

RShiny Bean Beetle Microbiome Tutorial

BeanBeetleMicrobiome is an RShiny app designed for community analysis of level-5 (family-level) datasets they have produced in DNA Subway. Although this app was designed for community analysis of data from the Bean Beetle Microbiome Project, it could be used for community analysis of any level-5 data. Before running the app, remove any unidentified taxa, chloroplasts, and mitochondria data from the level-5 spreadsheet. The level-5 file must be formatted with the first column as the taxa and the subsequent columns as samples with unique sample identifiers. Also, prepare a metadata file with the first column as sample names and the second column as treatments. Both files must be in .csv format. The order of the samples (i.e., rows in the spreadsheets) must be the same for the level-5 file and the metadata file.

Preparing files for analysis

Creating the metadata file

This metadata file is similar to the one created for DNA Subway, but 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 "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 your taxonomy file generated in 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) taxonomic names as the

headers of all the other columns. The numbers in each cell represent the “counts” for each bacterial type 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; k_Bacteria												
2	ADZU.3	2651	0	0	0	0	0	0	0	0	12	0	0	0
3	ADZU.4	2439	0	0	0	0	0	0	0	0	5	0	0	0
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	0	5	61	7	0	0

- In Excel, change the header title “Index” in cell A1 and replace it with the name “sampleID”
- 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
1	ia; k_Bacteria; k_Bacteria; k_Bacteria; k_Bacteria; k_Bacteria; k_Bacteria; k_Bacteria; BarcodeSeq							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		


- You will see that there are extra columns that contain information from your metadata file. We want to delete these columns because they do not count sequence count data. So, delete these columns entirely by highlighting all columns from the metadata file. Then right click to delete the 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; k_Bacteria; Verrucomicrobia; 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													

- Save your file so as to not lose what you have done up to this point.
- Create a new Excel worksheet by clicking on File, then New.
- A new spreadsheet should appear right in front of the spreadsheet you had just been working on.
- 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.
- Then, select all of the data and copy it.
- Next, click on the new spreadsheet you created (should be located behind your original taxonomy spreadsheet) to bring it back to the front.
- Click on the first cell A1 of your spreadsheet so that the first cell in outlined in green.
- 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.taxonon	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

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



	A	B	C	D	E
1	sum.taxonon	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

	A	B	C	D	E
1	sum.taxonomy	ADZU.3	ADZU.4	BEP.3	BEP.4
2	k_Bacteria; ; ; ; ;	2651	2439	11452	4168
3	k_Bacteria;p_Acidobacteria;c_Acidobacteria;o_Acidobacteriales;f_Koribacteraceae	0	0	5	0
4	k_Bacteria;p_Acidobacteria;c_DAO52;o_Ellin6513;f_	0	0	0	9
5	k_Bacteria;p_Acidobacteria;c_Solibacteres;o_Solibacterales;f_	0	0	0	3
6	k_Bacteria;p_Acidobacteria;c_[Chloracidobacteria];o_RB41;f_Ellin6075	0	0	9	0
7	k_Bacteria;p_Actinobacteria;c_Acidimicrobia;o_Acidimicrobiales;f_	0	0	6	0
8	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_	0	0	14	0
9	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Actinomycetaceae	0	0	0	5
10	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Actinopolysporaceae	0	0	0	6
11	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Corynebacteriaceae	12	5	83	61
12	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Dermabacteraceae	0	0	16	7
13	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Dermacoccaceae	0	0	14	0
14	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Intrasporangiaceae	0	0	30	0
15	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Microbacteriaceae	0	0	9	0
16	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Micrococcaceae	79	86	38	40
17	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Nocardiaceae	0	0	0	0
18	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Nocardioidaceae	7	0	0	0
19	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Promicromonosporaceae	0	0	0	0
20	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Pseudonocardaceae	0	0	14	9
21	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Sporichthyaceae	0	0	0	0

Before

After expanding column A

- 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).
- Now, read through your taxonomy names. There are few rows that must be deleted from our dataset.
- 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 taxa 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 really bacterial symbionts of plants, they get sequenced in Mi-Seq. Adult bean beetles retain some bean material in their gut, which is where the chloroplast sequence is from. Since they are not really part of the beetle microbiome, we want to delete these rows from our file.
 - Mitochondria: Similar to chloroplasts, mitochondria are bacterial symbionts of plants and are sequenced in Mi-Seq.
 - Make sure to delete every instance of unassigned, archaea, chloroplast, and mitochondria
18. Save your file by clicking the Save button and 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 .xlsx 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 perform 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 file and metadata file that you 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, which is what is 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. Because only a sub-set of the data collected is 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. Because there are arguments for and against rarefactions, 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 rarefied 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 taxa 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 rarefied 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 are 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 rarefied 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 taxa in each treatment type?
3. Do your conclusions change depending on whether you consider raw or rarefied 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 abundances, or the number of individuals (or sequence reads, in this case), belonging to the different taxa. This tab allows you to consider the Simpson's (specifically the complement 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-tests or an ANOVA tests for significant differences in diversity between treatments. In each graph, the median value for each treatment (the mid-point 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 taxa 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 abundance of those taxa. Both

Bray-Curtis and Morista-Horn incorporate the identity of taxa and the 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 ANOVA but with 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 rarefied data?
5. Do your conclusions change depending on what distance measure you use?
6. Do your conclusions change depending on what 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?

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.