Tutorial for Microbial Community Analysis Using Google Sheets

Goal

Once you or someone else has processed next generation sequencing data from one or more microbial community samples, the next step is to use the resulting processed data files to ask questions about the bacterial community composition of samples. Here, we will walk through the steps to analyze such data using tools available in Google Sheets. Specifically, you will:

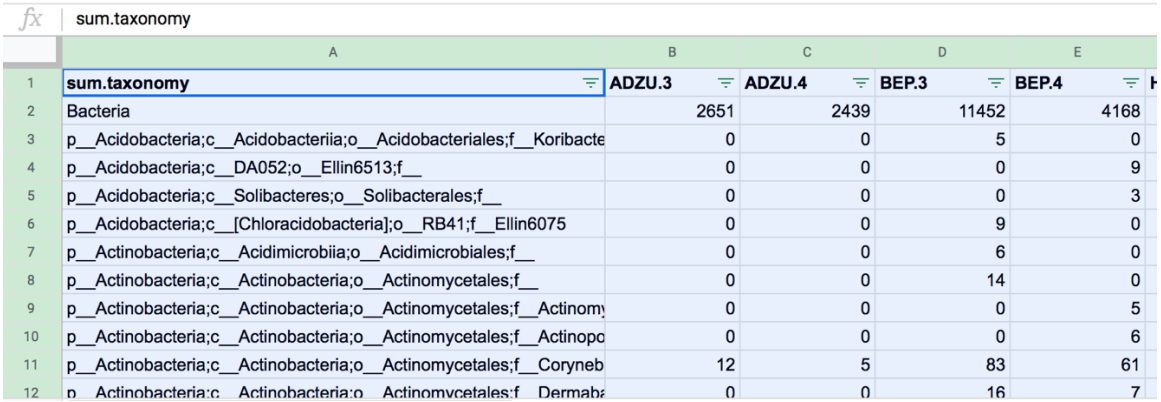
- Determine what are the taxa found in all samples.
- Calculate diversity indices that tell you something about species richness and evenness.
- Assess how similar or different communities are across samples.

The data file that you use in this tutorial should be the taxonomy file that was filtered previously in the **4. Preparing Files for Analyses** tutorial. In this file, the rows are the different bacterial taxa and the columns are the different samples.

What are the core taxa?

Core taxa are taxa that are found in all samples. In addition, we might consider taxa that are unique to only one treatment. The easiest way to determine these taxa in Excel is to use the Filter function for the columns.

1. To turn on filtering, click a cell in the top row
2. Click on `Data`, then click on `Create Filter`
3. You will now see small dashed inverted triangles in each cell of the top row like those seen below



fx sum.taxonomy

	A	B	C	D	E
1	sum.taxonomy	ADZU.3	ADZU.4	BEP.3	BEP.4
2	Bacteria	2651	2439	11452	4168
3	p__Acidobacteria;c__Acidobacteriia;o__Acidobacteriales;f__Koribacte	0	0	5	0
4	p__Acidobacteria;c__DA052;o__Ellin6513;f__	0	0	0	9
5	p__Acidobacteria;c__Solibacteres;o__Solibacterales;f__	0	0	0	3
6	p__Acidobacteria;c__[Chloracidobacteria];o__RB41;f__Ellin6075	0	0	9	0
7	p__Actinobacteria;c__Acidimicrobiia;o__Acidimicrobiales;f__	0	0	6	0
8	p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__	0	0	14	0
9	p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinom	0	0	0	5
10	p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinopo	0	0	0	6
11	p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Coryneb	12	5	83	61
12	p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Dermaba	0	0	16	7

4. To find the core taxa that are found in all samples, click on the dashed triangle from the first sample column.
5. Unselect “0” from the list. Then click OK.
6. Continue unselecting “0” for all samples sequentially until no cells in your table contain 0.
    - Note: “Bacteria” represents reads that were not classified to any taxonomic level other than to be identified as bacterial reads.
7. You should see many taxa get removed and fewer taxa remain. These taxa represent taxa that were present in all samples, in other words, the CORE taxa in the dataset.
8. Once you record all the CORE taxa, go to `Data` and click on `Turn off filter.` This will clear all your filters from each column so that you see all the taxa in the dataset again.

What are the unique taxa found in each treatment?

9. To find taxa that are unique to only one treatment, go back to `Data` and select `Create Filter`.
10. Select the dropdown for one of your samples for a particular treatment.
11. Click on `Filter by condition`
12. Scroll down and select `Is Equal To,` and input the value of 0. Then click ok.
13. Repeat this for all samples EXCEPT for the treatment for which you are trying to find the unique taxa.
    a. For example, if we were trying to find taxa unique to Adzuki bean diets from our sample dataset, we would do this by clicking on the dropdown for BEP and Hyacinth samples, and following steps 10-13. We would not do these steps for our Adzuki samples, only BEP and Hyacinth. The result will be taxa that are only found in Adzuki bean samples.
14. Once you record the taxa that are unique to one of your treatments, click on `Data` and `Turn off filter` to clear all filters so that you see all taxa again.
15. Repeat steps 9 through 14 for each of your treatments so that you record all unique taxa for each treatment.
16. `Turn off filter` from each column before proceeding to the next steps

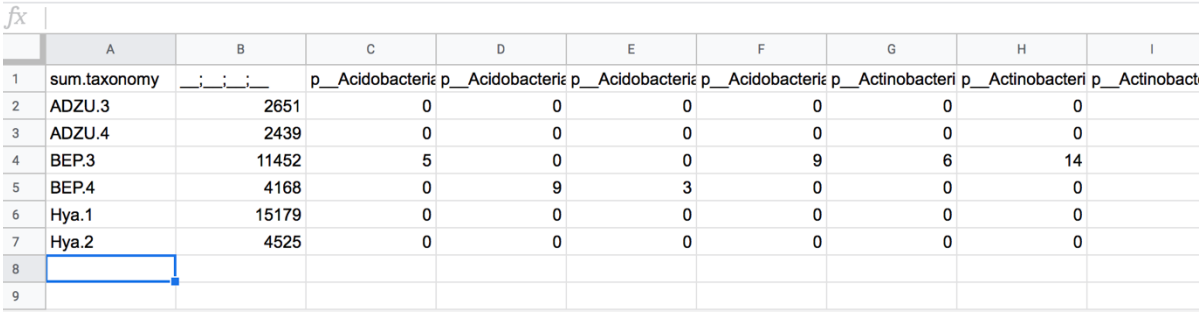
Questions Related to Core Taxa

1. How many taxa are in the core taxa?
2. Which are the most abundant taxa in the core taxa?
3. How many taxa are unique to each treatment and how many taxa are found in at least one sample in each treatment?
4. Are there particular taxonomic groups that tend to be found in one treatment and not the other?

Calculating Diversity Indices

Calculating diversity indices in Google Sheets is easier if the rows are the samples and the columns are the taxa. So, we need to transpose the data.

1. Make sure to clear all filters from the previous section before proceeding to the next steps.
2. Select all of the data, copy, and then transpose paste into a new worksheet (or new tab in your current worksheet) so that your data file looks like the photo below, with samples as rows and taxa as columns.

