BeanBeetleMicrobiome app: an online app for community analysis of microbiome data

Carolyne Huang[1], Anna J. Zelaya[1], Lawrence S. Blumer[2], Nicole M. Gerardo[1] and Christopher W. Beck[1]

[1]Emory University, Department of Biology, 1510 Clifton Rd., Atlanta GA 30322, USA
[2]Morehouse College, Department of Biology, 830 Westview Dr. SW, Atlanta GA 30314, USA
(carolyne.huang@emory.edu; anna.zelaya@emory.edu; lawrence.blumer@morehouse.edu; ngerard@emory.edu; christopher.beck@emory.edu)

With decreases in the cost of next-generation sequencing and advances in computational approaches, microbiome research has exploded. Consequently, microbiome research has become both feasible and accessible for undergraduate laboratory courses. Datasets from these research studies are ideal for introducing students to microbiology and community ecology in online classes. However, the coding skills necessary to analyze these data represent a substantial barrier to instructors and students. As a result, we developed the BeanBeetleMicrobiome app, an online app for the analysis of microbiome community data. While the app was designed for the Bean Beetle Microbiome Project, family-level taxonomy files from any microbiome project can be analyzed with the app. The app allows students to explore community ecology concepts such as species accumulation curves, alpha diversity, and beta diversity as they relate to microbial communities.

Introduction

Research on microbiomes, or the collection of bacteria, archaea, fungi and viruses residing in a given environment, has grown dramatically in recent years. As a result, microorganisms and their activities have increasingly been recognized as essential for maintaining both human and environmental health (Timmis et al. 2019; Martiny et al. 2020). For example, microbes in the digestive tracts of animals facilitate the breakdown and absorption of key nutrients (Dave et al. 2012). Symbiotic microbes in plants and animals aid in a variety of functions, such as development and defense (Compant et al. 2019; Douglas, 2011). Environmental microbes, with their vast metabolic capabilities and global contributions to the cycling of elements and nutrients, have the potential to mitigate the negative impacts of climate change (Cavicchioli et al. 2019).

Due to the increased interest in microbiomes and the decrease in costs of next-generation sequencing, microbiome research has become the basis for several course-based research experiences. For example, Authentic Research Experience in Microbiology – AREM project

(http://arem.cuny.edu/sites/) investigates natural microbial communities in urban environments. In contrast, the Bean Beetle Microbiome Project focuses on host-associated microbial communities (Zelaya et al. 2020). Students in these courses are involved in extracting DNA from samples, which is subsequently sequenced for the 16s rRNA gene. The resulting sequence data are then processed through a bioinformatics pipeline to identify the bacterial taxa in each sample. These data are then used for community analysis.

Both the bioinformatics and community analysis aspects of these projects can be a substantial barrier for instructors and students. Many sequencing centers will conduct basic bioinformatic and community analysis. However, this approach does not allow students to be involved in the analysis of the data that they collected. The Purple Line on the DNA Subway platform (dnasubway.cyverse.org) was developed by the DNA Learning Center at Cold Spring Harbor Laboratory and provides a user-friendly approach for implementing the QIIME2 pipeline for analysis of microbiome data. In addition, analysis of the resulting community data can be conducted on the Purple Line. However, the Purple Line does not allow filtering of the taxonomic data to exclude unidentified taxa, chloroplasts, mitochondria, or taxa that occur in low abundance. As a result, students need to download taxonomy files from the Purple Line, filter the data, and then analyze the data separately.

While basic community analysis of microbial community data could be done in a spreadsheet program (e.g., calculating species richness or plotting stacked bar graphs to visualize alpha diversity), more sophisticated approaches are needed to fully analyze the data. A variety of packages developed for R help facilitate this analysis. Yet, the coding is non-trivial. Ranacapa is a web-based analysis app for analyzing data from environmental DNA studies that runs using RShiny, a graphical user interface for running R code (Kandlikar et al. 2018). Although ranacapa can be used by students to analyze microbiome community data, it has some limitations in terms of the analyses that are available and the display on some data visualizations. As a result, we developed the BeanBeetleMicrobiome app, using ranacapa as an inspiration, to facilitate the community analysis of microbiome data generated by students participating in the Bean Beetle Microbiome Project. Although the app was developed for the Bean Beetle Microbiome project, it can be used to analyze any microbial community data as long as the taxonomy file and metadata file are formatted appropriately. The app is an RShiny app designed for community analysis of level-5 (family-level) datasets that have been produced in DNA Subway. Before running the app, students must re-format the taxonomy file and create a metadata file, as outlined below.

