

Colony-based Sanger Sequencing Analysis

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

Your group or class has collected data on the microbial community of bean beetles based on 16S rRNA sequencing of individual bacterial colonies cultured from bean beetle homogenates plated on different media. Since only a small number of colonies were sequenced from each plate, the data do not represent the entire microbial community for a particular sample. However, comparisons may be made based on host bean species, sex, and other sample or experimental variables with the assumption that the collection of taxa from a given treatment are representative of the bacterial microbiome communities of bean beetles in that treatment. There are no independent samples within treatment groups in this kind of analysis, however, we do have an identification for each picked colony in the dataset to the level of genus.

Dataset Creation

1. You and your fellow students will have conducted BLASTn searches on the individual 16S rRNA gene sequences and entered the genus and family of each sample in a spreadsheet. This spreadsheet should be created in or uploaded to Google Sheets to begin this analysis.
2. If your instructor shares a Google Sheets document with the class, be sure to **save your own copy of the document before beginning the data manipulation.**

	A	B	C	D	E	F	G
1	Sample_ID	Beetle stage	Sex	Host Bean	Strain	Family	Genus
2	T1-ADZ	adult	female	Adzuki	LB	Enterobacteriaceae	Enterobacter
3	T2-ADZ	adult	female	Adzuki	LB	Enterobacteriaceae	Enterobacter
4	T3-ADZ	adult	female	Adzuki	LB	Staphylococcaceae	Staphylococcus
5	T4-ADZ	adult	female	Adzuki	LB	Staphylococcaceae	Staphylococcus
6	T5-ADZ	adult	female	Adzuki	LB	Staphylococcaceae	Staphylococcus
7	T6-ADZ	adult	female	Adzuki	LB	Staphylococcaceae	Staphylococcus
8	T7-ADZ	adult	female	Adzuki	LB	Enterobacteriaceae	Enterobacter
9	T9-ADZ	adult	female	Adzuki	LB	Enterobacteriaceae	Enterobacter
10	T10-ADZ	adult	female	Adzuki	LB	Staphylococcaceae	Staphylococcus
11	T11-ADZ	adult	female	Adzuki	LB	Staphylococcaceae	Staphylococcus
12	T12-BEP	adult	female	Blackeye pea	LB	Staphylococcaceae	Staphylococcus
13	T13-BEP	adult	female	Blackeye pea	LB	Staphylococcaceae	Staphylococcus
14	T14-BEP	adult	female	Blackeye pea	LB	Staphylococcaceae	Staphylococcus
15	T16-BEP	adult	female	Blackeye pea	LB	Enterococcaceae	Enterococcus
16	T17-BEP	adult	female	Blackeye pea	LB	Staphylococcaceae	Staphylococcus
17	T18-BEP	adult	female	Blackeye pea	LB	Staphylococcaceae	Staphylococcus
18	T19-BEP	adult	female	Blackeye pea	LB	Staphylococcaceae	Staphylococcus
19	W1-BEP	adult	female	Blackeye pea	LB	Staphylococcaceae	Staphylococcus
20	W2-BEP	adult	female	Blackeye pea	LB	Staphylococcaceae	Staphylococcus
21	W3-BEP	adult	female	Blackeye pea	LB	Staphylococcaceae	Staphylococcus

Data manipulation

1. We need to consolidate the data by the variable of interest, for example, for each host type, each sex, or two manipulation treatments by the bacterial taxa. The easiest way to do this is with the Pivot Table function in Google Sheets.
2. When clicked on a cell within the data, create a Pivot Table (Insert -> Pivot Table) in a new sheet. Make sure that the data source includes the top row, which has the column headings. Set the treatment(s) that you are interested in as the rows and the taxonomic level you are interested in as the columns. The Values should be the Sample_ID, which is COUNTA of the Sample_ID as each row in the dataset represents a single sample.
3. Uncheck the box Show totals for Rows. Then copy the entire table and paste Special -> Values only in a new sheet.
4. Label this new sheet Community Data. You can add zeros to all of the empty cells in the Pivot Table but it is not necessary (You want to keep the Grand totals column and rename it Abundance).
5. Delete any extra rows at the top of the spreadsheet so cell A1 is Host Bean. The top row should be the names of the taxa.

