Microbial Community Analysis Using Colony Phenotypes

Student Handout

Objectives

- Manipulate student collected datasets to conduct community-level ecological analyses
- Use community-level data to address questions about insect microbiomes
- Use Google Sheets to calculate community ecology variables
- Compare microbial community using community ecology variables

Introduction

Microbiomes are the communities of microbes (bacteria, viruses, fungi and archaea) living symbiotically with all metazoans. In the past decade, both interest and research on microbiomes, including their implications for human health, have increased dramatically (Christian *et al.* 2015, Costello *et al.* 2012, McFall-Ngai *et al*. 2013, The Human Microbiome Consortium 2012, Young 2016). Insects have been used as model species to study the importance of microbiomes, because of their ease of use and the fact that microbial communities play diverse roles in insects (Engel and Moran 2013).

The data that are collected in any microbiome study consists of lists of the taxonomic units identified and their abundance. The same types of data are evaluated in an ecological community analysis, but now the communities are the collections of microbes that constitute different microbiomes. The community variables, "species" richness and relative abundance, are the same and the statistical methods used to compare communities, diversity and difference indices, also are the same. Perhaps the simplest measure of community structure used by ecologists is "species" or taxon richness, a count of the number of unique taxa in a sample. However, species richness does not consider the relative abundance of species in a community. Imagine two communities with five different species. In one community, all of the species have the same relative abundance. In the other community, one species dominates comprising 95% of individuals in the community. The other four species are very rare. Based on species richness as a measure of community structure, these two communities are the same, although they are clearly very different. As a result, ecologists use other species diversity indices that consider both the number of species and the relative abundance of species in a community. Two common indices are the Simpson Index and the Shannon-Weaver Index. Communities with greater numbers of species and higher evenness (i.e., similar relative abundance of species within a community) are considered more diverse.

A colony phenotype analysis does not permit us to identify the bacteria but only determine whether there are different taxa on a given culture dish. Measures of taxon richness and taxon diversity do not consider the identity of taxa in a community. Communities could have the same level of taxon diversity, but have completely different taxa. When it is possible to identify taxa, then a measure of community similarity, such as the Bray-Curtis Index, may be used to compare the similarity (or dissimilarity) between two communities based on the identity of taxa in the communities, as well as their relative abundances. In this initial analysis using colony phenotypes, community similarity measures

cannot be accurately calculated. For more information on indices of species diversity and measures of community similarity, see Krebs (1999).

In this example, you will have already collected bean beetle gut microbiome data by tabulating colony phenotypes from cultured bacteria. Now we are ready to analyze the colony phenotype data. In the example shown in the following pages, bean beetles were raised on either adzuki beans (*Vigna angularis*) or blackeye peas (BEP, *Vigna unguiculata*), but the analysis could be performed on any dataset in which two treatments or groups are compared such as beetle age, beetle sex, or a manipulation treatment.

Microbial Community Analysis Using Colony Phenotype

Questions

Using data students collect on colony phenotypes and the analyses described below, answer the following questions.

1. Based on the diversity indices that you calculated, which treatment had the highest (lowest) diversity?
2. Do the community ecology variables you calculated all lead to the same conclusion?
3. Is there a relationship between number of samples and taxonomic diversity? If so, what might explain this?

Dataset Creation

Your class or course will create a dataset containing information on the microbial community of bean beetles based on colony phenotypes of individual bacterial colonies cultured from bean beetle homogenates plated on different media. Different bacteria cultured from each beetle having different combination of phenotypic characters should be tabulated. A unique combination of phenotypic characters is taken to represent a unique bacterial taxon on a given culture plate.

The dataset you create should be a spreadsheet (Google Sheets) containing information on the bean host species, sex, life cycle stage, media on which bacteria were grown, and colony phenotype. Eosin methylene blue (EMB) and phenylethyl alcohol agar (PEA) plates select for gram negative and gram positive bacteria, respectively. So, a colony with a particular combination of phenotypic characters on an EMB plate is a different bacterial taxon than a colony with the same combination of characters on a PEA plate. Thus, the bacterial colonies tabulated from both EMB and PEA plates from a given beetle would represent the entire microbiome community sample from one beetle. The same is not true if we include nutrient agar plates in our sample. Bacterial colonies tabulated from nutrient agar plates should be evaluated independently of those tabulated from EMB and PEA plates since these would be redundant tabulations.