Student Outline

Objectives
- Analyze and visualize microbiome community data
- Interpret common ecological data and graphs related to diversity

Introduction
Microbiomes are the collection of bacteria, archaea, fungi and viruses residing in a given environment, including in or on other organisms. Research on microbiomes has grown dramatically in recent years. As a result, microorganisms and their activities have increasingly been recognized as essential for maintaining both human and environmental health (Timmis et al. 2019; Martiny et al. 2020). For example, microbes in the digestive tracts of animals facilitate the breakdown and absorption of key nutrients (Dave et al. 2012). Symbiotic microbes in plants and animals aid in a variety of functions, such as development and defense (Compant et al. 2019; Douglas, 2011). Environmental microbes, with their vast metabolic capabilities and global contributions to the cycling of elements and nutrients, have the potential to mitigate the negative impacts of climate change (Cavicchioli et al. 2019).

Research on microbiomes usually begins with the extraction of DNA from microbial community samples. This DNA is then sequenced by next-generation sequencing with primers that are specific to bacteria, but variable among taxa of bacteria, commonly the 16s rRNA gene. The resulting sequence data are run through a bioinformatics pipeline that ultimately identifies the bacterial taxa present in the samples. Researchers then use data analysis techniques similar to those used by community ecologists to explore different measures of diversity in the microbial community.

BeanBeetleMicrobiome is an online app designed for community analysis of level-5 (family-level) datasets produced in DNA Subway (a bioinformatics pipeline for microbiome data). Although this app was designed for

community analysis of data from the Bean Beetle Microbiome Project, it can be used for community analysis of any level-5 data. Before running the app, the taxonomy file needs to be formatted and a metadata file created, following the directions below.

Preparing files for analysis

Creating the metadata file
The metadata file only has the sample IDs and the treatment data.

To create the metadata file:

1. Open a new spreadsheet. The first column should have the header name "SampleID". The cells in the first column should then have all the ID's of the different samples that were run through DNA Subway. The second column should have the header name for the treatment variable. In this example, we use the treatment name "Host". All the cells under the Host column should contain the treatment data for the samples. In our example dataset, the treatments were the different host diets on which the beetles were raised. Your new file should look something like this:

	A	B
1	SampleID	Host
2	ADZU.3	Adzuki
3	ADZU.4	Adzuki
4	BEP.3	BEP
5	BEP.4	BEP

2. After modifying your file, save your file as a comma-delimited file (.csv)
3. It is important to note that the metadata file can have only a single treatment column.

Editing the level-5 file

In DNA Subway, you process raw data that is generated through sequencing of microbial community samples. Before using the BeanBeetleMicrobiome app for microbial community analyses, you need to format the taxonomy file generated by DNA Subway.

1. Once you have your taxonomy file (level-5.csv), open it in Microsoft Excel. It should look something like this, with sample ID's in the first column (rows labels) and taxonomic names as the headers of all the other columns. The numbers in each cell represent the "counts" for each bacterial taxon within each sample.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	index	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria
2	ADZU.3	2651	0	0	0	0	0	0	0	0	12	0	0	C
3	ADZU.4	2439	0	0	0	0	0	0	0	0	5	0	0	C
4	BEP.3	11452	5	0	0	9	6	14	0	0	83	16	14	30
5	BEP.4	4168	0	9	3	0	0	0	5	6	61	7	0	C

2. In Excel, change the header title "Index" in cell A1 and replace it with the name "sampleID"
3. Use your mouse to scroll all the way to the last column of your table. For example, in the table below we scrolled to column CU.

