BIOL 6750
Microbial metagenomics: analyzing environmental sequence data

Instructor: Dr. Bonnie Waring
Contact: bonnie.waring@usu.edu

Location: BNR 278
Schedule: MWF 8:30 – 10:20 am

Course Objectives
In this course, we will explore current best practices for the analysis of microbial community sequence data through hands-on experience with the QIIME2 pipeline. In addition to developing personalized workflows for sequence analysis, course participants will explore the evidence-based foundations of these best practices through literature discussion and review. By the end of the course, students will:

1) Gain familiarity with the QIIME 2 workflow and its data types, methods, visualizers, and pipelines
2) Understand the advantages and drawbacks of different bioinformatics methods for the analysis of microbiome sequence data
3) Practice basic analyses related to microbial diversity and community composition

Learning Objectives (USU IDEA Center)
1. Learn fundamental principles, generalizations, and theories related to analysis of environmental sequence data
2. Apply course materials through hands-on experience with a state-of-the-art bioinformatics pipeline
3. Learn to analyze and critically evaluate ideas, arguments and points of view

Course pre-requisites
- Basic familiarity with the fundamental principles of molecular genetics
- Experience with the R programming environment is strongly suggested but not required
- Course participants are encouraged to bring their own raw sequence data, but this is not required

Course Materials
You will have access to all assigned readings through the course Canvas website.

Important Note: You must have the QIIME2 Core 2018.X distribution installed on your computer before beginning the class. An easy-to-follow tutorial can be found here: https://docs.qiime2.org/2018.11/install/

QIIME 2 does NOT support Windows, so you will need to use a Virtual Machine (VirtualBox, Amazon Web, or Docker), described here: https://docs.qiime2.org/2018.11/install/virtual/

We will not devote class time to troubleshooting QIIME2 installation. Please contact the instructor with any installation questions before class starts!
Course Structure
This intensive seven-week course meets for two hours, three times a week. In the first hour of each class, we will explore key concepts in environmental metagenomics through a brief lecture and a group literature discussion. Although all participants are expected to read all assigned papers, each journal article will be assigned to a single ‘discussion leader’ who is expected to provide a brief critical summary of its content to the rest of the class.

In the second class hour, the instructor will guide students in the development of QIIME2 scripts for processing raw sequence data (FASTQ files) into Amplicon Sequence Variant (ASV) or Operational Taxonomic Unit (OTU) tables, which can then be subjected to further analysis in the R programming environment.

Grade Breakdown and Policy
In-class participation through group discussions and paper presentations will comprise 40% of your overall grade. Additionally, you will develop personalized QIIME2 scripts for the analysis of sequence data. Finally, you will also be required to write a short literature review and synthesis about a course-related topic of your choice; this assignment counts for 35% of your grade.

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Weight (% of Grade)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class participation (group discussions)</td>
<td>40</td>
</tr>
<tr>
<td>QIIME2 exercises</td>
<td>25</td>
</tr>
<tr>
<td>Literature review and synthesis</td>
<td>35</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Final course grades are based on the following scale and will not be rounded up to the nearest whole number: A (94-100%), A- (90-93.9%), B+ (87-89.9%), B (84-86.9%), B- (80-83.9%), C+ (77-79.9%), C (74-76.9%), C- (70-73.9%), D+ (67-69.9%), D (64-66.9%), D- (60-63.9%), F (below 60%).

How will class participation be graded?
It is expected that you will read all assigned papers each week, and be able to actively contribute to the class discussion of these papers. Here are some tips for effective class participation:

If you are leading a discussion…
- Take no more than 5-10 minutes to present an assigned paper to the rest of the class. You do not need to provide an exhaustive overview of every detail, nor should you just reiterate what everyone has read. Your main goal is to promote discussion: by linking your assigned paper to the larger themes of that week’s readings, and by providing your own critical evaluation (be it positive or negative) of the paper.
- Come prepared! Make a list of questions to promote discussion within the group (more than you think you will need).
- Listen! Part of leading a discussion is knowing when to direct the conversation, and when to let it flow. Let the conversation develop organically – I don’t expect
you to *dominate* the discussion by giving an in-depth analysis of every technique/analysis/finding. I do expect you to be familiar enough with the paper to *guide* discussions on these topics.