Collect your data from nutrient agar media (NA) and EMB and PEA media in separate spreadsheets (separate tabs for sheets in the same workbook) in a Google Sheets spreadsheet similar to that shown below:

Data manipulation

1. Choose one of your spreadsheets (let's work with the NA sheet), but the process is the same with either the dataset for nutrient agar or the dataset for EMB and PEA. Start by selecting all the information in one sheet, copy the entire dataset and then paste special as values only in a new sheet. The "+" at the lower left will create a new sheet tab. Double click the tab label to give it the name "NA Values". Then in the new sheet, delete Column L, the column tabulating the observed numbers of colony forming units (CFUs) for the full strength and 1/10 dilution cultures. All our analysis will be conducted on full strength culture or equivalent tabulations.

2. With our new "phenotype" dataset, we need to define a bacterial "Taxon" based on the combination of media and the four phenotypic characters (color, gloss, form and elevation). One way to do this is to create a "taxon" name by concatenating the media and the four different phenotypic traits. You can do this using the CONCATENATE function in Google Sheets to create a new variable in Column M "=CONCATENATE(G2,H2,I2,J2,K2)". After you create the "taxa" names, you might want to select the column and then re-paste it in the same column by pasting values (using paste special) to remove the formula.

The spreadsheet screenshot shows:

	Observer Name	Treatment	Source	Host	sex	stage	media	color	gloss	form	elevation	Full CFU#	Taxon
2	RR	M	M37 1/10	BEP	M	adult	NA	yellow	shiny	circular	flat	3170	NAyellowshinycircularflat
3	JJ	M	M35 Full	BEP	M	adult	NA	white	shiny	circular	flat	318	
4	PJ	M	M34 1/10	BEP	M	adult	NA	white	shiny	circular	raised	4000	
5	DG	M	M38 1/10	BEP	M	adult	NA	white	shiny	circular	raised	11020	
6	TX	M	M17 1/10	BEP	M	adult	NA	off-white	shiny	circular	flat	3320	
7	TX	M	M37 1/10	BEP	M	adult	NA	off-white	shiny	circular	flat	3010	
8	MJ	M	M36 1/10	BEP	M	adult	NA	white	shiny	rregular	raised	6010	
9	CJ	F	F38 1/10	BEP	F	adult	NA	white	shiny	irregular	flat	1600	
10	CJ	F	F38 1/10	BEP	F	adult	NA	white	shiny	circular	raised	1600	
11	MN	F	F37 1/10	BEP	F	adult	NA	off-white	shiny	circular	raised	14040	
12	LR	F	F34 1/10	BEP	F	adult	NA	white	shiny	irregular	flat	5120	
13	DS	F	F35 1/10	BEP	F	adult	NA	off-white	shiny	Irregular	raised	190	
14	ZC	F	F36 1/10	BEP	F	adult	NA	white	shiny	circular	flat	6010	

Formula bar: =CONCATENATE(G2:K2)

Callout box: Insert a new column for "taxon" then in the first blank sell of the column insert the formula =CONCATENATE(G2,H2,I2,J2,K2). Then, copy the formula down for the entire column.

3. Now, we need to consolidate the data for each treatment and source, by the bacterial taxa. The easiest way to do this is with the Pivot Table function in Google Sheets.

4. When clicked on a cell within the data, create a Pivot Table (Insert -> Pivot Table). Make sure that the data source includes the top row, which has the column headings. Select create a New sheet in the Create pivot table pop-up window.

   Set the Treatment and Source as the rows as these represent the treatment and individual communities, respectively. Click the Add button to add these rows. Check Repeat row labels for "Treatment" but uncheck Show totals for each. The new "taxon" column should be the columns in the pivot table and check Show totals. The Values should be the "Full CFU#" which will be the SUM of the CFU (colony-forming units, a measure of density), which will be shown as "Full CFU# SUM" using the Pivot table editor.

5. Create a new sheet and copy and paste the pivot table in the new sheet using Paste Special -> Paste values only. You may add zeros to all of the empty cells but it is not necessary. The empty cells indicate counts of zero. Delete all three rows of totals (Adzuki Total, BEP Total and Grand Total rows). Rename the Grand Total column "Abundance" and rename each taxon, "Taxon1", Taxon2 …. Delete any extra rows at the top of the spreadsheet, so cell A1 is "Treatment". Rename this sheet NA Community Data.