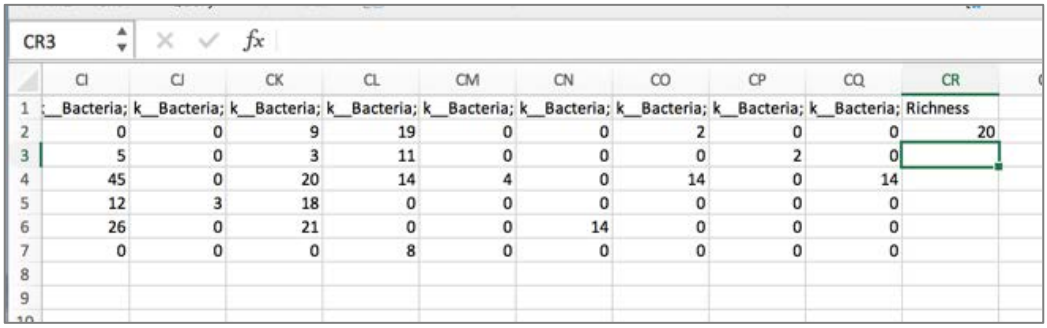


	A	B	C	D	E	F	G	H	I
1	sum.taxonomy	__;__;__;__	p__Acidobacteria	p__Acidobacteria	p__Acidobacteria	p__Acidobacteria	p__Actinobacteri	p__Actinobacteri	p__Actinobact
2	ADZU.3	2651	0	0	0	0	0	0	
3	ADZU.4	2439	0	0	0	0	0	0	
4	BEP.3	11452	5	0	0	9	6	14	
5	BEP.4	4168	0	9	3	0	0	0	
6	Hya.1	15179	0	0	0	0	0	0	
7	Hya.2	4525	0	0	0	0	0	0	
8									
9									

3. 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 sample, using the `COUNTIF` formula in Excel is easier
    b. Scroll to the very last column on your spreadsheet. For our **Beetle_Diet_Microbiome** sample dataset, the very last column with taxonomic data is cell CQ.
    c. In the very next column, (for our sample dataset that would be column CR1), type in `Richness` as the header of that column.
    d. Next, click on the cell directly below `Richness`, (in our sample dataset, that would be cell CR2)
    e. You will now type in a formula to count the number of cells that fulfill a specific requirement. The formula will use the first cell in your table with count data (B2) and the last cell of the first row of your table with count data (for example, CQ2). The formula should be structured as follows, but make sure to replace the highlighted cell value for the last cell of the first row that is specific to your data table.

i. `=COUNTIF(B2:CQ2,"">0"")`

f. Then hit RETURN on your keyboard to perform the calculation. If you are practicing with our sample dataset, you should see the value 20 has been computed in cell CR2.

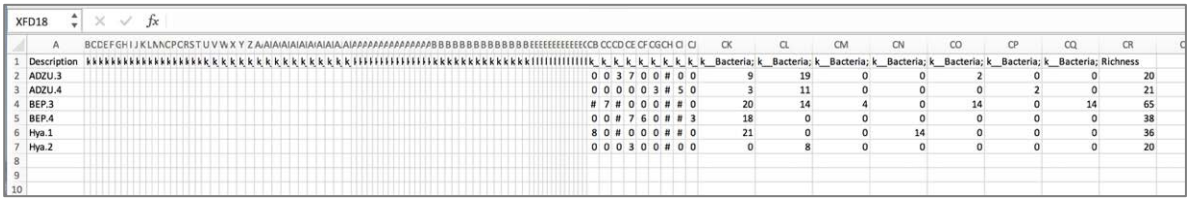


CR3

	CI	CJ	CK	CL	CM	CN	CO	CP	CQ	CR
1	;__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	Richness
2	0	0	9	19	0	0	2	0	0	20
3	5	0	3	11	0	0	0	2	0	
4	45	0	20	14	4	0	14	0	14	
5	12	3	18	0	0	0	0	0	0	
6	26	0	21	0	0	14	0	0	0	
7	0	0	0	8	0	0	0	0	0	
8										
9										

g. The `COUNTIF` formula will count the number of cells that fulfill a specific requirement. In our case, we are asking Excel to count cells only if they contain a value greater than zero ">0". We only want to count values greater than zero because we want to count the number of taxa present in a particular sample. The range B2:CQ2 represents the columns (B through CQ) that include all taxa in our dataset. The number 2 in B2:CQ2 represents data for our ADZU.3 sample, which is in row number 2. Our results show that ADZU.3 has a richness of 20, meaning out of all the taxa observed in our dataset, only 20 of them were present in ADZU.3

h. Let's repeat the process for all of our other samples. Remember that the column range (B through CQ) will stay the same, but the row we are performing the calculation for will change. B3:CQ3 will calculate richness for ADZU.4, B4:CQ4 will calculate richness for BEP.3, etc…



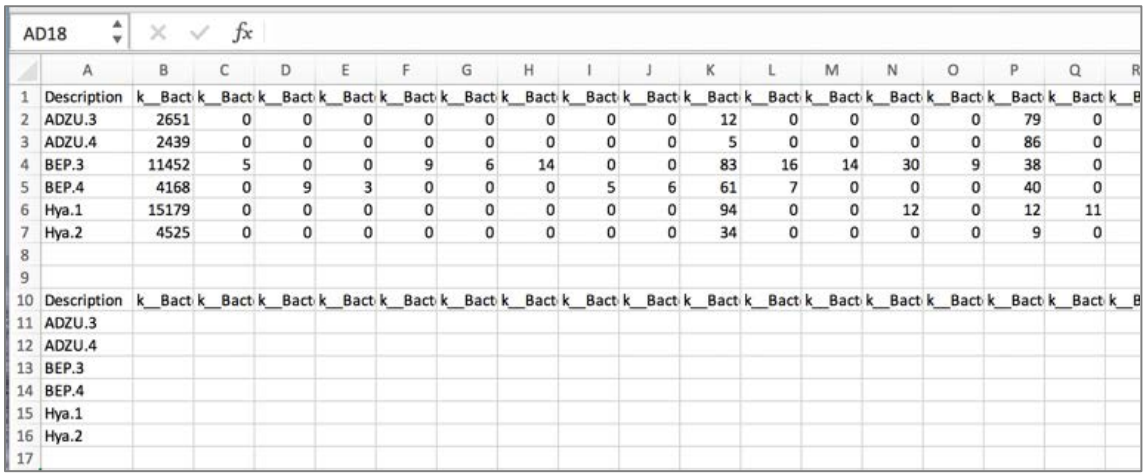
XFD18

	A	CK	CL	CM	CN	CO	CP	CQ	CR
1	Description	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	Richness
2	ADZU.3	9	19	0	0	2	0	0	20
3	ADZU.4	3	11	0	0	0	2	0	21
4	BEP.3	20	14	4	0	14	0	14	65
5	BEP.4	18	0	0	0	0	0	0	38
6	Hya.1	21	0	0	14	0	0	0	36
7	Hya.2	0	8	0	0	0	0	0	20
8									
9									
10									

Note: in the above photo, I collapsed the columns so that we could easily see which samples correspond to which richness values. You can also use the "Freeze First Column" option that will help you always view your samples as you scroll through your dataset.

