

Analysis of Microbiomes Using Free Web-Based Tools in Online and In-Person Undergraduate Science Courses

Anna J. Zelaya^{1*}, Nicole M. Gerardo², Lawrence S. Blumer³, and Christopher W. Beck²

¹Department of Biology, California State University, San Bernardino, CA, USA

²Department of Biology, Emory University, Atlanta, GA, USA

³Department of Biology, Morehouse College, Atlanta, GA, USA

Abstract

Our understanding of microbiomes, or the collection of microorganisms and their genes in a given environment, has been revolutionized by technological and computational advances. However, many undergraduate students do not get hands-on experiences with processing, analyzing, or interpreting these types of datasets. Recent global events have increased the need for effective educational activities that can be performed virtually and remotely. Here, we present a module that introduces STEM undergraduates to the bioinformatic and statistical analyses of bacterial communities using a combination of free, web-based data processing software. These lessons allow students to engage with the studies of microbiomes; gain valuable experiences processing large, high-throughput datasets; and practice their science communication skills. The lessons presented here walk students through two web-based platforms. The first (*DNA Subway*) is an easy-to-use wrapper of the popular QIIME (pronounced “chime”) pipeline, which performs quality control analysis of the raw sequence data and outputs a community matrix file with assigned bacterial taxonomies. The second, *ranacapa*, is an R Shiny App that allows students to compare microbial communities, perform statistical analyses and visualize community data. Students may communicate their findings with a written final report or oral presentation. While the lessons presented here use a sample dataset based on the gut-microbiome of the bean beetle (*Callosobruchus maculatus*), the materials are easily modified to use original next-generation amplicon sequence data from any host or environment. Additionally, options for alternative datasets are also provided facilitating flexibility within the curriculum.

Citation: Zelaya AJ, Gerardo NM, Blumer LS, Beck CW. 2022. Analysis of Microbiomes Using Free Web-Based Tools in Online and In-Person Undergraduate Science Courses. CourseSource 9. <https://doi.org/10.24918/cs.2022.35>

Editor: Charles Hauser, St. Edward's University

Received: 7/21/2020; **Accepted:** 5/4/2022; **Published:** 10/14/2022

Copyright: © 2022 Zelaya, Gerardo, Blumer, and Beck. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Conflict of Interest and Funding Statement: None of the authors have a financial, personal, or professional conflict of interest related to this work.

*Correspondence to: anna.zelaya@csusb.edu

Learning Goals

Students will:

- ◇ be introduced to the microbiological, molecular, and bioinformatic techniques used to study microbiome data.
- ◇ understand the use of model systems to infer general biological relationships.
- ◇ describe the importance of assessing microbial diversity and potential implications for human and environmental health.

Learning Objectives

Students will be able to:

- ◇ describe and perform the steps of data processing and analysis of metagenomic sequences.
- ◇ analyze 16S rRNA gene sequence data using common techniques used in microbiome research.
- ◇ generate and interpret figures communicating microbial community diversity.
- ◇ compare microbial taxonomy and community diversity within and between samples or treatments.
- ◇ communicate findings to their peers via oral or poster presentations and scientific writing (report).

INTRODUCTION

Research on microbiomes, or the collection of microorganisms and their genes within a given environment, has revolutionized our understanding of the impact that microbes have on the ecology and evolution of all living things on earth (1). By sequencing the 16S rRNA ribosomal gene, which is conserved in all bacterial and archaeal taxa, microbiome research has led to a greater appreciation for how microorganisms contribute to human (2) and environmental health (3, 4). For example, within the gastrointestinal tract of humans, it is estimated that about 100 trillion microbial cells reside (5). Studies have shown that these gut-microbiota perform a variety of beneficial functions, including development of a healthy immune system, digestion and absorption of nutrients, and biosynthesis of essential vitamins (5). While the last 10-15 years have seen an increase in the interest of microbiome research, much remains to be understood.

As interest in microbiomes has increased in the last 20 years across academia, government, and industry, there is a growing demand for computational skills required to analyze, process and communicate the type of high-throughput data generated by microbiome research (8). As such, exposing STEM undergraduates to microbiome research and introducing them to the bioinformatic skills required to process and visualize high-throughput datasets has become increasingly valuable (8). Since microbiome research is interdisciplinary in nature, encompassing aspects of molecular biology, microbiology, ecology, phylogeny, and computational sciences, it serves as an excellent topic to engage undergraduates in a variety of STEM disciplines (9).