Rename columns "Taxon 1", "Taxon 2", and so on.

Rename tab "NA Community Data"

Calculating diversity indices

1. Species (taxon) richness – the number of unique species (taxa) in a sample
   a. Although you could manually count the number of cells with values greater than zero for each treatment, using the COUNTIF formula in Google Sheets is easier (e.g., =COUNTIF(range,">0") where "range" the is range of cells in the spreadsheet containing the data, such as "C2:K2").

   b. After copying the richness formula to all the rows, calculate the average richness for each treatment category (in this example it is the average taxon richness among the four beetles from each of the two different host bean species, adzuki and BEP).

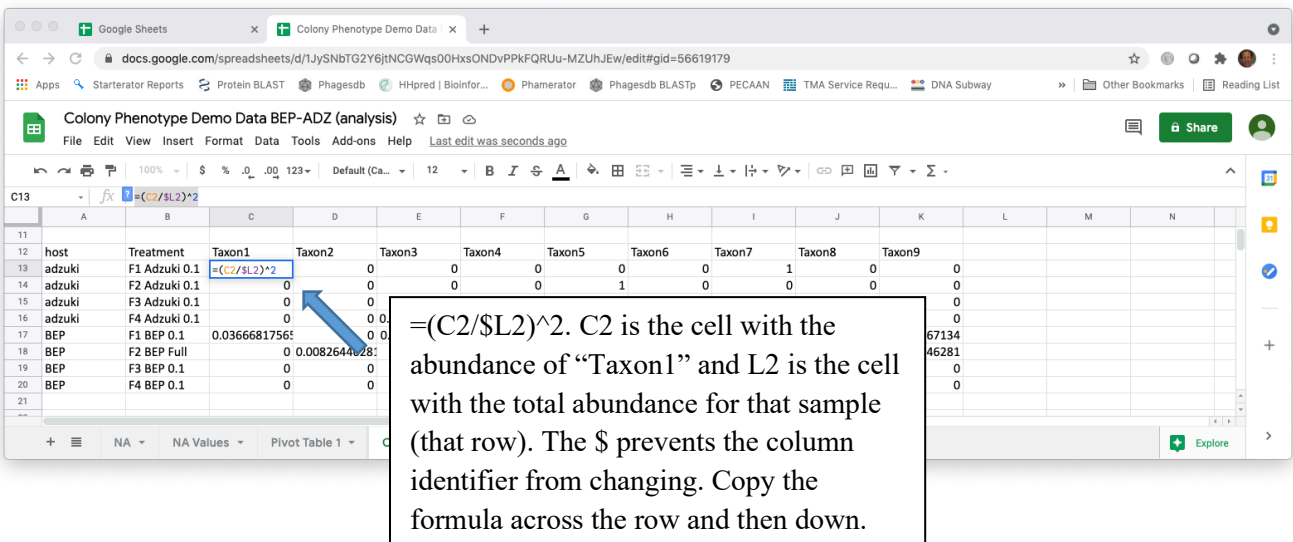2. Simpson Index – the Simpson Index incorporates both species (taxon) richness and species (taxon) evenness.
   a. $D = \Sigma(n/N)^2$, where n=number of individuals of a particular species (taxon) and N=total number of individuals in a sample. D increases as diversity decreases, which is

counterintuitive. A reciprocal or inverse index would be more intuitive and are easily calculated.

b. Reciprocal Simpson = 1/D and scales so the maximum value is the species richness of a community.

c. Inverse Simpson = 1-D and scales to a maximum value of 1.0.