4. Simpson Index – the Simpson Index incorporates both species richness and species evenness.
    - In ecology, D = Σ(n/N)^2, where n=number of individuals of a particular species and N=total number of individuals in a sample. In our case, each "species" is a taxa, and each "individual" is a sequence read.
    - In the above equation, D increases as diversity decreases, which is counterintuitive. That is why ecologists often use derivations of the index, such as the Reciprocal Simpson's (1/D) or the Inverse Simpson's (1-D), to describe diversity. These derivations allow diversity to be explained more intuitively, such that higher index values correspond to increases in diversity.
        a. Create a new data array below the original array by first copying the first row (that contains taxa names) and pasting it in an empty row below the dataset (for example row 10). You don't have to include the richness column, since we are done with that for now
        b. Then, copy the sample names and past them below row 10 (row 11, row 12, etc). In the end, your file should look something like this.

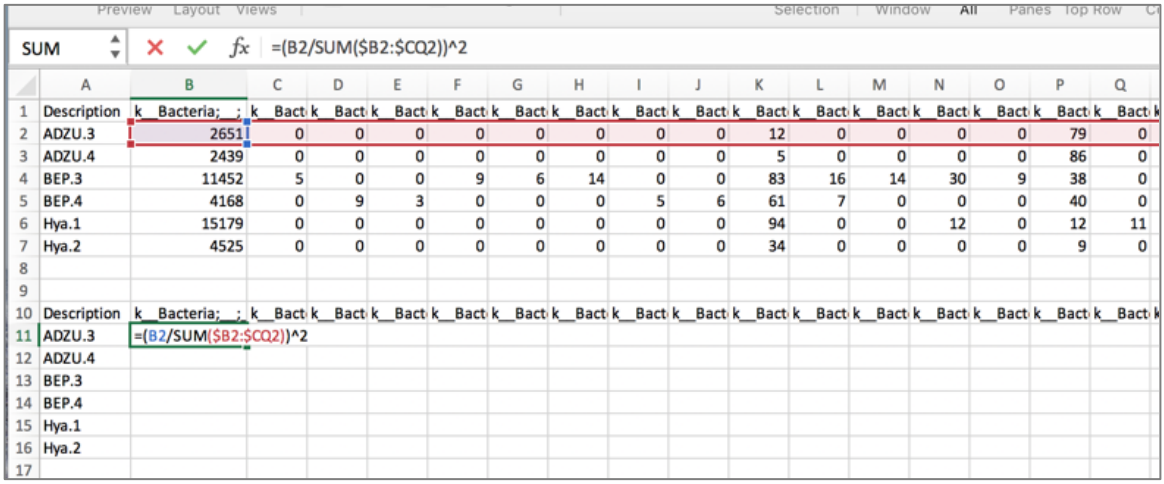
        c. To calculate the proportion squared for each taxa (*i.e.,* (n/N)^2), use the grand totals of reads for each taxa for each treatment.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Description	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__B
2	ADZU.3	2651	0	0	0	0	0	0	0	0	12	0	0	0	0	79	0	
3	ADZU.4	2439	0	0	0	0	0	0	0	0	5	0	0	0	0	86	0	
4	BEP.3	11452	5	0	0	9	6	14	0	0	83	16	14	30	9	38	0	
5	BEP.4	4168	0	9	3	0	0	0	5	6	61	7	0	0	0	40	0	
6	Hya.1	15179	0	0	0	0	0	0	0	0	94	0	0	12	0	12	11	
7	Hya.2	4525	0	0	0	0	0	0	0	0	34	0	0	0	0	9	0	
8																		
9																		
10	Description	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__B
11	ADZU.3																	
12	ADZU.4																	
13	BEP.3																	
14	BEP.4																	
15	Hya.1																	
16	Hya.2																	
17																		

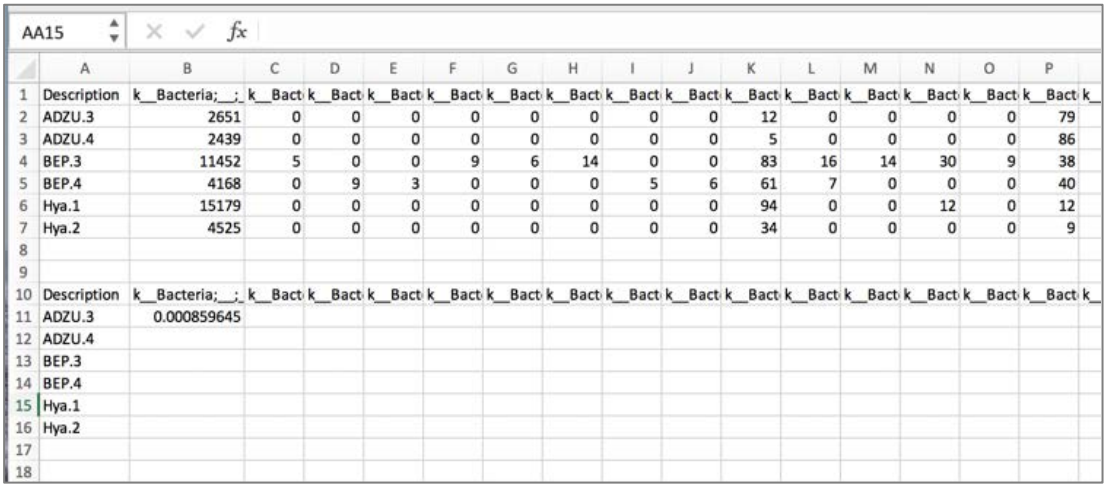
    - Using the Excel trick that $ before a column or row prevents Excel from iterating when copying a formula makes this easy. For example, `=(B2/SUM($B2:$CQ2))^2`

- Again, the value CQ2 in your formula will be specific to your dataset.



