

**Environmental Metabarcoding: an introduction to high-throughput
community sequencing and bioinformatics
PLP 6905
Spring 2019**

Course Instructors

Michelle A. Jusino, PhD
Office: 2531 Fifield Hall
Email: mjusino@ufl.edu

James Skelton, PhD
Office: 321 Newins Zeigler
Email: skelto3@ufl.edu

Course Time / Location

This course will begin on February 4th, 2019 and run through February 27th, 2019.
Mondays and Wednesdays, 1:55 PM - 3:50 PM (periods 7 and 8)

Location: 2306 Fifield Hall

Office Hours

Jusino: Mondays & Wednesdays 4p - 5p
Skelton: Mondays 10a - 12p

Course Description:

This course is intended to provide graduate students and postdocs the training necessary to successfully complete projects utilizing high throughput amplicon sequencing (HTS) of biological communities. HTS is powerful and increasingly used, but many studies suffer from avoidable methodological pitfalls. This course will help researchers get the most powerful and accurate results possible from their projects, while avoiding common and costly mistakes and misconceptions. During this course, students will: 1) get an in-depth overview of all of the steps required to sample biological communities using HTS, 2) learn to bioinformatically process HTS community data using the AMPtk pipeline, and 3) learn how to perform preliminary community analyses and data visualizations using R.

Course Objectives:

- 1) Introduction to methods and common pitfalls for high throughput sequencing (HTS) of biological communities
- 2) Gain competency in basic bioinformatics software for processing HTS data
- 3) Learn R-based multivariate statistical procedures for hypothesis testing using HTS output

Materials to bring to class:

It would be helpful to bring a laptop to all class periods. If you do not have one, or cannot bring yours to each class, we will work out a solution.

Required Text

Most of our readings will be from primary literature, so there is not a required text. We will have a few readings from a book that can be ordered from the publisher if you would like, but it is not required. We recommend ordering through the publisher rather than from a different site as prices are highly variable.

Book: Analysis of Ecological Communities. McCune and Grace. 2002

<https://www.wildblueberrymedia.net/store/analysis-of-ecological-communities>

Grading

Grading will be based on in-class participation and a group project.

Participation: 60%

Group project: 40%

Attendance:

This is a participation-based class, so attendance is mandatory. If you have an excused absence, you may make up missed class activities. In order to actively participate in class, you should 1) complete the readings required before class, or complete other appropriate preparation, 2) be on time, and 3) actively participate in class.

Course Schedule (subject to change)

All reading assignments are to be completed before attending class

Class 1 (Feb. 4): Introduction to community sequencing (lecture and discussion)

Readings:

Nilsson, R.H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P. and Tedersoo, L., 2018. Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nature Reviews Microbiology*.

<https://www.nature.com/articles/s41579-018-0116-y>

Lindahl, B.D., Nilsson, R.H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjølner, R., Kõljalg,

U., Pennanen, T., Rosendahl, S., Stenlid, J. and Kauserud, H., 2013. Fungal community analysis by high-throughput sequencing of amplified markers—a user's guide. *New Phytologist*, 199(1), pp.288-299.

<https://doi.org/10.1111/nph.12243>

Optional additional readings: Nguyen et al. 2015. *New Phytologist*

Group project assignment handed out.

Class 2 (Feb. 6): What is bioinformatics? An introduction to bioinformatics pipelines and the UF computing cluster (AMPTk, Python, and hypergator)

Readings:

Palmer JM, Jusino MA, Banik MT, Lindner DL. 2018. Non-biological synthetic spike-in controls and the AMPTk software pipeline improve mycobiome data. PeerJ 6:e4925
<https://doi.org/10.7717/peerj.4925>

Class 3 (Feb. 11): How to process raw sequence data: An introduction to demultiplexing, clustering, and turning sequences into operational taxonomic units (OTUs). Guest lecture / discussion leader: Dr. Craig Bateman
First AMPTk lab exercise.

Readings:

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. and Holmes, S.P., 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nature methods, 13(7), p.581.
<https://www.nature.com/articles/nmeth.3869>

Jusino, M.A., Banik, M.T., Palmer, J.M., Wray, A.K., Xiao, L., Pelton, E., Barber, J.R., Kawahara, A.Y., Gratton, C., Zachariah Peery, M. and Lindner, D.L., 2018. An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. Molecular ecology resources.
<https://doi.org/10.1111/1755-0998.12951>

Class 4 (Feb. 13): Making the OTU table: Sequence filtering, LULU, and taxonomy assignment.

Readings: LuLu paper, Optional reading: FUNGuild paper (?)

Class 5 (Feb. 18): I have an OTU table, now what??? Introduction to multivariate analyses and a brief introduction to R

Readings: *note these chapters will be handed out during class 4
McCune and Grace Chapters 2 and 6: Matrices and Distance Measures
McCune and Grace Chapter 13: Introduction to Ordination
McCune and Grace Chapter 16: Nonmetric Multidimensional Scaling

Class 6 (Feb. 20): More on R and multivariate statistics

Readings:

To be determined based on progress of the class.

Class 7 (Feb. 25): Other issues, and an open discussion about what comes next

Optional readings: Lofgren et al. 2018. *Molecular Ecology*, Tedersoo et al. 2018.
Molecular Ecology Resources

Class 8 (Feb. 27): Presentations