The reduced cost of next-generation sequencing combined with the increased availability of computational bioinformatic tools for sequence data analysis has led to greater accessibility and opportunity to engage undergraduate students in the study of microbiomes while learning the technical skills required to analyze microbiome data. However, despite the increased availability of tools and resources to incorporate microbiome research in undergraduate laboratories, many challenges remain that limit their widespread incorporation (10). Many institutions lack the computational resources for the bioinformatic analysis of microbiome data. Although these analyses can be done on personal computers, this requires the installation and configuration of specialized software (e.g., QIIME2 running on a virtual machine), which represents a sizable barrier to implementation. In addition, the lack of training in sequence analysis methods for faculty who are not experts in microbiome research poses significant challenges to widespread incorporation of the study of microbiomes in undergraduate courses. Additionally, while it is often ideal for students to analyze their own data, obtained from mentored or course-based research experiences, it is often not possible to do so. Some colleges and universities lack the monetary resources and research programs needed for generation of such data. More recently, with the global outbreak of COVID-19 altering current research programs, it is uncertain when such opportunities will become available again for undergraduates. The current situation facing colleges and universities underscores the need for classroom activities that can resemble undergraduate research opportunities while being performed in online and virtual formats.

Here we present teaching materials for computational data analysis using two free web-based analysis pipelines, *DNA Subway* (12) and *ranacapa* (13), to analyze 16S amplicon sequence data. Additional learning materials (Supporting Files S1, S4-15) were created to facilitate the online transition as a result of the COVID-19 pandemic. *DNA Subway* is a free-to-use web-based wrapper for the popular QIIME2 (12) platform and was produced by the DNA Learning Center at Cold Spring Harbor Laboratory. While QIIME2 is one of the most widely used platforms for metagenomic analysis, installation and navigation may be challenging for novice users, particularly in a classroom context, although lessons that emphasize these skills are available (14) if such learning outcomes are desired. Conversely, *DNA Subway* is an easy to use graphic-user-interface (GUI) that eliminates the need for sophisticated computational knowledge of command-line usage, as well as the infrastructural challenge of installing and updating QIIME2 software and its dependencies. Created in R statistical package, *ranacapa* is a Shiny web app that was designed to address the challenges associated with learning the suite of analysis tools necessary to adequately analyze, visualize, and interpret next-generation DNA sequencing datasets (13).

As a case study, we present how we use these tools as a part of the Bean Beetle Microbiome Project. While the tutorials presented here use 16S amplicons generated from the bean beetle microbiome as the sample dataset, *DNA Subway* and *ranacapa* may be used to analyze any set of 16S amplicon sequences. Instructors are encouraged to use their own datasets if they are available. Otherwise, a list of freely downloadable datasets from various other host systems is available (see S1. Analysis of Microbiomes - Table of next-generation sequence datasets). These freely available datasets provide research alternatives that increase the flexibility of the microbiome systems and topics that may be studied in undergraduate classes.

The Bean Beetle Microbiome Project CURE was designed as a full semester guided-inquiry research experience where students performed steps of the scientific process from hypothesis generation to final analysis and presentation of findings (Figure 1). While descriptions of the specific aims of the Bean Beetle Microbiome Project have been published elsewhere (11), the sudden transition to online teaching due to the COVID-19 pandemic resulted in the Bean Beetle Microbiome CURE transitioning from an in-person to a fully online experience. Consequently, we developed additional teaching materials for the microbiome data analysis component of the Bean Beetle Microbiome CURE that facilitated the online transition, including newly created assessments that aid in virtual learning of the material (Supporting Files S11, S15). In the Bean Beetle CURE, students analyze next-generation sequence data from the gut microbiome of bean beetles.

Intended Audience

The materials were designed for undergraduates of all levels (freshman to seniors), STEM disciplines, and institution types. As of Spring 2020, some of these materials have been implemented in 8 minority-serving universities implementing the Bean Beetle Microbiome CURE in courses ranging from introductory general biology, cell biology, genetics, and ecology/evolution. The materials were used in introductory,