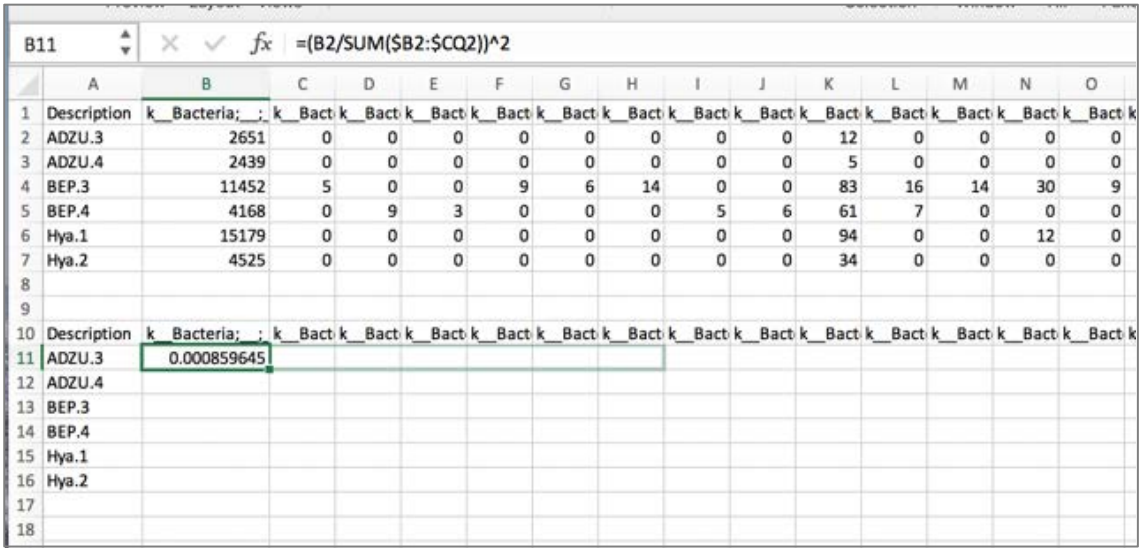
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Description	k__Bacteria;__;	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact
2	ADZU.3	2651	0	0	0	0	0	0	0	0	12	0	0	0	0	79	0
3	ADZU.4	2439	0	0	0	0	0	0	0	0	5	0	0	0	0	86	0
4	BEP.3	11452	5	0	0	9	6	14	0	0	83	16	14	30	9	38	0
5	BEP.4	4168	0	9	3	0	0	0	5	6	61	7	0	0	0	40	0
6	Hya.1	15179	0	0	0	0	0	0	0	0	94	0	0	12	0	12	11
7	Hya.2	4525	0	0	0	0	0	0	0	0	34	0	0	0	0	9	0
8																	
9																	
10	Description	k__Bacteria;__;	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact
11	ADZU.3	=(B2/SUM($B2:$CQ2))^2															
12	ADZU.4																
13	BEP.3																
14	BEP.4																
15	Hya.1																
16	Hya.2																
17																	

- Then hit RETURN to calculate the formula



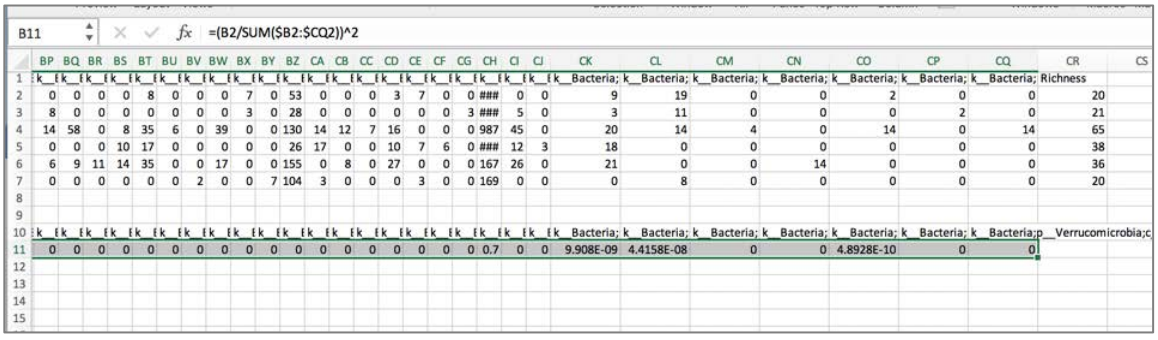
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Description	k__Bacteria;__;	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact
2	ADZU.3	2651	0	0	0	0	0	0	0	0	12	0	0	0	0	79
3	ADZU.4	2439	0	0	0	0	0	0	0	0	5	0	0	0	0	86
4	BEP.3	11452	5	0	0	9	6	14	0	0	83	16	14	30	9	38
5	BEP.4	4168	0	9	3	0	0	0	5	6	61	7	0	0	0	40
6	Hya.1	15179	0	0	0	0	0	0	0	0	94	0	0	12	0	12
7	Hya.2	4525	0	0	0	0	0	0	0	0	34	0	0	0	0	9
8																
9																
10	Description	k__Bacteria;__;	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact
11	ADZU.3	0.000859645														
12	ADZU.4															
13	BEP.3															
14	BEP.4															
15	Hya.1															
16	Hya.2															
17																
18																

- Copy the formula across the row by hovering your mouse over the bottom corner of the cell until you see a black cross, then click and drag across all cells in row 11. You should see a light blue outline across all the cells that are being dragged across. Keep dragging until you reach the last cell CQ2, then let go of your clicker to end click and drag. When you are finished, you should see values appear in all cells, indicating that the formula was copied and pasted in all cells of row 11 and the calculation was performed successfully.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Description	k__Bacteria;__;	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact
2	ADZU.3	2651	0	0	0	0	0	0	0	0	12	0	0	0	0
3	ADZU.4	2439	0	0	0	0	0	0	0	0	5	0	0	0	0
4	BEP.3	11452	5	0	0	9	6	14	0	0	83	16	14	30	9
5	BEP.4	4168	0	9	3	0	0	0	5	6	61	7	0	0	0
6	Hya.1	15179	0	0	0	0	0	0	0	0	94	0	0	12	0
7	Hya.2	4525	0	0	0	0	0	0	0	0	34	0	0	0	0
8															
9															
10	Description	k__Bacteria;__;	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact
11	ADZU.3	0.000859645													
12	ADZU.4														
13	BEP.3														
14	BEP.4														
15	Hya.1														
16	Hya.2														
17															
18															

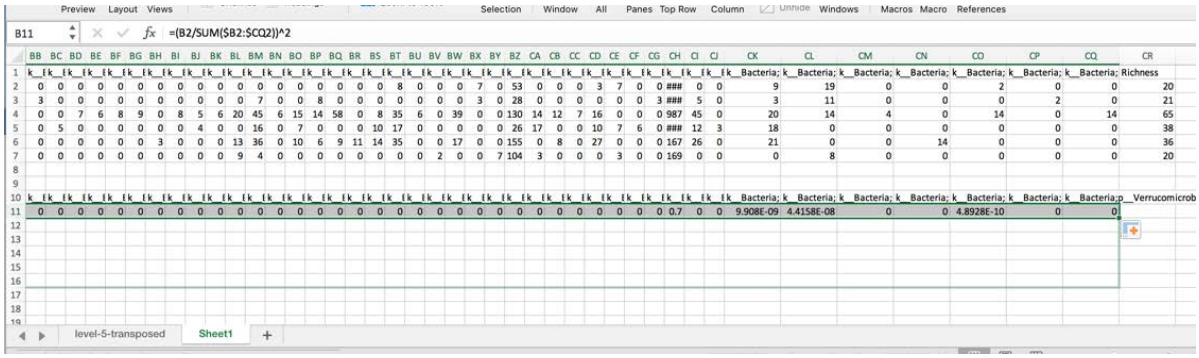
- You should also see that all the cells in this row are greyed out, and outlined in a dark blue outline.