	CK	CL	CM	CN	CO	CP	CQ	CR	CS	CT	CU	CV	CW
ia;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	BarcodeSequ	LinkerPrimer	Host	Description		
0	9	19	0	0	2	0	0			Adzuki	ADZU.3		
0	3	11	0	0	0	2	0			Adzuki	ADZU.4		
0	20	14	4	0	14	0	14			BEP	BEP.3		
3	18	0	0	0	0	0	0			BEP	BEP.4		

4. You will see that there are four extra columns that contain information that was added from your DNA Subway metadata file. We want to delete these columns because they do not contain sequence count data. So, delete these columns entirely by highlighting these four columns. Then, right click to delete each of the unwanted columns. Your sheet may now look something like this:

	CI	CJ	CK	CL	CM	CN	CO	CP	CQ	CR	CS	CT	CU
1	ia; k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	o__Verrucom	crobia;c__	[Pedosphaerae];o__	[Ped
2	05	0	0	9	19	0	0	2	0	0			
3	19	5	0	3	11	0	0	0	2	0			
4	87	45	0	20	14	4	0	14	0	14			
5	99	12	3	18	0	0	0	0	0	0			
6	67	26	0	21	0	0	14	0	0	0			
7	69	0	0	0	8	0	0	0	0	0			
8													
9													
10													

5. Save your file so as to not lose what you have done up to this point.
6. Create a new Excel worksheet by clicking on File, then New.
7. A window containing a new spreadsheet should appear right in front of the level-5 spreadsheet you had just been working on.
8. Click on the level-5 spreadsheet (which might be located behind the new spreadsheet) to bring it back to the front so you can continue working with it.
9. Then, select all of the data and copy it.
10. Next, click on the new spreadsheet (should be located behind your original taxonomy spreadsheet) to bring it back to the front.
11. Click on the first cell A1 of the new spreadsheet so that the first cell is outlined in green.
12. Right click on cell A1. From "Paste Options", paste a transposed version of the table that you just copied into this new spreadsheet. Different versions of Excel do this differently. Regardless of the method you use, the final table should look like the photo below. Make sure that the label for Column A is "sum.taxonomy" as shown.

	A	B	C	D	E
1	sum.taxonom	ADZU.3	ADZU.4	BEP.3	BEP.4
2	k__Bacteria;	2651	2439	11452	4168
3	k__Bacteria;	0	0	5	0
4	k__Bacteria;	0	0	0	9
5	k__Bacteria;	0	0	0	3
6	k__Bacteria;	0	0	9	0
7	k__Bacteria;	0	0	6	0
8	k__Bacteria;	0	0	14	0
9	k__Bacteria;	0	0	0	5
10	k__Bacteria;	0	0	0	6
11	k__Bacteria;	12	5	83	61
12	k__Bacteria;	0	0	16	7
13	k__Bacteria;	0	0	14	0
14	k__Bacteria;	0	0	30	0
15	k__Bacteria;	0	0	9	0
16	k__Bacteria;	79	86	38	40
17	k__Bacteria;	0	0	0	0
18	k__Bacteria;	7	0	0	0
19	k__Bacteria;	0	0	0	0
20	k__Bacteria;	0	0	14	9
21	k__Bacteria;	0	0	0	0
22	k__Bacteria;	0	0	29	0
23	k__Bacteria;	0	0	5	0

13. Notice that this table contains the same information as our original taxonomy table, but it is transposed, meaning that the rows and columns are switched: Sample ID's are now column headers, and the taxonomic names of the bacterial taxa are now located as rows.

14. Hover your mouse to the border between cell A and cell B on the spreadsheet. Click and drag the cell wall border to the right to expand the column, allowing you to view the contents within the cells in column A.

Before   *After expanding column A*

15. At this point, it may be a good idea to save your file to avoid losing any changes. Click on File, then on Save As. Save the files as a comma-delimited file (.csv).

16. Now, read through your taxonomy names. There are a few rows that must be deleted from our dataset.

17. Go to Edit and select the Find option.
    - In the Find What field, type "Unassigned". These are sequences that could not be assigned to a taxon at any level, so we want to remove them from our data sheet.
    - Then click Find Next.
    - If there are any "Unassigned" sequences found, select the entire row for each one found, right click, and select Delete Row.
    - Make sure to delete every row that contains unassigned bacteria.
    - Repeat the process to remove the following types:
        - Archaea: The primers that we use are more specific to bacteria, so we have not completely sampled all the Archaea.
        - Chloroplast: Since chloroplasts are actually bacterial symbionts of plants, they get sequenced in Mi-Seq. Adult bean beetles retain some bean material in their gut, which is the source of chloroplast sequences. Since they are not really part of the beetle microbiome, we want to delete these rows from our file.
        - Mitochondria: Mitochondria are bacterial symbionts of all eukaryotes and are sequenced in Mi-Seq.
    - Make sure to delete every instance of Unassigned, Archaea, Chloroplast, and Mitochondria

18. Save your file. Now, your files are ready to import into the BeanBeetleMicrobiome app. Remember, this file must be in CSV format.

