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.

	Demo San File Edit	ger Data BEP View Insert F	P-ADZ ☆ Format Data	⊡ ⊘ Tools Extension	s Help	Ū 🗏		
C	2 5 2 8	k ¶ 100%	▼\$%·	0, .00 123 A	rial -	- 14 + :		^
1:1		Sample_ID						
	A	В	с	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	Staphylococcacae	Staphylococcus	
5	T4-ADZ	adult	female	Adzuki	LB	Staphylococcacae	Staphylococcus	
6	T5-ADZ	adult	female	Adzuki	LB	Staphylococcacae	Staphylococcus	
7	T6-ADZ	adult	female	Adzuki	LB	Staphylococcacae	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	Staphylococcacae	Staphylococcus	
11	T11-ADZ	adult	female	Adzuki	LB	Staphylococcacae	Staphylococcus	
12	T12-BEP	adult	female	Blackeye pea	LB	Staphylococcacae	Staphylococcus	
13	T13-BEP	adult	female	Blackeye pea	LB	Staphylococcacae	Staphylococcus	
14	T14-BEP	adult	female	Blackeye pea	LB	Staphylococcacae	Staphylococcus	
15	T16-BEP	adult	female	Blackeye pea	LB	Enterococcaceae	Enterococcus	
16	T17-BEP	adult	female	Blackeye pea	LB	Staphylococcacae	Staphylococcus	
17	T18-BEP	adult	female	Blackeye pea	LB	Staphylococcacae	Staphylococcus	
18	T19-BEP	adult	female	Blackeye pea	LB	Staphylococcacae	Staphylococcus	
19	W1-BEP	adult	female	Blackeye pea	LB	Staphylococcacae	Staphylococcus	
20	W2-BEP	adult	female	Blackeye pea	LB	Staphylococcacae	Staphylococcus	
21	W3-BEP	adult	female	Blackeye pea	LB	Staphylococcacae	Staphylococcus	
22								

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 San File Edit				∎ Too		ensions	Help				3		
C	< 5 e E	k 🖥 100%	6 🕶 \$	%	.0,	.0 <u>0</u> 123	:		^		Pivot tab	le edit	or	
I	▼ fx C	OUNTA of Sam	ple_ID							1.				
	А	В		С		D		Е			Suggested			~
	COUNTA of Sar													
	Host Bean	Enterobacteri	ace Entero	coccace	a Stap	phylococo	caca Grar	nd Total			Rows		Add	
	Adzuki		4				6		10		Rows			
	Blackeye pea				1		9		10		Host Bean		×	
	0													
											Order	Sort by		
											Ascen 🔻	Host	B 🔻	
											Show tota	als		
2											Columns		Add	
3														
ł											Family		×	
5											Order	Sort by		
ì											Ascen 🔻	Famil	y 👻	
,											Show tota	ale		
											Show tota			
)														
)											Values		Add	
1														
2											Sample_ID		×	
3											Summarize by	Show a	s	
+ 5											COUN		ilt 👻	
5												Serae	··· ·)	
,										* *				
									4 F		Filters		Add	

	Google Sheets	× 📔 Demo	Sanger Data BEP-ADZ - 🗙	Denio Sanger		+ (
\leftarrow	→ C	/spreadsheets/d/10v	fbvbP0sR7RfC3ZS5VO7u	uD2sU9EXy-RwPtPs	iLsg ☆ 🛞	o 🗯 🌑
A	Apps 🔍 Starterator Reports 😪	Protein BLAST 🏟 P	hagesdb 🕜 HHpred Bio	vinfor »	Dther Bookmarks	📰 Reading Li
	Demo Sanger Data Ana	alysis BEP-ADZ	☆ ⊡ ⊘			Share
■	File Edit View Insert Fo	ormat Data Tools	Add-ons Help Last e	edit was seconds age		
k	> ~ 🖶 🔁 100% → \$	% .0 .00 123▼	Default (Ca 👻 10	- В <i>І</i> - А		^
•		← →			-	8
8	- <i>f</i> x			-		
	A B	С	D E	F	G H	
1		interococcaceae Staphy				
2	Adzuki 4	6	10			
3	Blackeye pea 1	9	10			
4	-					-
5						
						- >
	+	Pivot Table 1 👻	Community Data 👻	▲ ▶		

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")).