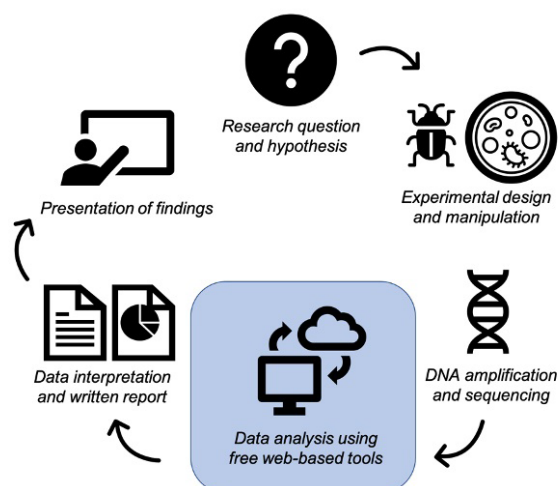


Figure 1. Simplified graphic showing the steps of the bean beetle microbiome CURE. Students begin with an introduction on bean beetles, and through guided inquiry generate a research question and hypotheses. Students design and conduct experiments and collect DNA for amplification and sequencing. Once the sequence data are received, they use free web-based tools to analyze their data (in a shaded box to denote that it is the topic being presented here). After analysis is complete, students interpret their data and write a final report, followed by presentations of their findings. For courses that are not conducting a CURE, the process could be simulated via freely downloadable datasets from various sources (S1. Analysis of Microbiomes - Table of next-generation sequence datasets). The lesson presented here will work with a variety of datasets, allowing for flexibility in the curriculum. Students may still go through the steps of the scientific process, such as hypotheses formation, data analysis, and communications of findings.

lower-division, and upper-division level courses. Additionally, since the computational tools described here utilize a graphical user interface (GUI) and do not require previous experience with coding language using command line or other computational data-science skills, these lessons are easily transferable to non-majors science courses.

Required Learning Time

The lessons presented here prepare students for the use of the web-based software of *DNA Subway* and *ranacapa* using a small subset of authentic microbiome sequence data. In-class time required to cover the lessons is approximately 6 hours, which could be split between two 3-hour laboratory periods. For example, the first 3-hour lesson could act as an introduction to microbiomes and their significance, the data sets that will be used in class, the experimental design and research question to be asked, hypothesis formation, and introduction to *DNA Subway* and *ranacapa* as tools for sequence analysis. While students will likely be able to begin their *DNA Subway* analysis in this first lesson, their analysis may not be complete by the end of the session. Therefore, out-of-class time to complete worksheets and *DNA Subway* analysis may be required. Since the analysis speed of *DNA Subway* often depends on the number of users on the site at any given time, completion times for this assignment may vary. For example, *DNA Subway* analysis can be completed in as little as 2 hours with small datasets (e.g., less than 10 samples) and little web-host traffic. However, students have reported up to 4 days for completion of *DNA Subway* assignment due to multiple users and larger datasets. For this reason, it may be best to provide a full week for completion of the *DNA Subway* activity. The speed of the *ranacapa* analysis is less susceptible to number of users using

the application, as such this assignment can be done during a second 3-hour class period. If written reports or presentations will be used as final assessment for the activities, those may be addressed in a third class session.

Prerequisite Student and Instructor Knowledge

Students should have a basic working knowledge of how to use computers (e.g., file and folder creation, data input and processing using Microsoft Excel or Google Sheets). Also, they should feel comfortable using graphical user interfaces (GUI) and basic web-browser navigation. The authors and our collaborators have successfully implemented this curriculum in advanced and introductory undergraduate courses. Implementation in introductory courses is facilitated by providing mini-lectures on the bean beetle life cycle and the biology of microbiome-animal interactions (see Supporting Files S2, S3). In addition, we assigned reading one or two published studies on microbiome-host interactions prior to discussing new research questions with students. Students completed a one-page worksheet (see Supporting Files S4, S5) that provided questions to discuss in class and helped focus on next questions. Just prior to having students conduct a community ecology analysis of microbiome data, we reviewed the basic metrics for alpha and beta diversity in a mini-lecture (see S6. Analysis of Microbiomes - Review of community ecology analyses). Most of these presentations and scaffolding activities could be used as is or with slight modification if students are analyzing microbiome data from other systems (S1. Analysis of Microbiomes - Table of next-generation sequence datasets).

These scaffolding activities facilitated the successful completion of microbiome analysis by students in their first college-level biology laboratory course. In introductory laboratory courses, we perform the *DNA Subway* processing as a demonstration using a small sample dataset and then ask students to perform the same processing on their own data. Similarly, we demonstrate the community ecology analysis of a sample dataset (the output of the *DNA Subway* processing of demonstration data) using *ranacapa* and then ask students to perform an analysis on their own data.

SCIENTIFIC TEACHING THEMES

Active Learning

Throughout the lesson, students perform steps of the analysis along with the instructor. Students remain engaged at each step of data processing and analysis. With each step, students may ask conceptual questions regarding decision making for each step of the sequence analysis. This lesson has the goal of preparing students to conceptually understand the steps involved in analyzing and interpreting amplicon sequence data and graphical output summarizing the analysis and results. This lesson can be used with microbiome sequence data generated by students as part of a longer laboratory study. However, in an online course or a course in which instructors are introducing students to the analysis of microbiome data rather than having them generate their own sequence data, sequence datasets from a variety of host and environmental sample types are available elsewhere (see S1. Analysis of

Microbiomes - Table of next-generation sequence datasets). In addition, sample datasets from the [Bean Beetle Microbiome Project](#) are available online.