19. If you are using Excel, save an Excel format copy of this file using Save As and selecting the .xlxs extension so you can perform community analysis in the excel spreadsheet to calculate the Alpha Diversity values (see section on Alpha Diversity).

Community Analysis with the BeanBeetleMicrobiome app

- Go to https://beanbeetles.shinyapps.io/BeanBeetleMicrobiome/
- From the Welcome! Homepage, you will see a series of tabs. Each of these tabs performs different steps of community analysis and should be visited in sequential order.

Step 1: Import your data into the BeanBeetleMicrobiome app

Click on the Data Upload tab. Using the Browse buttons, select your transposed level-5 taxonomy CSV file and your metadata CSV file created in the previous steps. Then, click the Run App button. Your files will upload and display. Double check the tables to make sure that they look correct before proceeding.

Step 2: Core taxa

Click on the Core Taxa tab. Core taxa are taxa that are found in all samples. The table generated in this tab shows the core taxa found in all samples listed in descending order from the most to least abundant. The values are the abundances of the core taxa found in each sample. The last column is the abundance across all samples. The caption at the top of the table tells you how many core taxa there are relative to the total number of taxa. You can select the taxonomic level of interest using the dropdown menu on the left.

Questions
1. How many taxa are in the core taxa at each taxonomic level?
2. Does the proportion of the taxa that are core taxa change with taxonomic level? Why?
3. Which taxa are the most abundant taxa in the core taxa?

Step 3: Unique taxa

Click on the Unique Taxa tab. Unique taxa are those taxa found only in a single treatment. The tables generated in this tab show the unique taxa found in particular treatments. The caption at the top of the table tells you how many unique taxa there are relative to the total number of taxa. You can select the taxonomic level of interest using the dropdown menu on the left. If there are fewer tables than treatments, this is because some treatments might not contain unique taxa.

Questions
1. How many taxa are unique to each treatment at each taxonomic level?
2. Are there particular taxonomic groups that tend to be found in one treatment and not the others?

Step 4: Rarefaction

Click on the Rarefaction tab. Taxa accumulation curves will be displayed, showing the number of different taxa as a function of the number of sequences evaluated. In the context of microbiome community analysis, rarefaction relates to sampling depth or the number of sequences in a particular sample. Differences in the number of sequences might not be due to differences in the abundance of bacteria in a sample, but due to differences in DNA extraction and sequencing. Because the number of taxa increases with the number of sequences, up to a point, we might want to standardize the data to a constant number of sequences. Typically, the data are standardized to the minimum number of sequences across all samples, as represented in the bottom graph. Rarefaction is done by random sampling of the data. Some scientists argue that rarefaction is a bad idea when comparing samples. Since only a sub-set of the data collected are analyzed, you may be losing important information. Other scientists argue that certain statistical methods that are used to help us compare differences between samples produce biased results when the samples being compared contain different number of total sequences. Given that there are arguments for and against rarefaction, in all subsequent tabs, you can select whether to use the raw data or the rarified data.

Using the dropdown menus on the left, you can select the taxonomic level of interest and whether to plot the data by sample or treatment.

Questions

1. How does treatment affect microbiome richness in the bean beetle? For example, which treatment resulted in higher richness in your dataset? Which resulted in lowest richness?
2. Does the number of total sequences per sample differ between treatments? For example, does one treatment tend to have higher sequence counts than another? If so, how might this affect whether you choose to analyze raw or rarefied data in subsequent analyses?

Step 5: Taxonomy Bar Graphs

Click on the Taxonomy Bar Graphs tab. The stacked bar graphs show the abundance of different taxa for each sample separated by treatment. Using the dropdown menus on the left, you can select the taxonomic level of interest, whether to use raw or rarified data, and whether to plot the absolute or relative abundances.

Questions

1. Compare bar graphs at different taxonomic levels (Phylum, Class, Order, Family). Are replicate samples similar in terms of the
    a. taxa present?
    b. their relative abundances in each replicate sample?
2. What is the most dominant taxon in each treatment type?
3. Do your conclusions change depending on whether you consider absolute or relative abundance?
4. Do your conclusions change depending on whether you consider raw or rarified data?