$\leftarrow \rightarrow$				beine ounger b	ata BEP-ADZ - 🗙	🛨 Demo San	ger Data Analysis	BE × +		0
	> C 🔒	docs.google.com	n/spreadsheet	s/d/10vfbvbP0s	R7RfC3ZS5VO7	uD2sU9EXy-Rwl	PtPsLsgW6Q/	☆ ⑧ 0) 🛪 🌘) :
Ap	ps 🔍 Star	terator Reports 🛭 😒	Protein BLAST	Phagesdb	⊘ HHpred Bio	pinfor	» 🗎 Oth	ner Bookmarks	🖽 Read	ing List
		Sanger Data An t View Insert F	,			edit was seconds	ago	🗎 🔒 SI	hare	9
5	~ 6 7	≞ 100% - \$	% .000_	123 - Default (Ca 10	- В <i>І</i> -5	A 🗟 🖽	22 ····	^	31
F2	- fx	COUNTIF B2:D2	2,">0")							-
	A	В	С	D	E	F	G	н		
1 H	Host Bean	Enterobacteriacea	Enterococcaceae	Staphylococcaca	e Abundance	Richness				
2 A	Adzuki	4		6	10	=COUNTIF(B2:D2	,">0")			
3 B	Blackeye pea		1	9	10	2				
4									*	-
5									*	
									4 1	
-	+ ≣	Raw ID Data 👻	Pivot Table	1 - Comm	unity Data 🔻				+	>

- 2. Simpson Index the Simpson Index incorporates both species richness and species evenness.
 - a. D = $\Sigma(n/N)^2$, 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.

	Goog	gle Sheets	× 🖬	Demo Sanger Da	ta BEP-ADZ - 🗙	🚹 Demo San	ger Data Analysis	BE × +		0
\leftarrow	→ C 🏻	docs.google.cor	n/spreadsheets,	/d/10vfbvbP0sR	R7RfC3ZS5VO7	ID2sU9EXy-Rwl	PtPsLsgW	☆ ⓒ ♀	* 6) i
A	Apps 🔍 Starte	erator Reports 🗧	Protein BLAST	Phagesdb	MHpred Bio	infor	» 🗎 Othe	r Bookmarks	🔝 Readi	ing List
		anger Data Ar View Insert F				dit was seconds		🗏 🔒 Sh	are	0
1		100% 🚽 💲	% .000_ 1	23 - Default (0	Ca 👻 10	• В <i>І</i> \$	Α		^	31
36:D7	• <i>f</i> X	?=(B2/\$E2)^2								-
	A	В	С	D	E	F	G	Н		
1	Host Bean	Enterobacteriacea	Enterococcaceae	Staphylococcacae		Richness				
2	Adzuki	4		6	10	2				
3	Blackeye pea		1	9	10	2				•
4	Host Bean	Entorobactoriacoa	Enterococcaceae	Stanbulacoccasa						
6	Adzuki		0	0.36	1					
7	Blackeye pea	0	0.01	0.81						
8					1					+
9	1								*	
									4 1	
	+ ≡ R	law ID Data 👻	Pivot Table 1	Commi	unity Data 🔻			Sum: 1.34 👻	+	>

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

	> C 🔒	docs.google.co	m/spreadsheets	s/d/10vfbvbP0s	R7RfC3ZS5VO7	uD2sU9EXy-Rw	/PtPsLsg ☆	· © 0	* 🚱
Ap	ops 🔍 Starte	rator Reports	Protein BLAST	Phagesdb	💮 HHpred Bio	pinfor	» 🗎 Other B	ookmarks 🛛 🔝	Reading Li
	Demo Sa	anger Data A	nalysis BEP-A	ADZ 🕁 🗈	Ø			🔒 Shar	
	File Edit	View Insert	Format Data	Tools Add-on	s Help Last	edit was second	<u>s ago</u>		
ŝ	~ 8 7	100% - \$	%.000_1	123 v Default	(Ca 👻 10	- B I S	<u>A</u>		^ 31
•	- fx	<pre>2=SUM(B6:D6)</pre>							
5	 <i>f</i>X 	<pre>B</pre> B	С	D	E	F	G	Н	
	0	В	_	_		F	G	Н	
4	0	B	C a Enterococcaceae	_	e Simpson D	F reciprocal D	G inverse D	Н	
4 5 H	A	В	_	_			_	Н	
4 5 6	A Host Bean	B	a Enterococcaceae	Staphylococcaca	e Simpson D	reciprocal D	inverse D	Н	
4 5 6	A Host Bean Adzuki	B Enterobacteriace 0.16	a Enterococcaceae 0	Staphylococcaca 0.36	e Simpson D =SUM(B6:D6)	reciprocal D 1.923076923	inverse D 0.48	Н	