Demo Sanger Data BEP-ADZ

File Edit View Insert Format Data Tools Extensions Help

100% | \$ % .0_ .00 123

A1 COUNTA of Sample_ID

	A	B	C	D	E
1	COUNTA of Sample_ID	Family			
2	Host Bean	Enterobacteriaceae	Enterococcaceae	Staphylococcaceae	Grand Total
3	Adzuki	4		6	10
4	Blackeye pea		1	9	10

Pivot table editor

Suggested

Rows Add

Host Bean ×

Order: Ascen... Sort by: Host B...

Show totals

Columns Add

Family ×

Order: Ascen... Sort by: Family

Show totals

Values Add

Sample_ID ×

Summarize by: COUN... Show as: Default

Filters Add

Sheet1 Pivot Table 1

	A	B	C	D	E	F	G	H
1	Host Bean	Enterobacteriaceae	Enterococcaceae	Staphylococcaceae	Abundance			
2	Adzuki	4		6	10			
3	Blackeye pea		1	9	10			
4								
5								

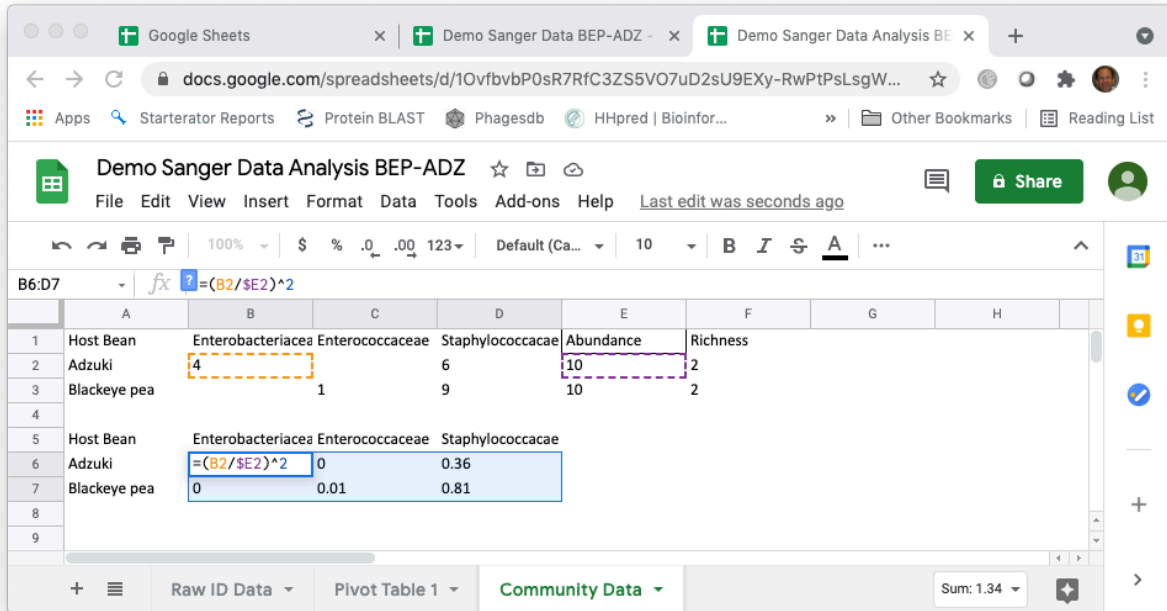
Calculating diversity indices

1. Species richness – the number of unique species 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")).

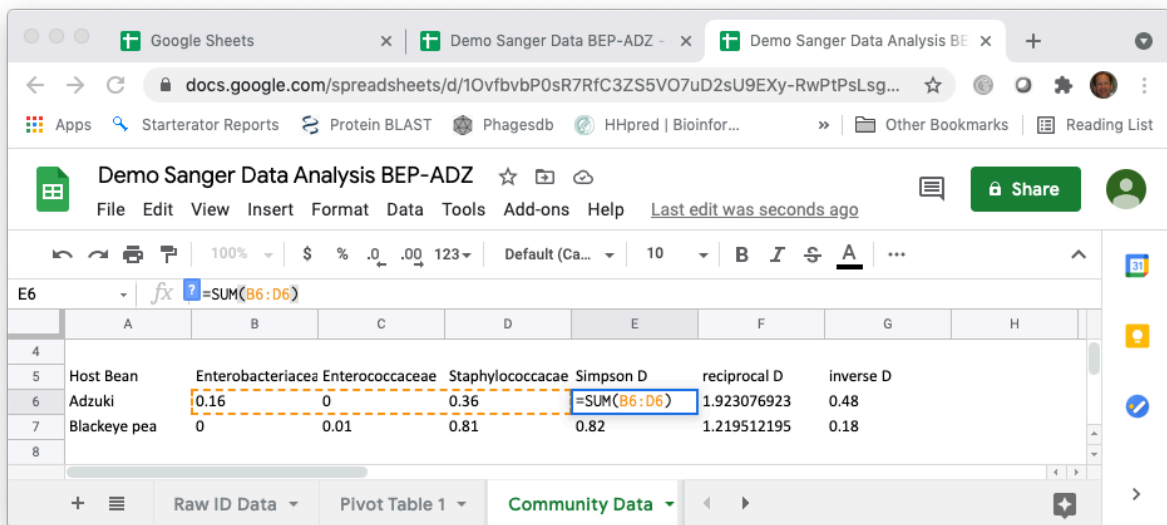