Step 6: Taxonomy Heatmaps

Click on the Taxonomy Heatmaps tab. Another way to visualize and compare taxon abundances across samples is with a heat map. Abundance is observed as a color gradient with lighter colors representing greater abundance. The taxa are ordered with the most abundant taxa at the top of the heatmap. When there is a large number of taxa, you might need to scroll down to see all of the taxa. Using the dropdown menus on the left, you can select the taxonomic level of interest and whether to use raw or rarified data.

Questions

1. Compare heatmaps at different taxonomic levels (Phylum, Class, Order, Family). Are replicate samples similar in terms of the
    a. taxa present?
    b. their relative abundances in each replicate sample?
2. What is the most dominant taxon in each treatment type?
3. Do your conclusions change depending on whether you consider raw or rarified data?

Step 7: Alpha Diversity

Click on the Alpha Diversity tab. Alpha diversity is a measure of diversity within a particular sample. A diversity index is a mathematical measure of the diversity in a given community. Richness is a count of the number of taxa present in a given community – for example, the number of different families in each microbiome sample. While counting the number of different taxa is a first step in characterizing the diversity of a community, it is also important to examine the relative abundances, or the proportion of individuals (or sequence reads, in this case), belonging to different taxa in a given community. This tab allows you to consider the Simpson's (specifically the inverse of the Simpson Index, 1-D) and the Shannon's diversity indices. Both indices incorporate richness and evenness (how evenly taxa are represented in a particular community). The Simpson index is more sensitive to the influence of the dominant, or highly abundant taxa, as compared to the Shannon index. Box plots are used to summarize the data for each treatment and either a t-test or an ANOVA test for significant differences in diversity

between treatments. In each graph, the median value for each treatment (the midpoint of the data) is indicated by the dark horizontal line in the box plot. Using the dropdown menus on the left, you can select the taxonomic level of interest, whether to use raw or rarified data, and what diversity measure to use.

You may calculate the raw data median values (shown in the box plots) for each of the alpha diversity indices by using the Excel spreadsheet of the level-5 data that you saved in excel format (when you were preparing a clean CSV file for the RShiny app. Open the Excel file, select the entire datafile, open a new worksheet, click in cell A1 and Paste Transposed. This will return the clean dataset to a form in which the rows are the individual microbiome samples and the columns are the taxa (Families). Now you may perform the same alpha diversity calculations you performed when you did the Colony Phenotype Analysis and the Colony Sequence Analysis. Calculate the bacterial cell (sequence) abundance for each microbiome. Then calculate the Family richness for each microbiome and then the median richness for each treatment in your study. Calculate Simpson's Index and the Inverse Simpson's Index for each microbiome, then calculate the median value of the Inverse Simpson's Index for each treatment. Finally, calculate Shannon's index for each microbiome and the median values for each treatment.

Questions
1. Do the treatments differ significantly in terms of diversity?
2. Do your conclusions change depending on taxonomic level you consider?
3. Do your conclusions change depending on whether you consider raw or rarified data?
4. Do your conclusions change depending on what measure of diversity you consider?

Step 8: Beta Diversity
Click on the Beta Diversity tab. Beta diversity is a measure of diversity between samples. Measures of beta diversity include the identity of taxa in each community and in some cases the abundance of those taxa. Samples might have the same alpha diversity (e.g., richness) but be very different communities if they have different taxa. Measures of beta diversity capture this difference. This tab includes three different distance measures (i.e., measures of beta diversity or how different communities are) – Jaccard, Bray-Curtis, and Morista-Horn. Jaccard just considers the presence and absence of taxa and ignores the relative abundance of those taxa. Both Bray-Curtis and Morista-Horn incorporate the identity of taxa and the relative abundances of the taxa. Bray-Curtis is the most commonly used measure of beta diversity in research on insect microbiomes. Morista-Horn is preferred in ecological studies in which communities might be incompletely sampled. Incomplete sampling would be indicated by a taxa accumulation curve that did not become horizontal in either the raw data or the rarefied data. This tab also allows you to visualize the data in two ways – Non-metric Multidimensional Scaling (NMDS) and Principal Coordinates Analysis (PCoA). PCoA has traditionally been used in studies of microbial communities. However, NMDS is gaining in popularity. In these visualizations, communities that are more similar are closer together on the graph. In the PCoA plots, the numbers in brackets on the axes tell how much of the variation is explained by that axis. In the NMDS plots, a "stress" value is given below the plot. This is a measure of how well the data are explained in a reduced number of dimensions. Lower values of stress are better. A stress value below 0.1 suggests that the NMDS does a very good job of explaining the data in two dimensions. If you graph by treatment group, confidence ellipsoids will be shown for each treatment. Below the graphs, the results of PERMANOVA are displayed. A PERMANOVA is similar to an ANOVA but uses community data as the response variable.