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

	Goog	gle Sheets	× 🕇	Denio Saliger Da	ata BEP-ADZ - 🗙	Denio San	ger Data Analysis B	× +	•
<u></u>	→ C 🏔	docs.google.com	n/spreadsheets	/d/10vfbvbP0sF	R7RfC3ZS5VO7	uD2sU9EXy-Rw	PtPsLsg 🕁	🛞 O 🗯 🍓	
A	pps 🔍 Starte	erator Reports 🗧 🗧	Protein BLAST	🏟 Phagesdb	MHpred Bio	oinfor	» 🛅 Other Book	kmarks 🛛 🔝 Readir	ng Lis
					-				-
	Demo Sa	anger Data Ar	alysis BEP-A	DZ 🕁 🗊	\odot			0.01.000	
⊞		View Insert I	Format Data	Toola Add-one		odit was 2 minut		🛱 Share	÷
	File Eult	view insert i	-onnat Data	TOOIS Add-Olis	s neip <u>Last</u>	euit was z minute	<u>es ago</u>		
lic.		100% - \$	% .000_ 1	23 - Default (Ca 10	- В <i>І</i> -Ş	A	~	_
			← →						31
,	- <i>fx</i>	?=1/E6							
	A	В	С	D	E	F	G	н	
4									2
5	Host Bean	Enterobacteriacea	Enterococcaceae	Staphylococcacae	Simpson D	reciprocal D	inverse D		
6	Adzuki	0.16	0	0.36	0.52	=1/E6	0.48		•
7	Blackeve pea	0	0.01	0.81	0.82	1.219512195	0.18		V
	Diacite / C pea	•			0.01		0.20	-	
R								4 1	
В			Pivot Table		unity Data 👻				>
8	+ ≡ R	aw ID Data 👻							

- 3. Shannon-Weaver (Shannon-Weiner) Index also incorporates species richness and species evenness
 - a. H=-Σ*p*ln*p*, where p is the proportion of individuals of each species in a community (i.e., n/N).
 - b. 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.
 - c. Note that lnp 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)),"")

🔍 🛨 Goo	gle Sheets	× 🖬	Demo Sanger Da	ta BEP-ADZ – 🗙	🛨 Demo San	ger Data Analysis	BE × +	0
→ C 🏻	docs.google.com	n/spreadsheets	/d/10vfbvbP0sR	R7RfC3ZS5V07u	uD2sU9EXy-Rwl	PtPsLsg ☆	© 0 \$	() :
pps 🔍 Start	erator Reports 🗧	Protein BLAST	Phagesdb	🖉 HHpred Bio	infor ×	Other Boo	okmarks 🛛 🖽 R	eading List
File Edit	View Insert F	ormat Data	Tools Add-ons	Help Last e			â Share	9
~ 8 7	100% 🚽 \$	% .000_ 1	23 v Default (C	Ca + 10	• В <i>І</i> \$	A		^ 🛐
- <i>f</i> X	2=IF(82>0,(82/	\$E2)*LN((<mark>B2</mark> /\$E	2)),"")					
A	В	С	D	E	F	G	н	
Host Bean	Enterobacteriacea	Enterococcaceae	Staphylococcacae	Simpson D	reciprocal D	inverse D		
Adzuki	0.16	0	0.36	0.52	1.923076923	0.48		
Blackeye pea	0	0.01	0.81	0.82	1.219512195	0.18		
					Shannon-Weaver			
	=IF(B2>0,(B2/\$							
Blackeye pea		-0.2302585093	-0.09482446409					
Біаскеуе реа								A
blackeye pea								*
	Raw ID Data 👻	Pivot Table 1		unity Data 👻			•	×
	C C C C C C C C C C C C C C C C C C C	→ C docs.google.cor pps Starterator Reports Demo Sanger Data Ar File Edit View Insert F T 100% ✓ \$ fx ?=IF(B2>0, (B2/4 A Blackeye pea O Host Bean Enterobacteriacea Adzuki Enterobacteriacea Adzuki Enterobacteriacea Adzuki Enterobacteriacea Adzuki Enterobacteriacea	→ C	→ C doos.google.com/spreadsheets/d/10vfbvbP0sF ppps Starterator Reports Protein BLAST Phagesdb Demo Sanger Data Analysis BEP-ADZ ☆ File Edit View Insert Format Data Tools Add-ons a a b c fix f=IF(B2>0, (B2/\$E2)*LN((B2/\$E2)), "") A B c D Host Bean Enterobacteriacea Enterococcaceae Staphylococcacae Adzuki 0.16 0 0.36 Blackeye pea 0 0.01 0.81	→ C docs.google.com/spreadsheets/d/10vfbvbP0sR7RfC3ZS5V07u pps Starterator Reports Protein BLAST Phagesdb W HHpred Bio Demo Sanger Data Analysis BEP-ADZ Pagesdb HHpred Bio File Edit View Insert Format Data Tools Add-ons Help Last.et A B C D E Host Bean Enterobacteriacea Enterococcaceae Staphylococcacae Simpson D Adzuki 0.16 0 0.36 0.52 Blackeye pea 0 0.01 0.81 0.82 Host Bean Enterobacteriacea Enterococcaceae Staphylococcacae Staphylococcacae Staphylococcacae Adzuki 0.16 0.81 0.82 Host Bean Enterobacteriacea Enterococcacaeae Staphylococcacae Staphylococcacae Staphylococcacae Adzuki e=1F(82>0, (82/\$E2)*LN((82/\$E2))*LN((82/\$E2))*, "") 953743	→ C docs.google.com/spreadsheets/d/10vfbvbP0sR7RfC3ZS5VO7uD2sU9EXy-Rwl pps Starterator Reports Protein BLAST Phagesdb HHpred Bioinfor Demo Sanger Data Analysis BEP-ADZ Pagesdb HHpred Bioinfor Pile Edit View Insert Format Data Tools Add-ons Help Last edit was seconds → D Pile Pile Edit View Insert Format Data Tools Add-ons Help Last edit was seconds → D Pile Pile Edit View Insert Format Data Tools Add-ons Help Last edit was seconds Pile Edit View Insert Format Data Tools Add-ons Help Last edit was seconds Pile Edit View Insert Format Data Tools Add-ons Help Last edit was seconds Pile Edit View Insert Format Data Tools Add-ons Help Last edit was seconds Pile Edit View Insert Format Data Tools Add-ons Help Last edit was seconds Pile Pile Edit View Insert Format Data Tools Add-ons Help Last edit was seconds Pile Pile Pile Pile Pile Pile Pile Pile	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	$\begin{array}{c c c c c c c c c c c c c c c c c c c $