Assessment

To meet the learning objectives, 4 written tutorials and 2 video tutorials may be assigned and graded as complete or incomplete (See Timeline of *DNA Subway* and *ranacapa* lessons). The written tutorials have been included here as Supporting Files S7-S10, and the links for the video tutorials can be found in the corresponding written tutorial. One of the tutorials (S10. Analysis of Microbiomes - Community analysis with *ranacapa*) contains conceptual questions that can be treated as a low-stakes assignment intended to encourage students to research the tool and explain their understanding in writing. Example answers for conceptual questions are included in S11. Analysis of Microbiomes - Rubric for Community analysis with *ranacapa* questions. There is also an optional quiz that instructors may assign to assess understanding of the major steps involved in sequence processing using *DNA Subway* (S15. Analysis of Microbiomes - Quiz on *DNA Subway* tutorial). At the completion of the module, instructors may assign a written/formal report, oral presentation, or both as graded assignments. Reports and presentations should include background and rationale for the samples being compared, results of community analysis, conclusions about implications of differences in diversity between samples, and future directions. Example rubrics for assessing students' oral and written presentations are also available (see Supporting Files S12, S13).

Inclusive Teaching

The written and video tutorials presented here are designed to provide students with visual materials they can reference at their own pace. Most of the assignments are low stakes to help minimize student stress and to facilitate student learning. By using free, web-based bioinformatics tools that have a GUI format, we aim to increase the accessibility and comfort level of students towards conceptual understanding of the material. Students are also allowed ample time outside of class to complete assignments. Students may perform the analyses alone if they choose, or in groups of 2-4 students to encourage discussion and peer mentoring. Additionally, as this entire lesson plan may be performed online, these activities facilitate remote learning.

The web-based tools presented here are well suited for remote inclusive teaching because they do not require high-end or expensive computer systems, nor do they require high-speed internet. Although microbiome datasets can be quite large, students do not need to upload or download any large datasets or files, as all file transfers are done virtually via the CyVerse infrastructure connected to *DNA Subway*. Similarly, all computational processes performed by *DNA Subway* are run virtually via the CyVerse infrastructure, decreasing the dependence on reliable internet connectivity (once a job is started it runs regardless of whether one is connected to the internet or logged onto *DNA Subway*). It also eliminates the need for large memory or processing requirements of personal computers. In fact, the analysis performed by *DNA Subway* can be successfully completed using a smartphone with a web browser, which would allow students without a computer or internet access to complete this activity.

We also acknowledge the barriers that remain for many students and institutions relating to lack of adequate technology or internet connectivity, and that each college or university tackles such challenges differently. We also acknowledge that there is no single answer that will accommodate all students and institutions. For educators who may be interested in incorporating this activity with their students, but who are concerned about the lack of access to technology for their students, one possible alternative may be for the instructor to perform the *DNA Subway* and *ranacapa* analyses that their students recommend (based on initial discussion and hypothesis generation step) and print the resulting graphs and figures of each step performed along the way to show students the process. These hardcopies may then be disseminated to students, where they can follow the written tutorial and learn to interpret each of the graphs. In this way, students may still be engaged with the data analysis and graph interpretation process, even if they did not go through the steps on a computer themselves. In this case, the learning outcomes would emphasize graph interpretation and communication of results as opposed to the technical aspects of using the web tools.

LESSON PLAN

Classroom Setup

If performing activities in a classroom environment with computers, the instructor should ensure internet connectivity is available and at least one computer with internet access per group. Students may work alone or in groups of 2-4 students. A computer lab would be ideal for these activities, where each student can have their own computer or where groups of students can share one computer. Additionally, a classroom where one computer is attached to a large monitor or projector where the instructor can demonstrate certain steps would be beneficial. An instructional assistant or peer helper who can freely walk around the classroom to help students troubleshoot problems is highly beneficial.

Teacher Preparation

Prior to the start of the activity, instructors will need to obtain sample microbiome datasets. Instructors may use their own datasets if available, or they may download freely available datasets online (S1. Analysis of Microbiomes - Table of next-generation sequence datasets). It is important to note that while sequence files may come in a variety of file formats, as of the time of this writing *DNA Subway* only accepts FASTQ files that are in Casava Illumina 1.8 file format. The files may contain either single-end or paired-end reads. If the files are in a different format, they will need to be converted to the above format to be processed via *DNA Subway*. The sequence provider may offer alternative options or resources to either convert the files or otherwise obtain the data in the appropriate format. Additionally, instructors may refer to the [online tutorial for *DNA Subway*](#) for any updates related to file formatting or other pipeline issues.