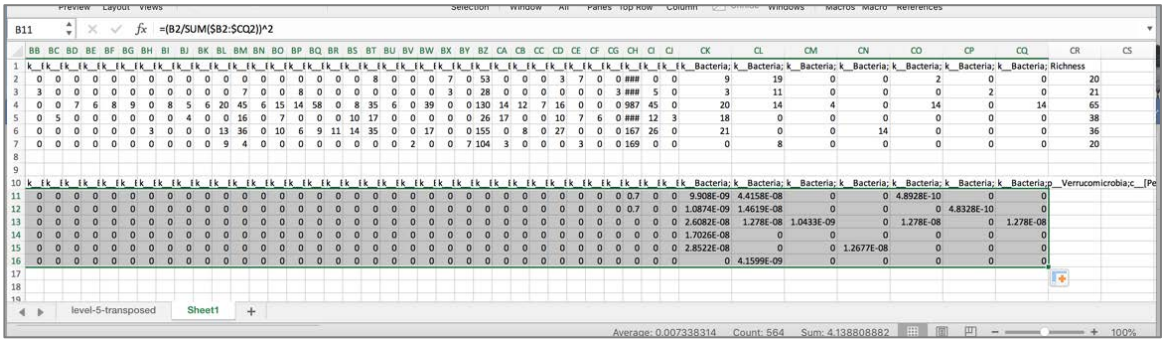


	BP	BQ	BR	BS	BT	BU	BV	BW	BX	BY	BZ	CA	CB	CC	CD	CE	CF	CG	CH	CI	CJ	CK	CL	CM	CN	CO	CP	CQ	CR	CS
1	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	k__Bacteria;	Richness	
2	0	0	0	0	8	0	0	0	7	0	53	0	0	0	3	7	0	0	###	0	0	9	19	0	0	2	0	0	20	
3	8	0	0	0	0	0	0	0	3	0	28	0	0	0	0	0	0	3	###	5	0	3	11	0	0	0	2	0	21	
4	14	58	0	8	35	6	0	39	0	0	130	14	12	7	16	0	0	0	987	45	0	20	14	4	0	14	0	14	65	
5	0	0	0	10	17	0	0	0	0	0	26	17	0	0	10	7	6	0	###	12	3	18	0	0	0	0	0	0	38	
6	6	9	11	14	35	0	0	17	0	0	155	0	8	0	27	0	0	0	167	26	0	21	0	0	14	0	0	0	36	
7	0	0	0	0	0	0	2	0	0	7	104	3	0	0	0	3	0	0	169	0	0	0	8	0	0	0	0	0	20	
8																														
9																														
10	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;	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;	p__Verrucomicrobia;c	
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0	0	9.908E-09	4.4158E-08	0	0	4.8928E-10	0	0		
12																														
13																														
14																														
15																														

- This means that the formula for all of these cells is still copied on the clipboard. It also means we can copy and paste this formula to the cells of the remaining samples.

- To do this, hover your mouse over the bottom corner of your dataset that would be equivalent to CQ11 (the last cell of the first row of the second data table you created).
- Next, click and drag down until the last cell of the last row in your new datat able (in our example, CQ16), which will copy the formula across all samples of our dataset. Then let go.
    - You should now see that the proportion squared has been calculated for each taxa in the dataset.



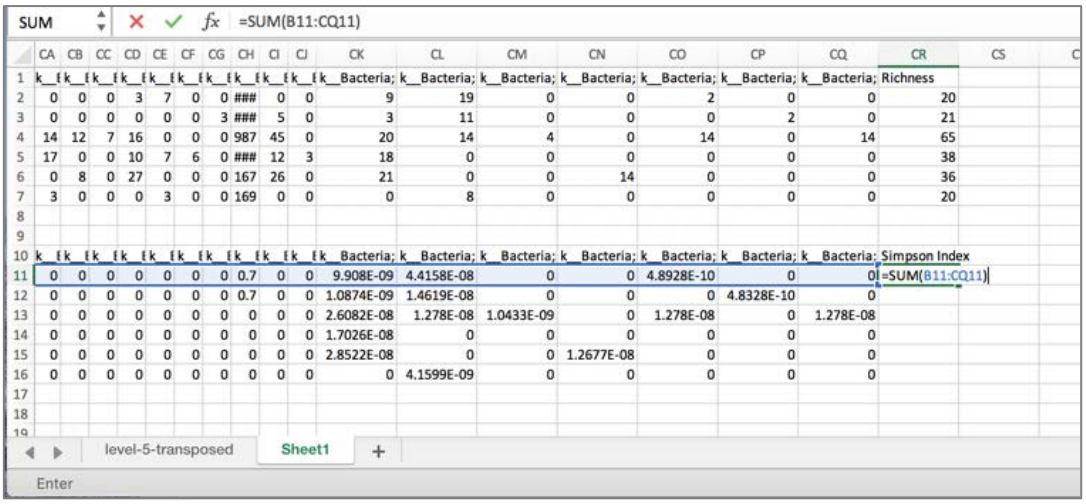


d. In a new column (in our example, cell CR10 would be a good choice), type in `Simpson Index`.

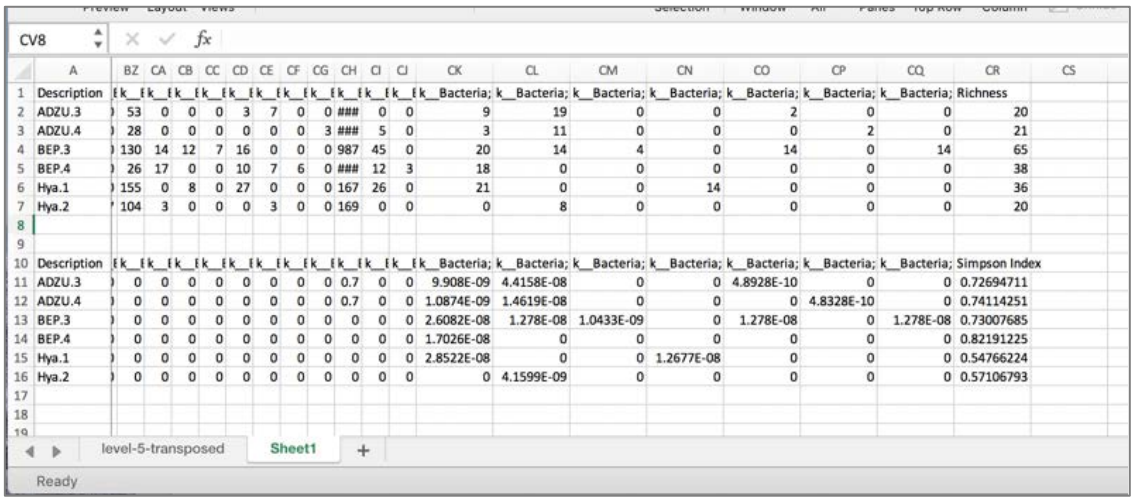
e. Then in the cell directly below that (C11), type in the following formula to calculate the sum of the proportions squared. The values B11 and CQ11 for those specific to your dataset.