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

\leftarrow	→ C 🗎	docs.google.com	n/spreadsheets	/d/10vfbvbP0sR	R7RfC3ZS5V07	7uD2sU9EXy-Rw	PtPsLsg ☆		F 🛞 🗄
A	Apps 🔍 Starte	erator Reports 😪	Protein BLAST	🏟 Phagesdb	🛞 HHpred Bi	ioinfor	» 🗎 Other Bo	okmarks 🔛	Reading Lis
	Demo Sa	anger Data Ar	alysis BEP-A	ADZ ☆ 🖻	\odot			O Chara	
▦		View Insert F	ormat Data	Tools Add-ons	B Help Last	edit was seconds		🛱 Share	
		100%	° 0 00 1	Defecto (10	D 7 0	•		
lic.		100% 🚽 \$	% .0 .00 1	23 v Default (0	Ga ≠ 10	- B I S	- A		^ 31
10	- fx	=-SUM(B10:D10)						
	A	В	С	D	E	F	G	н	
5	A Host Bean			D Staphylococcacae		F reciprocal D	G inverse D	Н	
5 6							-	н	
-	Host Bean	Enterobacteriacea	Enterococcaceae	Staphylococcacae	Simpson D	reciprocal D	inverse D	Н	
6	Host Bean Adzuki	Enterobacteriacea	Enterococcaceae 0	Staphylococcacae	Simpson D 0.52	reciprocal D 1.923076923	inverse D 0.48	н	
6 7	Host Bean Adzuki	Enterobacteriacea 0.16 0	Enterococcaceae 0 0.01	Staphylococcacae	Simpson D 0.52 0.82	reciprocal D 1.923076923	inverse D 0.48 0.18	н	
6 7 8 9	Host Bean Adzuki Blackeye pea	Enterobacteriacea 0.16 0	Enterococcaceae 0 0.01	Staphylococcacae 0.36 0.81	Simpson D 0.52 0.82	reciprocal D 1.923076923 1.219512195	inverse D 0.48 0.18	Н	
6 7 8	Host Bean Adzuki Blackeye pea Host Bean	Enterobacteriacea 0.16 0 Enterobacteriacea	Enterococcaceae 0 0.01	Staphylococcacae 0.36 0.81 Staphylococcacae	Simpson D 0.52 0.82	reciprocal D 1.923076923 1.219512195 Shannon-Weaver	inverse D 0.48 0.18	н	
6 7 8 9 10	Host Bean Adzuki Blackeye pea Host Bean Adzuki	Enterobacteriacea 0.16 0 Enterobacteriacea	Enterococcaceae 0 0.01 Enterococcaceae	Staphylococcacae 0.36 0.81 Staphylococcacae -0.3064953743	Simpson D 0.52 0.82	reciprocal D 1.923076923 1.219512195 Shannon-Weaver =-SUM(B10:D10)	inverse D 0.48 0.18	н	

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.