d. Create a new data array below the original using the same row labels (treatment variables) and the same column labels (species).
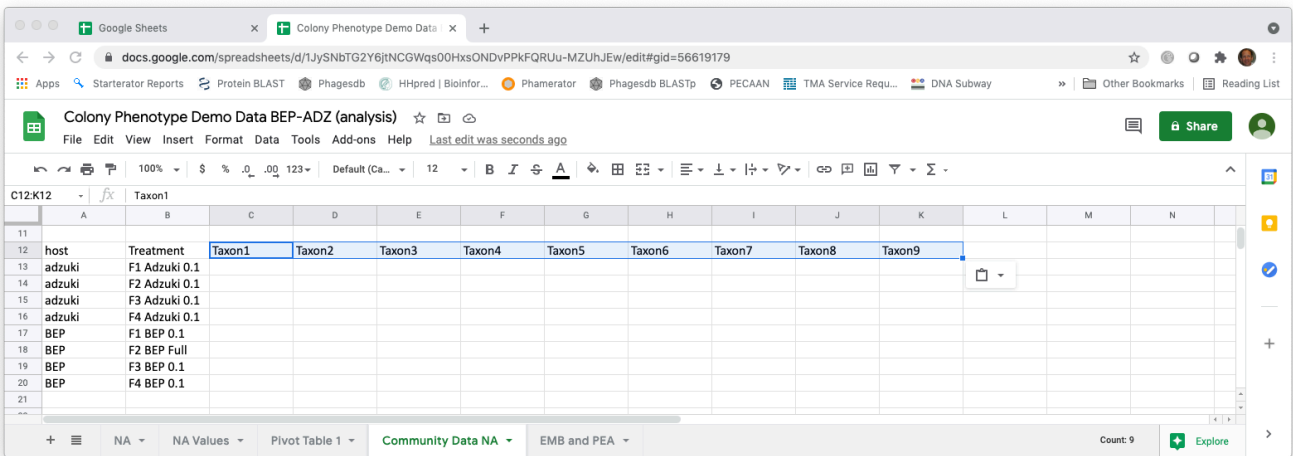
e. To calculate the proportion squared for each taxon, use the grand totals for each treatment. Using the Google Sheets (and Excel) trick that $ before a column or row prevents Google Sheets from iterating when copying a formula makes this easy. For example, =(C2/$L2)^2. Copy the formula across the row and then down.

=(C2/$L2)^2. C2 is the cell with the abundance of "Taxon1" and L2 is the cell with the total abundance for that sample (that row). The $ prevents the column identifier from changing. Copy the formula across the row and then down.

f. Calculate the sum of the proportions squared (=SUM in Google Sheets for each row, a different microbial community) to calculate the Simpson Index.
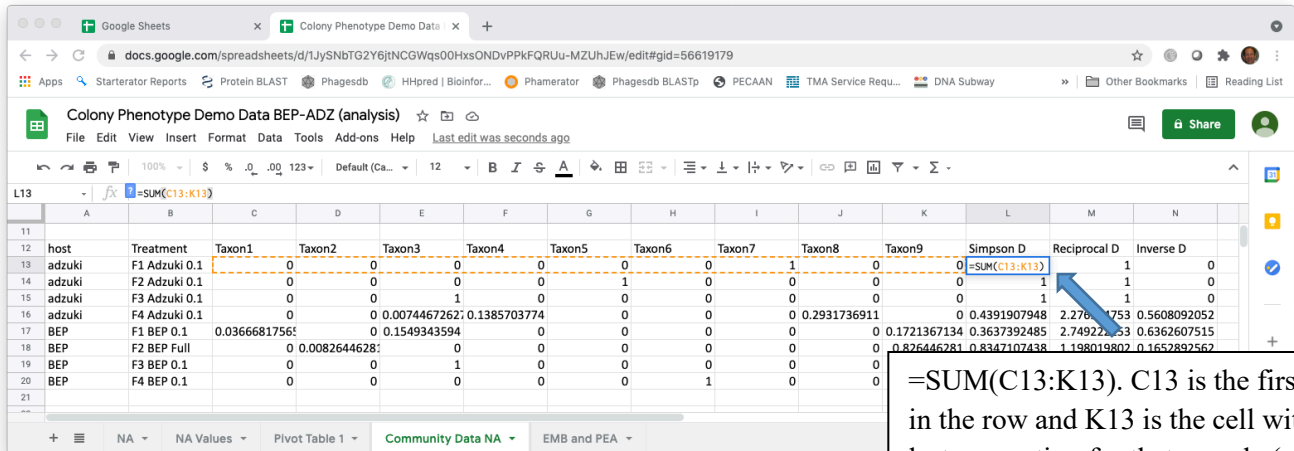
=SUM(C13:K13). C13 is the first cell in the row and K13 is the cell with the last proportion for that sample (row).

g. Calculate the reciprocal (e.g., =1/L13) and inverse Simpson (e.g., =1-L13) using formulas in Google Sheets.

h. Now, calculate the average of either the reciprocal or the inverse Simpson for each of the treatment groups (comparing the two host bean species).