- `=SUM(B11:CQ11)`

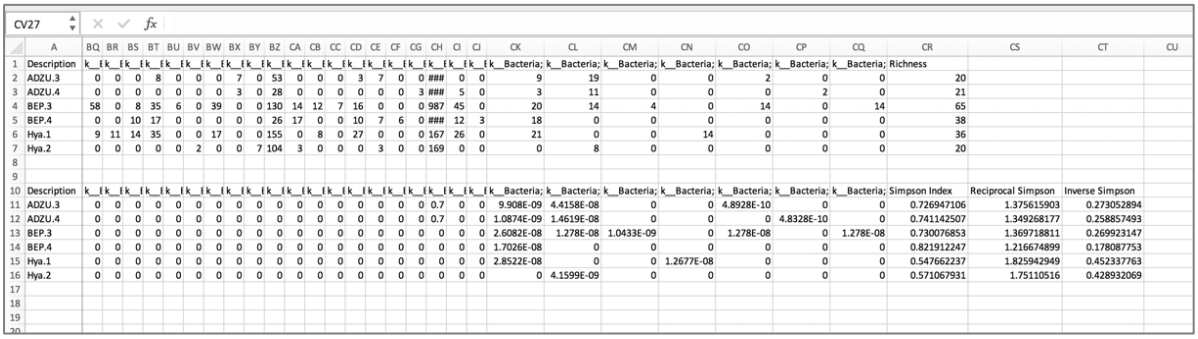
f. Then hit ENTER. The resulting value is the calculated Simpson Index for the first sample in your data set.



g. Copy and paste the formula into the rest of the cells under the `Simpson Index` column to calculate the Simpson index for all other samples. You can use the click/drag short-cut we learned previously, or you can manually copy and paste the formula, making sure to adjust the formula for each specific row (for example, `=SUM(B12:CQ12)` to calculate for sample ADZU.4, etc.)



5. Calculate the reciprocal and inverse Simpson using formulas in Excel.
    - Reciprocal Simpson
        - in cell CS10 name the column `Reciprocal Simpson`
        - in cell CS11, type in `=(1/CR11)` then hit ENTER
        - click and drag the formula into cells CR12 through CR16 to calculate for all other samples
    - Inverse Simpson
        - In cell CT10, name the column `Inverse Simpson`
        - In cell CT11, type in `=(1-CR11)` then hit ENTER
        - click and drag the formula into cells CT12 through CT16 to calculate for all other samples

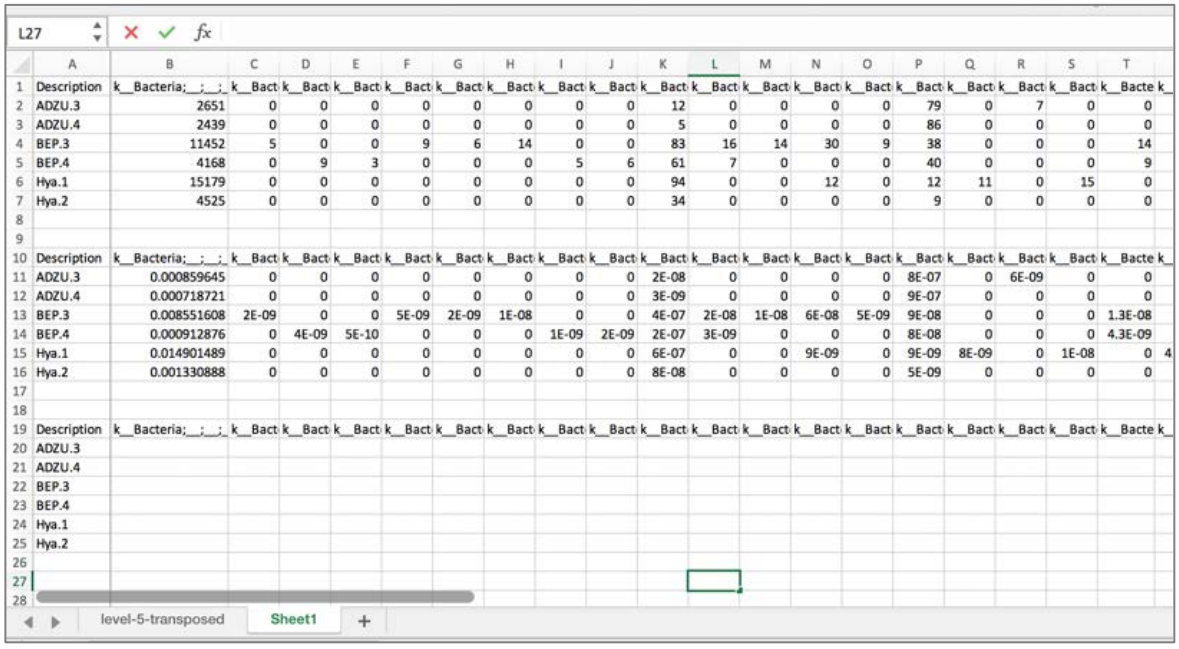


CV27

	A	BQ	BR	BS	BT	BU	BV	BW	BX	BY	BZ	CA	CB	CC	CD	CE	CF	CG	CH	CI	CJ	CK	CL	CM	CN	CO	CP	CQ	CR	CS	CT	CU
1	Description	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;	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;	Richness			
2	ADZU.3	0	0	0	8	0	0	0	7	0	53	0	0	0	3	7	0	0	###	0	0	9	19	0	0	2	0	0	20			
3	ADZU.4	0	0	0	0	0	0	0	3	0	28	0	0	0	0	0	0	3	###	5	0	3	11	0	0	0	2	0	21			
4	BEP.3	58	0	8	35	6	0	39	0	0	130	14	12	7	16	0	0	0	987	45	0	20	14	4	0	14	0	14	65			
5	BEP.4	0	0	10	17	0	0	0	0	0	26	17	0	0	10	7	6	0	###	12	3	18	0	0	0	0	0	0	38			
6	Hya.1	9	11	14	35	0	0	17	0	0	155	0	8	0	27	0	0	0	167	26	0	21	0	0	14	0	0	0	36			
7	Hya.2	0	0	0	0	0	2	0	0	7	104	3	0	0	0	3	0	0	169	0	0	0	8	0	0	0	0	0	20			
8																																
9																																
10	Description	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;	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;	Simpson Index	Reciprocal Simpson	Inverse Simpson	
11	ADZU.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0	0	9.908E-09	4.4158E-08	0	0	4.8928E-10	0	0	0.726947106	1.375615903	0.273052894	
12	ADZU.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0	0	1.0874E-09	1.4619E-08	0	0	0	4.8328E-10	0	0.741142507	1.349268177	0.258857493	
13	BEP.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.6082E-08	1.278E-08	1.0433E-09	0	1.278E-08	0	1.278E-08	0.730076853	1.369718811	0.269923147	
14	BEP.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.7026E-08	0	0	0	0	0	0	0.821912247	1.216674899	0.178087753	
15	Hya.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.8522E-08	0	0	1.2677E-08	0	0	0	0.547662237	1.825942949	0.452337763	
16	Hya.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.1599E-09	0	0	0	0	0	0.571067931	1.75110516	0.428932069	
17																																
18																																
19																																
20																																

6. Shannon-Weaver (Shannon-Weiner) Index – also incorporates species richness and species evenness
    - H=-Σ*p*ln*p*, where p is the proportion of individuals of each species in a community (i.e., n/N). In our case, it is the proportion of reads of each taxa in a community