Before the first class session, instructors will want to prepare a brief mini-lecture that explains what microbiomes are and why they are important to study (for examples, see S3. Analysis of Microbiomes - What is a microbiome). It would also be important to introduce necessary information related to the datasets that will be analyzed by students. For example, it will be

helpful for students to know the source of the samples, aspects of experimental design such as treatments administered or environmental conditions, and the ultimate research question or topic of interest. The worksheets and handouts provided walk students through a sample dataset from bean beetles fed on different bean types. If this dataset will be used for the activity, instructors may refer to information presented in S14. Analysis of Microbiomes - Does the composition of the Bean Beetle microbiome depend on bean species. Once students have been introduced to microbiomes and information related to the samples they will be analyzing, instructors may use a guided-inquiry approach to help students think about what possible differences they may see among the various samples, and eventually generate hypotheses that they may then investigate as they perform their data analysis.

Instructors should run through the entire set of tutorials ahead of time before the beginning of each lesson. Beginning with S7. Analysis of Microbiomes - CyVerse tutorial, instructors must create their own CyVerse account and upload all necessary files to their home library in the CyVerse Data Store. Instructors that do not have their own sequence datasets may use sample datasets (see S1. Analysis of Microbiomes - Table of next-generation sequence datasets). These sample datasets must be uploaded to the CyVerse Data Store prior to laboratory lesson #1. Additionally, it is recommended that instructors share the data sets with their students via the CyVerse Data Store. Within the Data Store, datasets can be shared between users with CyVerse accounts. Therefore, students should create their CyVerse accounts prior to the start of the first laboratory lesson with ample time for instructors to share datasets with students (see Student Pre-Lesson Activity section below).

Student Pre-Lesson Activity

As an out-of-class assignment prior to the start of laboratory lesson #1, students should set up accounts using CyVerse using the “CyVerse Tutorial” tutorial to ensure that they can be ready to begin during the class period. This tutorial guides students through the steps necessary to open their CyVerse accounts, which will allow them to gain access to the sample data set and the *DNA Subway* application via CyVerse. This activity should ideally be done at least one week prior to the start of the in-class teaching lesson to ensure that all students have working accounts and work out any difficulties prior to start of the in-class teaching lesson. CyVerse automatically sends a confirmation email when a new account is initiated. It is necessary to respond to the confirming email for the CyVerse account to become active.

In-Class DNA Subway Lesson

Prior to starting the *DNA Subway* lesson, a brief introduction to microbiomes and their importance, the datasets and samples being investigated, the research question that is being addressed, and the tools that will be used to analyze the dataset will help to set the stage for the *DNA Subway* and *ranacapa* activities. As mentioned above, instructors may opt to guide students through the process of generating hypotheses for the research question. Once students have been introduced to the topic and have all their materials, they are ready to begin the lesson. If metadata files have been pre-made and shared with students, then they may proceed directly to S8. Analysis of Microbiomes - *DNA Subway* tutorial.

In total there are 5 steps in the *DNA Subway* pipeline. The time needed for each step varies from a few seconds to several hours. These times also vary based on the number of users logged in with actively running processes on the servers. If this activity will be done in class and not as homework, additional activities should be prepared for students to participate in while they wait for processes to finish running. It is likely that the entire analysis will not be completed prior to the end of the day's class meeting. As a result, instructors may want to prepare slides with examples of the outputs of each step where they can discuss how to interpret outputs/results while processes are running. A video tutorial, [which can be viewed here](#), runs through a sample analysis that instructors can use as example data to discuss.

After all steps of the *DNA Subway* lesson have been completed, the final output is an amplicon sequence variant (ASV) count table in csv format classified at the taxonomic level of Family (a Level-5 taxonomy file). This table lists all samples in the dataset and corresponding ASVs present in each sample, as well as the sequence counts contained in each sample. However, the taxonomy file will likely contain sequences that need to be filtered or removed prior to continuing with downstream analysis. For example, mitochondrial and chloroplast DNA that originated from the host sample may have been sequenced and will be present in the final count table. As these DNA are not bacterial, they need to be removed prior to the bacterial community analysis. (Note that this provides an excellent opportunity to ask students why these sequences might occur in the dataset and remind them of the endosymbiotic origins of mitochondria and chloroplasts). This final filtering is described in detail in S9. Analysis of Microbiomes - Preparing files for downstream analysis. The final filtered taxonomy file, along with the appropriately formatted metadata file, are the two files required for further downstream community analysis using *ranacapa*. (Note: It is important to ensure that the metadata file used here is the file formatted for *ranacapa* analysis.)

