# Incorporating community-level 16S rRNA gene sequencing of local rivers into a semester-long undergraduate course: benefits and challenges

### Emma Wear, Lewis & Clark College, Portland, OR emmawear@lclark.edu

### Summary

- Summer field courses focused on microbial ecology for advanced students often include 16S rRNA gene amplicon sequencing, with varying emphasis from sample collection and preparation to bioinformatics analysis.
- This poster discusses successes and challenges of incorporating this technique into a traditional, semester-long undergraduate microbial ecology course at a small liberal arts college.
- Hands-on exposure to modern, molecular microbial ecology is valuable for the students, most of whom will go on to other subfields of biology but will nonetheless benefit from familiarity with next-gen sequencing and basic bioinformatics. An advantage of a semester-long course is that students can take their own samples from collection and wet-lab preparation through analysis.
- However, the more general audience of an undergraduate class on a college campus presents challenges in course design. In practice, across multiple iterations of this material, students without a strong ecology background have struggled with relating community composition to environmental variability.

# Methods

Sampling sites were selected to favor high physicochemical contrasts within reasonable proximity to campus.

Students conducted the physical preparation of samples:

- Water collection and filtration of samples from the Willamette River
- DNA extraction using the ZymoBIOMICS DNA Miniprep Kit
- PCR amplification using standard 16S rRNA gene primers (515F-Y and 926R)
- Amplicon visualization and clean-up with agarose gels, followed by quantitation with a NanoDrop spectrophotometer

Willamette Riv Sites sampled in: 2022 and 2023 2022 2023

Samples were sequenced at a commercial facility using MiSeq PE250 chemistry. Resulting sequences were clustered into amplicon sequence variants using the DADA2 algorithm applied through Nephele (https://nephele.niaid.nih.gov), including chimera removal and taxonomic classification using the Silva database. Non-target sequences including chloroplasts and mitochondria were manually removed.

Students were provided with quality-controlled, normalized relative abundance data for analysis.

### Acknowledgments

The foundation for the R tutorials comes from the Skills for Teaching Data Sciences workshop (University of Arizona and Lewis & Clark College), with particular thanks to Jeremy McWilliams and Ethan Davis. I thank Lindy Gewin for help with lab logistics. DNA sequencing and supplies were funded by the LC Biology Department. R figures were produced using the packages vegan and phyloseq, by Microbial Ecology students Palmer Bassett and Madeline Jones.

The example data used in the R tutorial are from: Wear et al. 2021, https://doi.org/10.3389/fmars.2021.634803



### Multivariate statistics and data visualization

Students completed two different labs in which they used R packages to analyze and visualize community composition data.

• In the first analysis-focused lab, students were provided with previously published relative abundance data from open-ocean water column and sediment samples. Using published samples ensured that the students saw predictable, interpretable patterns as they worked through a series of guided exercises, from calculating Shannon diversity indices to generating NMS ordinations and stacked bar charts, using an R Markdown file:



sample. plot\_bar(gp.cyano) rochlorococcus\_MIT9313 which they were collected. Tychonema\_CCAP\_1459-11B

Cyanobium PCC-6307

Synechococcus\_CC9902

Left: Stacked bar chart of *Cyanobacteria* relative abundances produced by P. Bassett, class of 2022, with sample type annotations added for clarity. Right: screenshot of R Markdown instructions walking students through the example data visualization and interpretation.

• In the second analysis-focused lab, students were provided with quality-controlled relative abundance data from class-generated sequences. They were prompted with questions guiding them through visualizing and interpreting the data and were given example R scripts but not exact code:

Left: NMS ordination of the six communities, using Bray-Curtis dissimilarity. Right: stacked bar chart of relative abundance of Archaea across the six samples. Both figures produced by M. Jones, class of 2025.







### Advantages of the semester-long context for a general undergraduate audience

- encounter such wet-lab focused techniques.
- themselves to their own data.

# Small-school challenges that can be addressed

- number of samples and limited budgets.

# **Disadvantages presented by the semester-long format**

- time in the field.
- information included in the course.



Above: The Sandy River, approximately half of which is regulated as a National Wild and Scenic River, and which has its headwaters in the glaciers of Mt. Hood. Right: The Columbia River off of Sauvie Island, a major shipping thoroughfare.



**Poster #** 

In preparing the physical samples themselves, students get hands-on practice with molecular tools, from micropipettes and quantitation instruments to standard commercial extraction kits, and with protocols for techniques like PCR. This is particularly valuable with a flexible biology curriculum where ecology-focused students might not otherwise

The semester-long format allows sufficient time for sequencing, so that students can analyze the samples they collected and prepared.

The longer course duration also allows for lecture and lab activities to reinforce one another. For example, multivariate statistics and visualization approaches like ordinations are valuable tools across fields. In a full-length course, students can learn to interpret these figures and statistics while reading published papers and then apply the tools

A major challenge at a small school without a subsidized DNA sequencing facility is the **cost of community-level sequencing**, especially when compared with the cost of more familiar Sanger sequencing. • This challenge is improved by the recent increase in commercial NGS sequencing services. For example, Genewiz's Amplicon-EZ service allows for independently submitted samples to be bundled into a MiSeq run for < \$100 per sample. While bundling samples into a full run is of course more economically efficient per sample, this approach accommodates a small

Students have struggled with relating the community composition results to variability across the sites from which they were collected, even with intentional sampling from contrasting sites. To an extent, this is a downside to a semester-based course, in which lab schedules are set into constrained blocks of time. At least in an urban environment like Portland. many sites can take a prohibitively long amount of time to access from campus, and so students lack an in-person understanding of some ecosystems sampled. In a short course, students would be less likely to have the scheduling constraints of jobs and extracurriculars limiting their

An on-campus course also attracts a more mixed, less specialized student population than an in-depth field course, with correspondingly variable experience in interpreting environmental data. While not a disadvantage per se, this does necessitate careful consideration of the background