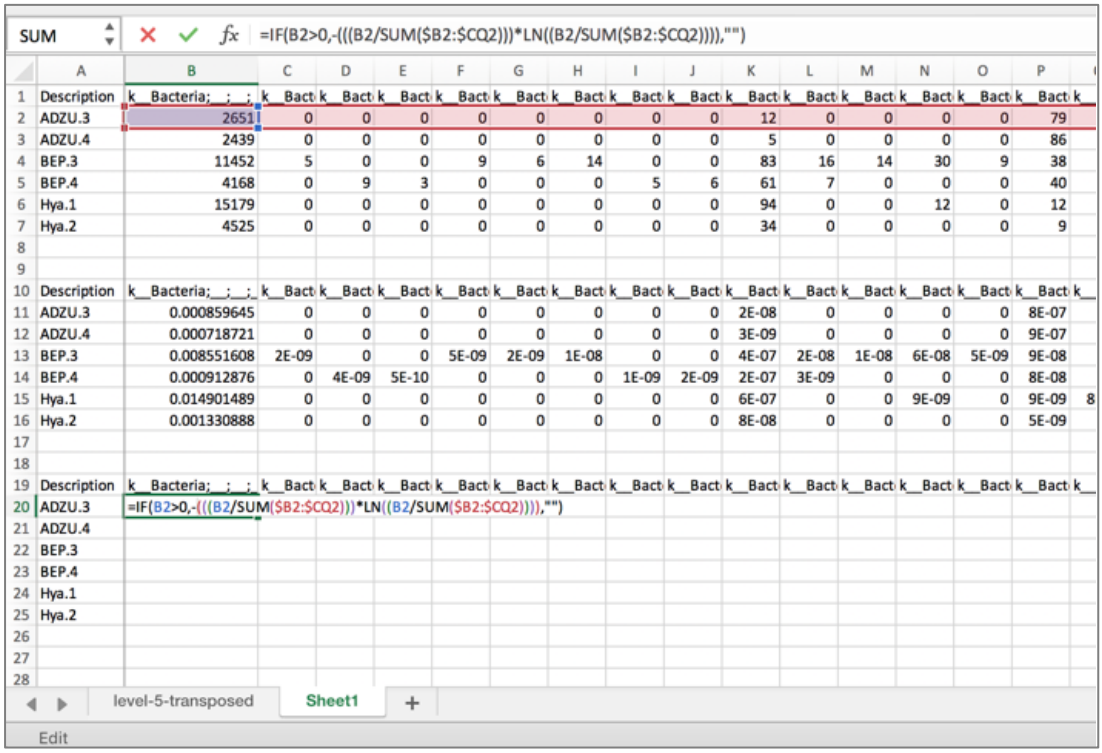
a. As we did before, create a new data array below the second data array using the same row labels (treatments/samples) and the same column labels (species). In the example below, I am starting the new array on row 19.



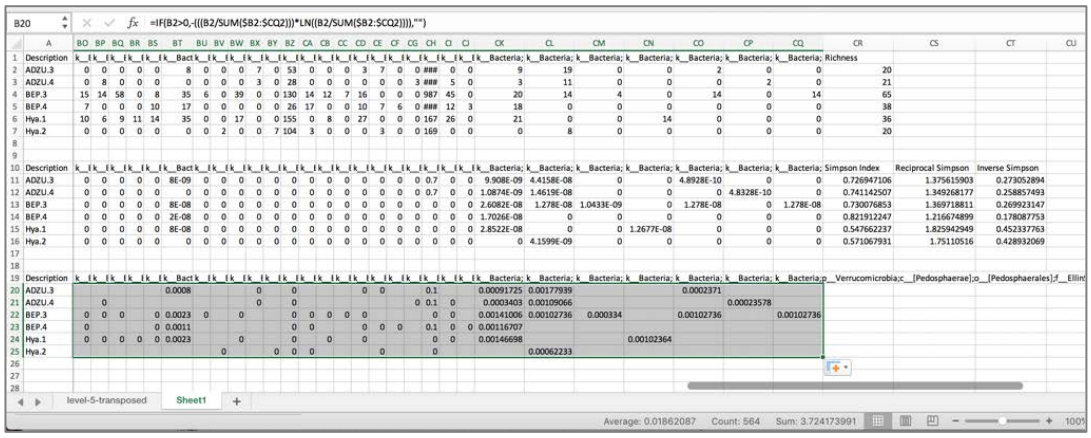
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Description	k__Bacteria;__;__;	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bacte
2	ADZU.3	2651	0	0	0	0	0	0	0	0	12	0	0	0	0	79	0	7	0	0
3	ADZU.4	2439	0	0	0	0	0	0	0	0	5	0	0	0	0	86	0	0	0	0
4	BEP.3	11452	5	0	0	9	6	14	0	0	83	16	14	30	9	38	0	0	0	14
5	BEP.4	4168	0	9	3	0	0	0	5	6	61	7	0	0	0	40	0	0	0	9
6	Hya.1	15179	0	0	0	0	0	0	0	0	94	0	0	12	0	12	11	0	15	0
7	Hya.2	4525	0	0	0	0	0	0	0	0	34	0	0	0	0	9	0	0	0	0
8																				
9																				
10	Description	k__Bacteria;__;__;	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bacte
11	ADZU.3	0.000859645	0	0	0	0	0	0	0	0	2E-08	0	0	0	0	8E-07	0	6E-09	0	0
12	ADZU.4	0.000718721	0	0	0	0	0	0	0	0	3E-09	0	0	0	0	9E-07	0	0	0	0
13	BEP.3	0.008551608	2E-09	0	0	5E-09	2E-09	1E-08	0	0	4E-07	2E-08	1E-08	6E-08	5E-09	9E-08	0	0	0	1.3E-08
14	BEP.4	0.000912876	0	4E-09	5E-10	0	0	0	1E-09	2E-09	2E-07	3E-09	0	0	0	8E-08	0	0	0	4.3E-09
15	Hya.1	0.014901489	0	0	0	0	0	0	0	0	6E-07	0	0	9E-09	0	9E-09	8E-09	0	1E-08	0
16	Hya.2	0.001330888	0	0	0	0	0	0	0	0	8E-08	0	0	0	0	5E-09	0	0	0	0
17																				
18																				
19	Description	k__Bacteria;__;__;	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bact	k__Bacte
20	ADZU.3																			
21	ADZU.4																			
22	BEP.3																			
23	BEP.4																			
24	Hya.1																			
25	Hya.2																			
26																				
27																				
28																				