In-Class ranacapa Lesson

The second in-class session can be dedicated to community analysis using *ranacapa* (S10. Analysis of Microbiomes - Community analysis with *ranacapa*). To begin, students will need the filtered output taxonomy file from *DNA Subway*. In practice, *ranacapa* seems to be less sensitive to multiple users, such that instructors may opt to have each student work individually if desired, or, students may continue to work in their groups.

In total there are 7 steps in the *ranacapa* analysis. A brief lecture prior to the start of the activity may be beneficial to introduce students to the different tabs as well as how to obtain and interpret the output of each analysis. After the mini-lecture, students can begin to work on their own data. Students should work through each tab, download the generated figures, and write a brief interpretation of results either in their laboratory notebooks or electronic notebooks. These figures can comprise much of the results section of a written report or oral presentation. A video tutorial is also available for the *ranacapa* analysis and [can be viewed here](#).

TEACHING DISCUSSION

As of Spring 2020, these lessons have been implemented at 8 minority-serving institutions across the United States and have introduced students with limited previous coding experience to the steps involved in amplicon sequence analysis. In this lesson, the specific dataset that was used focused on processing 16S rRNA gene amplicon sequences from insect samples (bean beetles). The steps outlined in these lessons may use amplicons that target other regions (as permitted by the *DNA Subway* pipeline), however it is important that instructors and students be knowledgeable about appropriate parameters for analysis necessary for the specific amplicons they choose to use. In total, the activities presented are best completed as a combination of in-class and out-of-class activities and assignments. In our previous implementations, these activities comprised a module as part of a Course Based Undergraduate Research Experience. The activities outlined here were completed over the course of two laboratory meetings (that met once a week for 4 hours). Between in-person meetings, students completed some of the activities as out-of-class assignments. Final oral and written presentations were done at the end of the semester. Example grading rubrics are in Supporting Files S12 and S13.

As video tutorials are available that walk students step-by-step through each analysis, students may work on these assignments completely remotely (e.g., outside of a traditional classroom environment). Students may choose to watch the video tutorial or follow the written handout. The end-product of the analysis, a Level-5.csv taxonomy file, may be collected as evidence that students followed the tutorial and completed the processing of the amplicon data. Additionally, a quiz may be administered to assess deeper comprehension of the material (see Supporting Files S11, S15). Instructors may always tailor the provided assignments to address aspects that are specific to the samples being investigated by their students.

Our experiences conducting these activities highlighted that while incoming freshmen may be highly skilled in computers, they may not have much experience with data management, data entry, and spreadsheet manipulation. In our experience, students can fall behind quickly during demonstrative lectures if it is assumed that students are already competent in the above skills and information is presented too quickly. The presence of a knowledgeable and engaged instructional assistant or peer helper can be critical here, as they can move around the classroom freely and check-in with students who may need the additional attention to stay on task. For remote learning, additional on-line office hours may be helpful for students who need to work-out or talk through misunderstandings. For instructors who may wish to begin the *DNA Subway* lesson in class, it is important to note that CyVerse servers slow with the increased traffic of multiple users. While performing the *DNA Subway* portion in class ensures that instructors can check that all starting materials (e.g., metadata files) are properly formatted, the increased traffic may cause delays in typical run times. Therefore, it is critical that instructors have ample material to cover in class while processes are running. As mentioned above, it is likely that students will not finish the *DNA Subway* portion in-class, which would require them to finish from home. On rare occasions, the server is down for maintenance. In such cases, it is best to contact the hosts of *DNA Subway* directly to get

updates (contact information is found directly on the *DNA Subway* website). Analysis using *ranacapa*, however, is less susceptible to traffic limitations. One notable limitation we experienced while using *ranacapa* was the inability for certain statistics (e.g., beta-diversity) to be calculated when only 2 groups were compared for analysis (e.g., 1 treatment and 1 non-treatment). Therefore, additional calculations outside of the *ranacapa* pipeline were necessary for students to report statistically meaningful results. As comparisons between only 2 groups is not uncommon, particularly for course-based instruction, the ability to perform such analyses is beneficial. Currently, we are developing our own Shiny based app that allows for comparisons between 2 groups along with additional features not currently available in *ranacapa*.