	A	B	C	D	E	F	G	H
1	Host Bean	Enterobacteriaceae	Enterococcaceae	Staphylococcaceae	Abundance	Richness		
2	Adzuki	4		6	10	=COUNTIF(B2:D2, ">0")		
3	Blackeye pea		1	9	10	2		
4								
5								

2. Simpson Index – the Simpson Index incorporates both species richness and species evenness.
 - a. $D = \sum (n/N)^2$, where n=number of individuals of a particular species and N=total number of individuals in a sample. D increases as diversity decreases, which is counterintuitive.
 - b. Reciprocal Simpson – $1/D$

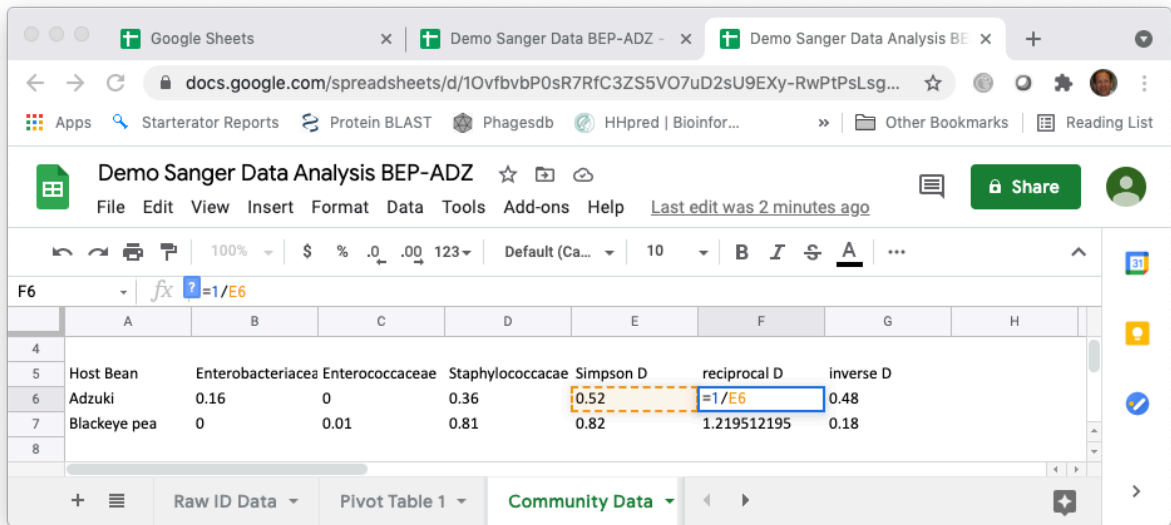
- c. Inverse Simpson – 1-D
- d. Using the total abundance for each treatment, calculate the proportions squared. Using the Google Sheets trick that \$ before a column or row prevents the program from iterating when copying a formula makes this easy.



- e. Calculate the sum of the proportions squared ($=SUM(B6:D6)$) for the first row) to calculate the Simpson Index.