3. Shannon-Weaver (Shannon-Weiner) Index – also incorporates species (taxon) richness and species (taxon) evenness

   a. H=-$\Sigma p \ln p$, where p is the proportion of individuals of each species (taxon) in a community (i.e., n/N).

   b. Create a new data array below the original using the same row labels (treatment variables) and the same column labels (species).

   c. Using the grand totals for each community, calculate the proportions ($p \ln p$). Using the Google Sheets (and Excel) trick that $ before a column or row prevents Google Sheets from iterating when copying a formula makes this easy.

   d. Note that ln*p is* undefined if *p*=0, so you can use an "IF" statement in Google Sheets. For example, =IF(C2>0,(C2/$L2)*LN(C2/$L2),"")

The image shows a Google Sheets screenshot. The formula bar shows cell C24 with `=IF(C2>0,(C2/$L2)*LN((C2/$L2)),"")`. The spreadsheet has columns: host, Treatment, Taxon1 through Taxon9, Shannon-Weaver.

host	Treatment	Taxon1	Taxon2	Taxon3	Taxon4	Taxon5	Taxon6	Taxon7	Taxon8	Taxon9	Shannon-Weaver
adzuki	F1 Adzuki 0.1	=IF(C2>0,(C2/$L2)*LN((C2/$L2)),"")						0			
adzuki	F2 Adzuki 0.1				0						
adzuki	F3 Adzuki 0.1										
adzuki	F4 Adzuki 0.1			-0.21142049	.367						
BEP	F1 BEP 0.1	-0.3165171749		-0.3669994061							
BEP	F2 BEP Full		-0.2179904793								
BEP	F3 BEP 0.1			0							
BEP	F4 BEP 0.1										

=IF(C2>0,(C2/$L2)*LN(C2/$L2),""). C2 is the cell with the abundance of "Taxon1" and L2 is the cell with the total abundance for a sample. The $ prevents the column identifier from changing. Copy the formula across the row and then down.

e. Calculate the negative sum of the proportions ($p\ln p$) (=-SUM in Google Sheets for each row, a different microbial community) to calculate the Shannon-Weaver Index.

The image shows a Google Sheets screenshot. The formula bar shows cell L24:L31 with `=-SUM(C24:K24)`.

host	Treatment	Taxon1	Taxon2	Taxon3	Taxon4	Taxon5	Taxon6	Taxon7	Taxon8	Taxon9	Shannon-Weaver
adzuki	F1 Adzuki 0.1							0			=-SUM(C24:K24)
adzuki	F2 Adzuki 0.1				0						0
adzuki	F3 Adzuki 0.1			0							0
adzuki	F4 Adzuki 0.1			-0.21142049	-0.3678535764				-0.3321800455		0.911454112
BEP	F1 BEP 0.1	-0.3165171749		-0.3669994061					.9956629		1.048512244
BEP	F2 BEP Full		-0.2179904793						.086645618		0.3046360973
BEP	F3 BEP 0.1			0							0
BEP	F4 BEP 0.1										0

= -SUM(C24:K24). C24 is the first cell in the row and K24 is the cell with the last proportion for that sample.

f. As you did for Taxon Richness and either the reciprocal or inverse Simpson Index, calculate the average Shannon-Weaver values for each treatment group, each host bean species.

Cited References

Christian N, Whitaker BK, Clay K. 2015. Microbiomes: unifying animal and plant systems through the lens of community ecology theory. Front. Microbiol. 6:1–15.

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).

Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA. 2012. The application of ecological theory toward an understanding of the human microbiome. Science. 336:1255–1262

Engel P, Moran NA. 2013. The gut microbiota of insects – diversity in structure and fuction. FEMS Microbiol. Rev. 37:699-735.

Krebs CJ. 1999. Ecological Methodology, 2nd edition. New York: Benjamin Cummings.

McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll AH, Kremer N, Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A, Ruby EG, Rumpho M, Sanders JG, Tautz D, Wernegreen JJ. 2013. Animals in a bacterial world, a new imperative for the life sciences. Proc. Natl. Acad. Sci. U.S.A. 110:3229–3236.

The Human Microbiome Consortium. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486: 207-214.

Young E. 2016. I Contain Multitudes: The Microbes Within Us and a Grander View of Life. New York: HarperCollins Publishers.