*If you are participating in a discussion…*

- Discussing a paper does not necessarily mean that you need to find its flaws, or point out where the authors may be wrong. Of course, if you find an issue or inconsistency, it is totally appropriate for you to bring this up! However, you may also point out areas where the authors excelled, or aspects of the paper you particularly liked. In other words, ‘paper discussion’ and ‘paper criticism’ are not synonymous.
- If you find some aspect of the paper to be extremely challenging, it is fine to ask for clarification from the rest of the group. Asking questions also counts as ‘discussion.’ However, I do expect you to make an effort – if there is a recurring, important concept that you are not familiar with, explore the literature on your own to provide context (I am happy to guide you.)

*How will QIIME2 exercises be graded?*

Your QIIME2 analyses will begin with sequence data and end with an OTU/ASV table that can be subjected to ecological analysis. Although the inputs and outputs to the QIIME 2 pipeline will always have the same format, there are many decisions to be made as your sequences are being processed. The goal of this course is to guide you through the process of making those informed decisions. Your work in QIIME will be evaluated on the extent to which your analyses are grounded in an evidence-based decision-making framework, guided by course readings.

**Course Schedule**

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
<th>Programming Task</th>
<th>Papers</th>
</tr>
</thead>
</table>
| 1    | Next-generation sequencing: principles and platforms | Familiarization with QIIME2 workflow and terminology | Caporaso et al. 2010 *Nat Meth*
Mardis 2013 *Ann Rev An Chem*
Van Dijk et al. 2014 *Trends Genet*ics |
| 2    | Design of NG sequence libraries | De-multiplexing | Kozich et al. 2013 *Appl Env Micro*
Smith and Peay 2014 *PLoS One*
Kim et al. 2017 *Microbiome* |
| 3    | NG sequencing errors and how to address them | Denoising and chimera removal | Bokulich et al. 2012 *Nat Meth*
Callahan et al. 2016 *Nat Meth*
Amir et al. 2017 *mSystems* |
| 4    | OTUs vs. ASVs: should sequence reads be clustered? | OTU clustering | Kopylova et al. 2016 *mSystems*
Nguyen et al 2016 *NPJ Biofilms*
Callahan et al. 2017 *ISME* |
| 5    | Taxonomic assignment and analyses | Taxonomy assignment | Werner et al. 2012 *ISME*
Cole et al. 2013 *Nucl Acid Res*
Bokulich et al. 2018 *Microbiome* |
| 6    | Analyzing the diversity of | Rarefaction and diversity estimation | Lundin et al. 2012 *Env Micro Rep*
Gihring et al. 2012 *Env Microbiol* |
Analyzing the composition of microbial communities

Community analyses and visualization

<table>
<thead>
<tr>
<th>Microbial communities</th>
<th>Morris et al. 2014 Ecol Evol</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Analyzing the composition of microbial communities</td>
</tr>
<tr>
<td></td>
<td>Knight et al. 2018 Nat Rev Microbiol</td>
</tr>
</tbody>
</table>

The Honor System and Plagiarism

To enhance the learning environment at Utah State University and to develop student academic integrity, each student agrees to the following Honor Pledge: "I pledge, on my honor, to conduct myself with the foremost level of academic integrity." A student who lives by the Honor Pledge is a student who does more than not cheat, falsify, or plagiarize. A student who lives by the Honor Pledge:

- espouses academic integrity as an underlying and essential principle of the Utah State University community;
- understands that each act of academic dishonesty devalues every degree that is awarded by this institution;
- is a welcomed and valued member of Utah State University.

Plagiarism includes knowingly "representing, by paraphrase or direct quotation, the published or unpublished work of another person as one's own in any academic exercise or activity without full and clear acknowledgment. It also includes the unacknowledged used of materials prepared by another person or agency engaged in the selling of term papers or other academic materials." The penalties for plagiarism are severe. They include warning or reprimand, grade adjustment, probation, suspension, expulsion, withholding of transcripts, denial or revocation of degrees, and referral to psychological counseling.

Violation of the Honor System: If you are found to be cheating on an exam or written assignment, you immediately forfeit your grade on that assignment. There are no exceptions to this policy.

Students with Disabilities

The Americans with Disabilities Act states: "Reasonable accommodation will be provided for all persons with disabilities in order to ensure equal participation within the program. If a student has a disability that will likely require some accommodation by the instructor, the student must contact the instructor and document the disability through the Disability Resource Center (797-2444), preferably during the first week of the course. Any request for special consideration relating to attendance, pedagogy, taking of examinations, etc., must be discussed with and approved by the instructor. In cooperation with the Disability Resource Center, course materials can be provided in alternative format, large print, audio, diskette, or Braille."