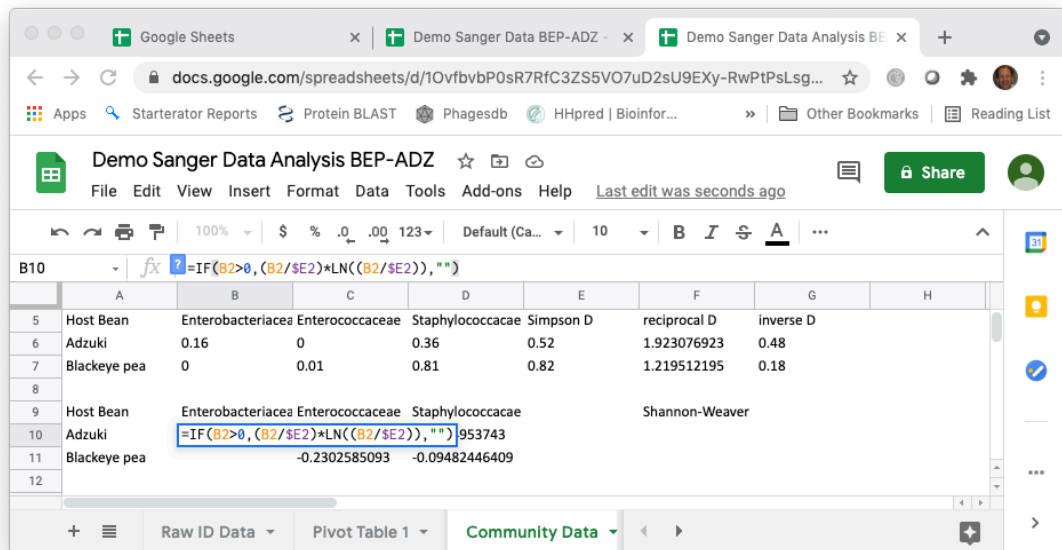


- f. Calculate the reciprocal and inverse Simpson using formulas in Google Sheets.



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

- $H = -\sum p \ln p$, where p is the proportion of individuals of each species in a community (i.e., n/N).
- Using the grand totals for each treatment, calculate the proportions. Using the Excel trick that \$ before a column or row prevents Excel from iterating when copying a formula makes this easy.
- Note that $\ln p$ is undefined if $p=0$, so you can use an "IF" statement in Excel to prevent the calculation of undefined values. For example, `=IF(B2>0,(B2/$E2)*LN((B2/$E2)),"")`



- d. Complete the calculation of the Shannon-Weaver Index by calculating the negative sum of each row ($=-SUM(B10:D10)$ for the first row).

The screenshot shows a Google Sheets spreadsheet titled "Demo Sanger Data Analysis BEP-ADZ". The spreadsheet contains a table with the following data:

	A	B	C	D	E	F	G	H
5	Host Bean	Enterobacteriaceae	Enterococcaceae	Staphylococcaceae	Simpson D	reciprocal D	inverse D	
6	Adzuki	0.16	0	0.36	0.52	1.923076923	0.48	
7	Blackeye pea	0	0.01	0.81	0.82	1.219512195	0.18	
8								
9	Host Bean	Enterobacteriaceae	Enterococcaceae	Staphylococcaceae				Shannon-Weaver
10	Adzuki	-0.3665162927		-0.3064953743				$=-SUM(B10:D10)$
11	Blackeye pea		-0.2302585093	-0.09482446409				0.3250829734
12								

Questions

1. Based on the diversity indices that you calculated, which treatment had the highest (lowest) diversity?
2. Does the answer depend on the measure of diversity that you use?
3. Do your answers to the questions above depend on the taxonomic level of analysis?
4. Do your conclusions based on the analysis of picked colony Sanger sequences agree with your conclusions based on the colony phenotype analysis? If they do not agree, what could cause the difference between these analyses?

This study is based on Blumer LS, Beck CW 2020. **Introducing community ecology and data skills with the bean beetle microbiome project.** Article 24 In: McMahon, K editor. *Advances in Biology Laboratory Education* Volume 41 . Publication of the 41st Conference of the Association for Biology Laboratory Education (ABLE) <https://doi.org/10.37590/able.v41.art24>

Last edited by LSB 18 May 2023.