• Baker_2012_de_novo_genomeAssembly  
• Ayling_2019_metagenome_assembly_with_short_reads  
• Poretsky_2014_16S_vs_metagenomics  
• Laurence_2014_contaminants_metagenomics  

**Suggested readings on microbial species’ definition:**  
• deQueiroz_2005_concept_of_species  
• Reeder_2009_rare_biosphere  
• Xu_2014_who_or_what  
• Robinson_2010_structure_to_function_in_HAM  
• Prosser_2007_ecological_theory_microbial_ecology