Questions
1. Do the microbial communities differ significantly between the treatments?
2. How variable are the communities within a particular treatment?
3. Do your conclusions change depending on taxonomic level you consider?
4. Do your conclusions change depending on whether you consider raw or rarified data?
5. Do your conclusions change depending on the distance measure you use?
6. Do your conclusions change depending on the ordination method you use?

Synthesis questions

1. Why might particular bacteria be core taxa that are common to all treatments?
2. Why might particular bacteria be unique to certain treatments?
3. In general, considering the results of your various analyses, what might explain the differences between your treatments, if they exist?
4. Do your conclusions depend on the parameters that you chose during the analysis the data? For example, are your conclusions different if you used the raw data or the rarefied data? Are your conclusions different for different distance measures when analyzing beta diversity? If the parameters chosen do affect your conclusions, does this change your overall interpretation of the results?

Using Images from the BeanBeetleMicrobiome app

The graphs produced in the BeanBeetleMicrobiome app may be inserted in slides (for an oral presentation) or a text document (for a written report). On a MacOS computer, you may click and drag an image from the BeanBeetleMicrobiome app to the window of the application in which you wish to insert that image. Alternatively, a selective screen shot will work. On a Windows computer, right click on an image to select it and copy, then paste it in a slide or text document.

Cited References

Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, Behrenfeld MJ, Boetius A, Boyd PW, Classen AT, Crowther TW. 2019. Scientists' warning to humanity: microorganisms and climate change. Nat Rev Microbiol 17:569–586.

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.

Dave M, Higgins PD, Middha S, Rioux KP. 2012. The human gut microbiome: current knowledge, challenges, and future directions. Transl Res 160:246-257.

Douglas AE. 2011. Lessons from studying insect symbioses. Cell Host Microbe 10:359-367.

Martiny JB, Whiteson KL, Bohannan BJ, David LA, Hynson NA, McFall-Ngai M, Rawls JF, Schmidt TM, Abdo Z, Blaser MJ, Bordenstein S. 2020. The emergence of microbiome centres. Nature Microbiology 5:2-3.

Timmis K, Cavicchioli R, Garcia JL, Nogales B, Chavarría M, Stein L, McGenity TJ, Webster N, Singh BK, Handelsman J, de Lorenzo V. 2019. The urgent need for microbiology literacy in society. Environ Microbiol 21:1513-1528.

Materials

The BeanBeetleMicrobiome app can be run online in the cloud at https://beanbeetles.shinyapps.io/BeanBeetleMicrobiome/. Alternatively, the app can be downloaded and run locally. Running the app locally requires R and RStudio. Details on downloading and running the app locally are available in the Supplementary Materials.

Notes for the Instructor

We have used the BeanBeetleMicrobiome app in both introductory biology and upper-level laboratory courses. The student handout above is aimed more at upper-level students and encourages them to explore microbiome community data in detail by considering different choices that can be made in the analysis of these data. In introductory courses, assignments should be more scaffolded with students given more direction as to what options to select. For example, instructors might have students just consider the data at the family level, only use rarefied data, and only consider visualizations of taxonomic diversity and analysis of alpha diversity.

The app can be used after bioinformatic analysis of next-generation sequencing data from microbiome samples using the Purple Line of the DNA Subway platform (dnasubway.cyverse.org), as is done in introductory biology laboratory courses at

Morehouse College where an entire semester is dedicated to the Bean Beetle Microbiome Project (Zelaya et al. 2020). Alternatively, instructors could just focus on community analysis by providing students with a level-5 taxonomy file for analysis. In the second-semester introductory biology lab at Emory University, students use culture-based methods to identify bacteria from bean beetles (Cole et al. 2018) and then analyze level-5 taxonomy files (Cole et al. 2021). We have included several level-5 taxonomy files from bean beetle microbiome studies in the Supplementary Materials for this purpose. In addition, the app may be used to conduct community analysis on data from other published microbiome studies. As an example, we have included data from a study on the microbiome difference in coffee berry borers (*Hypothenemus hampei*) (Ceja-Navarro et al. 2015) in the Supplementary Materials.