SUPPORTING MATERIALS

- S1. Analysis of Microbiomes - Table of next-generation sequence datasets
- S2. Analysis of Microbiomes - Bean beetles as a model system
- S3. Analysis of Microbiomes - What is a microbiome
- S4. Analysis of Microbiomes - Worksheet on reading assignment 1
- S5. Analysis of Microbiomes - Worksheet on reading assignment 2
- S6. Analysis of Microbiomes - Review of community ecology analyses
- S7. Analysis of Microbiomes - CyVerse tutorial
- S8. Analysis of Microbiomes - *DNA Subway* tutorial
- S9. Analysis of Microbiomes - Preparing files for downstream analysis
- S10. Analysis of Microbiomes - Community analysis with *ranacapa*
- S11. Analysis of Microbiomes - Rubric for Community analysis with *ranacapa* questions
- S12. Analysis of Microbiomes - Rubric for written reports
- S13. Analysis of Microbiomes - Rubric for oral presentations
- S14. Analysis of Microbiomes - Does the composition of the Bean Beetle microbiome depend on bean species
- S15. Analysis of Microbiomes - Quiz on *DNA Subway* tutorial

ACKNOWLEDGMENTS

The authors would like to acknowledge the faculty and student participants of the 2019-2020 Bean Beetle Microbiome Project for providing feedback on instructional materials presented here.

REFERENCES

1. Knight R, Callewaert C, Marotz C, Hyde ER, Debelius JW, McDonald D, Sogin ML. 2017. The microbiome and human biology. *Annu Rev Genomics Hum Genet* 18:65-86. doi.org/10.1146/annurev-genom-083115-022438.
2. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight, R. 2018. Current understanding of the human microbiome. *Nat Med* 24(4):392-400. doi.org/10.1038/nm.4517.
3. Compant S, Samad A, Faist H, Sessitsch A. 2019. A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. *J Adv Res* 19:29-37. doi.org/10.1016/j.jare.2019.03.004.
4. Jansson JK, Hofmockel KS. 2020. Soil microbiomes and climate change. *Nat*

- Rev Microbiol 18(1):35-46. doi.org/10.1038/s41579-019-0265-7.
5. Dave M, Higgins PD, Middha S, Rioux KP. 2012. The human gut microbiome: current knowledge, challenges, and future directions. Transl Res 160(4):246-257. doi.org/10.1016/j.trsl.2012.05.003.
6. Douglas AE. 2011. Lessons from studying insect symbioses. Cell Host Microbe 10(4):359-367. doi.org/10.1016/j.chom.2011.09.001.
7. Newton IL, Sheehan KB, Lee FJ, Horton MA, Hicks R.D. 2013. Invertebrate systems for hypothesis-driven microbiome research. Microbiome Science and Medicine 1(1). doi.org/10.2478/micsm-2013-0001.
8. Porter SG, Smith TM. 2019. Bioinformatics for the Masses: The need for practical data science in undergraduate biology. OMICS 23(6):297-299. doi.org/10.1089/omi.2019.0080.
9. Wang JT, Daly JN, Willner DL, Patil J, Hall RA, Schembri MA, Tyson GW, Hugenholtz P. 2015. Do you kiss your mother with that mouth? An authentic large-scale undergraduate research experience in mapping the human oral microbiome. J Microbiol Biol Educ 16:50-60. doi.org/10.1128/jmbe.v16i1.816.
10. Williams JJ, Drew JC, Galindo-Gonzalez S, Robic S, Dinsdale E, Morgan WR, Triplett EW, Burnette III JM, Donovan SS, Fowlks ER, Goodman AL, Grandgenett NF, Goller CC, Hauser C, Jungck JR, Newman JD, Pearson WR, Ryder EF, Sierk M, Smith TM, Tosado-Acevedo R, Tappich W, Tobin TC, Toro-Martinez A, Welch LR, Wilson MA, Ebenbach D, McWilliams M, Rosenwald AG, Pauley MA. 2019. Barriers to integration of bioinformatics into undergraduate life sciences education: A national study of US life sciences faculty uncover significant barriers to integrating bioinformatics into undergraduate instruction. PLoS One 14(11):e0224288. doi.org/10.1371/journal.pone.0224288.
11. Zelaya AJ, Gerardo NM, Blumer LS, Beck CW. 2020. The Bean Beetle Microbiome Project: A Course-Based Undergraduate Research Experience in Microbiology. Front Microbiol 11:577621. doi.org/10.3389/fmicb.2020.577621.
12. Boylen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, AlGhalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Carabeallo-Rodriguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keef CR, Kelley ST, Knights D, Koester I, Kosciorek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Priesse E, Rasussen LB, Rivers A, Robeson II MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hoof JJJ, Vargas F, Vazquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Wills AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME2. Nat Biotechnol 37:852-857. doi.org/10.1038/s41587-019-0209-9.
13. Kanlikar GS, Gold ZJ, Cowen MC, Meyer RS, Freise AC, Kraft NJB, Moberg-Parker J, Sprague J, Kushner DJ, Curd EE. 2018. *ranacapa*: An R package and Shiny web app to explore environmental DNA data with exploratory statistics and interactive visualizations. F1000Research 7:1734. doi.org/10.12688/f1000research.16680.1.
14. Stevens JL, DeHority R, Goller CC. 2017. Using QIIME to Interpret Environmental Microbial Communities in an Upper Level Metagenomics Course. CourseSource 4. doi.org/10.24918/cs.2017.3.