The amount of time required to complete the exercise depends on the duration of introduction prior to students beginning the exercise and the degree of sophistication of analysis required. We have typically introduced community analysis of microbiome data to students (see Supplementary Materials for an introductory presentation) and then have them begin their analysis in class. They complete their analysis for homework.

Cited References

Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, Behrenfeld MJ, Boetius A, Boyd PW, Classen AT, Crowther TW. 2019. Scientists' warning to humanity: microorganisms and climate change. Nat Rev Microbiol 17:569–586.

Ceja-Navarro JA, Vega FE, Karaoz U, Hao Z, Jenkins S, Lim HC, Kosina P, Infante F, Northen TR, Brodie EL. 2015. Gut microbiota mediate caffeine detoxification in the primary insect pest of coffee. Nature Communications 6:1-9.

Cole MF, Acevedo-Gonzalez T, Gerardo NM, Harris EV, Beck CW. 2018. Effect of Diet on Bean Beetle Microbial Communities. Article 3 In: McMahon K, editor. Tested studies for laboratory teaching. Volume 39. Proceedings of the 39th Conference of the Association for Biology Laboratory Education (ABLE).

Cole MF, Hickman MA, Morran L, Beck CW. 2021. Assessment of course-based research modules based on faculty research in introductory biology. J Microbiol Biol Educ 22:e00148-21.

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.

Dave M, Higgins PD, Middha S, Rioux KP. 2012. The human gut microbiome: current knowledge, challenges, and future directions. Transl Res 160:246-257.

Douglas AE. 2011. Lessons from studying insect symbioses. Cell Host Microbe 10:359-367.

Kandlikar GS, Gold ZJ, Cowen MC, Meyer RS, Freise AC, Kraft NJ, 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.

Martiny JB, Whiteson KL, Bohannan BJ, David LA, Hynson NA, McFall-Ngai M, Rawls JF, Schmidt TM, Abdo Z, Blaser MJ, Bordenstein S. 2020. The emergence of microbiome centres. Nature Microbiology 5:2-3.

Timmis K, Cavicchioli R, Garcia JL, Nogales B, Chavarría M, Stein L, McGenity TJ, Webster N, Singh BK, Handelsman J, de Lorenzo V. 2019. The urgent need for microbiology literacy in society. Environ Microbiol 21:1513-1528.

Zelaya AJ, Gerardo NM, Blumer LS, Beck CW. 2020. The bean beetle microbiome project: a course-based undergraduate research experience in microbiology. Frontiers in Microbiology. 11:2274.

Acknowledgments

About the Authors

Carolyne Huang is a senior biology major at Emory University, where she is continuing her work on the Bean Beetle Microbiome Project with her mentor, Dr. Anna Zelaya. Dr. Anna Zelaya is a FIRST (Fellowships in Research and Teaching) post-doctoral fellow at Emory University, where she works as part of the Bean Beetle Microbiome Project. Dr. Larry Blumer is a Professor in the Department of Biology at Morehouse College, where he teaches the Bean Beetle Microbiome CUREs for introductory biology. Dr. Nicole Gerardo is a Professor in the Department of Biology at Emory University, where she teaches CUREs on fungus-growing ant mutualisms. Her lab also researches insect-microbe interactions. Dr. Christopher Beck is a Professor of Pedagogy in the Department of Biology at Emory University, where he recently co-taught an upper-level lab course based on the Bean Beetle Microbiome CURE with Dr. Anna Zelaya.

Mission, Review Process & Disclaimer

The Association for Biology Laboratory Education (ABLE) was founded in 1979 to promote information exchange among university and college educators actively concerned with teaching biology in a laboratory setting. The focus of ABLE is to improve the undergraduate biology laboratory experience by promoting the development and dissemination of interesting, innovative, and reliable laboratory exercises. For more information about ABLE, please visit http://www.ableweb.org/.

Advances in Biology Laboratory Education is the peer-reviewed publication of the conference of the Association for Biology Laboratory Education. Published articles and extended abstracts are evaluated and selected by a committee prior to presentation at the conference, peer-reviewed by participants at the conference, and edited by members of the ABLE Editorial Board. Published abstracts are evaluated and selected by a committee prior to presentation at the conference.

Citing This Article