Table 1. Timeline of *DNA Subway* and *ranacapa* lessons.

Activity	Description	Estimated Time	Notes
Instructor preparation prior to Lesson #1			
Create CyVerse Account	Instructors should create a CyVerse account in order to share datasets with students.	10 min	S7. Analysis of Microbiomes - CyVerse tutorial
Upload and share sequence datasets via CyVerse	Instructors will want to upload all sequence datafiles to their CyVerse accounts. These files will then need to be shared with their students for their <i>DNA Subway</i> lesson. If instructors need sample datasets, a list of freely available datasets is available.	20 min	S1. Analysis of Microbiomes - Table of next-generation sequence datasets
<i>DNA Subway</i> Video Tutorial	A video tutorial is available as a resource for instructors who would like to familiarize themselves with the <i>DNA Subway</i> pipeline prior to the in-class lesson.	60 min	S8. Analysis of Microbiomes - <i>DNA Subway</i> tutorial
Pre-Lesson Activity (optional)			
Pre-Lesson Activity #1 for students	Instructors may opt to assign the creation of a CyVerse account and metadata files as pre-lesson assignments to students. This will also allow instructors to share the datasets via CyVerse with students prior to Lesson #1.	30 min	- S7. Analysis of Microbiomes - CyVerse tutorial (optional) S9. Analysis of Microbiomes - Preparing files for downstream analysis
(optional) Pre-Lesson Reading	Instructors may opt to assign reading assignments to students to prepare them to think about the potential research questions relating to microbiomes.	30-60 min per reading assignment of work outside of class	- S4. Analysis of Microbiomes - Worksheet on reading assignment 1 - S5. Analysis of Microbiomes - Worksheet on reading assignment 2
Lesson #1			
Introduction	Instructors may provide a brief background on environmental microbiomes and their system of investigation. If using the sample data set, two slide decks are provided.	15-20 min	(optional) S2. Analysis of Microbiomes - Bean beetles as a model system - S3. Analysis of Microbiomes - What is a microbiome
<i>DNA Subway</i>	Students analyze their next-generation sequence data using <i>DNA Subway</i> . If pre-lesson activities were not assigned, these must be done first (e.g., create their CyVerse accounts, create and upload metadata files, ensure that students can access sequence datasets via CyVerse). This lesson may not be completed in class, and will likely need to be continued as an out-of-class assignment.	Multiple steps required. May take several days to complete the analysis. Typically takes no longer than 3 days, working 10-60 min daily.	- S8. Analysis of Microbiomes - <i>DNA Subway</i> tutorial (optional) DNA Subway Video Tutorial
Pre-lesson #2			
Formatting metadata file for <i>ranacapa</i> analysis	While <i>ranacapa</i> also requires a metadata file, it needs to be formatted slightly differently than the one used for <i>DNA Subway</i> . As a pre-lesson, students can create a new metadata file for <i>ranacapa</i> analysis.	10-15 min	S9. Analysis of Microbiomes - Preparing files for downstream analysis
Lesson #2			
Community Analysis with <i>ranacapa</i>	Students analyze their community data using <i>ranacapa</i> tutorials. A brief lecture or introduction may be appropriate before having students perform analysis.	60 min	- S6. Analysis of Microbiomes - Review of community ecology analyses - S10. Analysis of Microbiomes - Community analysis with <i>ranacapa</i> (optional) ranacapa Video Tutorial

Activity	Description	Estimated Time	Notes
Assessments (optional)			
Quiz and Rubrics	Student content knowledge and skills may be assessed using the accompanying rubrics (for <i>ranacapa</i> analysis, written, and oral presentations) and quiz on <i>DNA Subway</i> .	May be assigned in or outside of class	• S11. Analysis of Microbiomes - Rubric for Community analysis with <i>ranacapa</i> questions • S12. Analysis of Microbiomes - Rubric for written reports • S13. Analysis of Microbiomes - Rubric for oral presentations • S15. Analysis of Microbiomes - Quiz on <i>DNA Subway</i> tutorial