level-5-transposed | Sheet1 | +

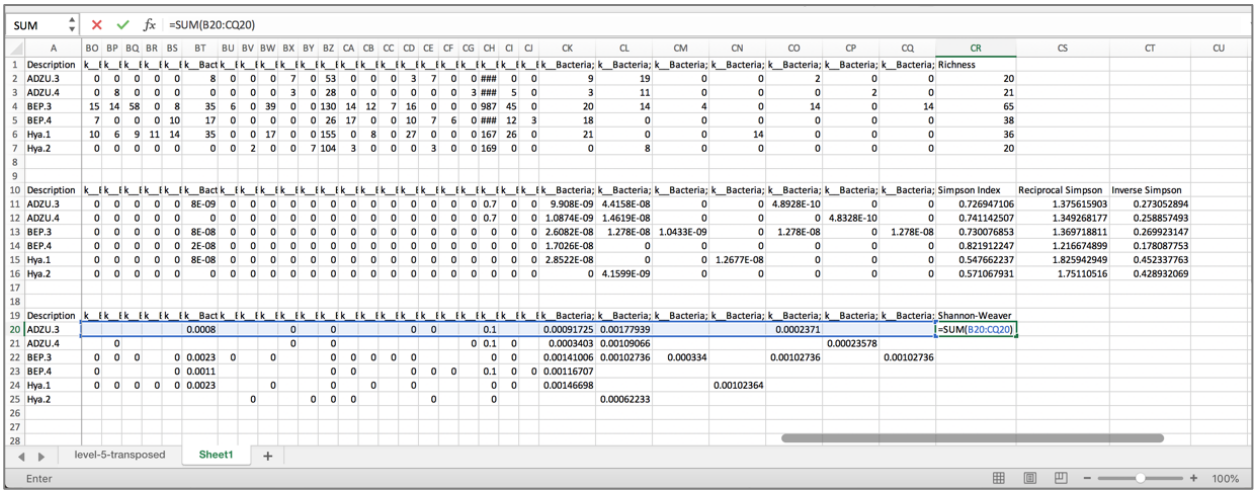
b. Using the grand totals for each treatment, calculate the proportions ($p \ln p$). Using the Excel trick that $ before a column or row prevents Excel from iterating when copying a formula makes this easy.
   - In the first empty cell of the array (in our example this cell is B20), type in the following formula with the appropriate values for each term from your dataset.
     =if(B2>0,-(((B2/SUM($B2:$CQ2)))*ln((B2/SUM($B2:$CQ2)))),"")

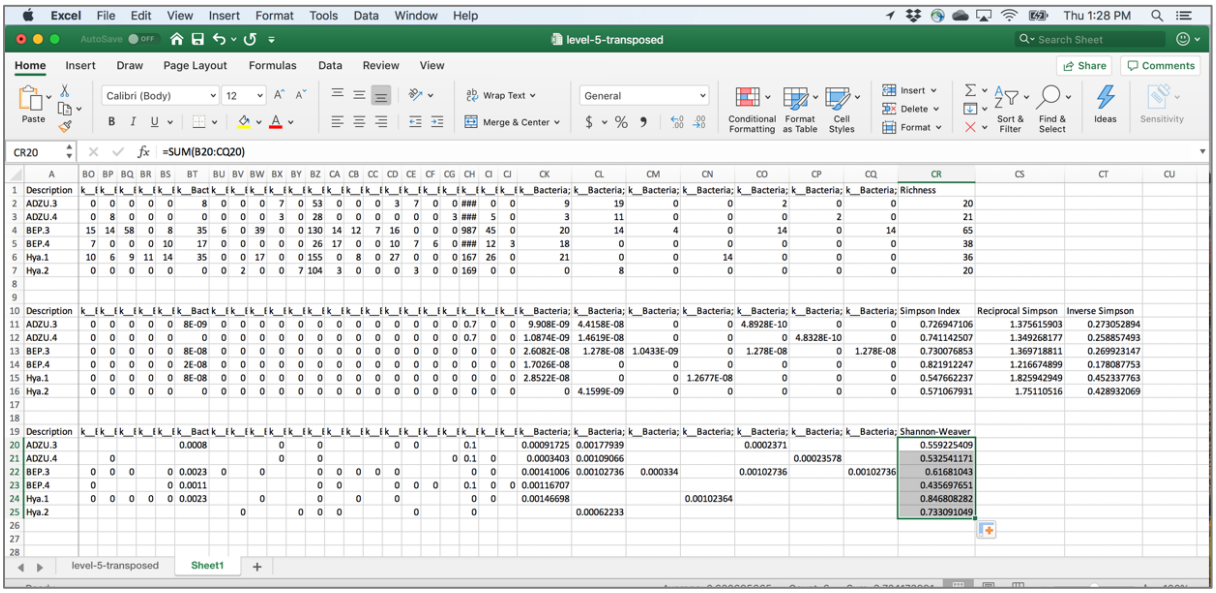


- Because *ln* of 0 is undefined, we use the `if` statement before the formula like to check whether the abundance of a taxon is greater than 0. If so, then the formula calculates *plnp*. If not, it makes the cell blank using two sets of quotation marks.
- Click and drag the formula into the other cells as we did before, first across for all cells in row 20, then down for all cells until row 25. You should see values filled in for all cells, unless they are blank due to 0 count values.



- Calculate the Shannon-Weaver index:
    i. in cell CR19, type in `Shannon-Weaver` as the column header
    ii. in cell CR20, type in the following formula `=SUM(B20:CQ20)` then hit ENTER to execute the calculation
    iii. Click and drag the formula down to the rest of the cells in that column (cells CR21 through CR25)





Questions Related to Diversity

1. Which treatment resulted in higher richness in our example dataset? Which resulted in lowest richness?
2. Which treatment resulted in highest and lowest Simpson's diversity?
3. Which treatment resulted in highest and lowest Shannon-Weaver diversity?
4. Do the samples that have highest richness also have highest diversity?
5. Based on the Simpson's and Shannon-Weaver diversities, infer which samples may be more even (have greater representation of many different types of species) and which samples may be less even (dominated by only a few species). If you performed Ranacapa analysis, reference your taxa-bar plots from Ranacapa. Do your inferences match the taxa-bar-plot data? Why or why not?

Calculating community similarity (distance)

Sometimes we are interested in how similar (or different) two communities are based on what species are present and the relative abundance of those species in the two communities. One of the most common measures of distance is the Bray Curtis Dissimilarity. Similarity can be measured as 1-BC.

$$BC_{ij} = 1 - \frac{2C_{ij}}{S_i \div S_j}$$

Where:

- i & j are the two sites,
- S_i is the total number of specimens counted on site i,
- S_j is the total number of specimens counted on site j,
- C_{ij} is the sum of only the lesser counts for each species found in both sites.

In Excel, S_i and S_j are just the grand totals for a particular community. To calculate C_{ij}, we need to find the taxa that are present in both samples and then find the minimum. We can use the following formula for a particular taxa:

=IF(AND(B2>0,B3>0),MIN(B2:B3),0)

Where B2 is the cell with the number of individuals of the taxa for one sample and B3 is the cell with the number of individuals of the taxa for the other sample. The formula first checks that the number of individuals is greater than zero for both samples. If this is true, it finds the minimum. If not, it returns a value of 0. The formula can be copied for all of the taxa and then SUM can be used to add up the values to calculate C_{ij}.

Although Bray-Curtis Dissimilarity is often used in community ecology, it is not robust to incomplete sampling of the community (all taxa are not sampled) or unbalanced sampling (all treatments are not equally sampled). An alternative is the Morista-Horn Index of Dissimilarity (1-C_H). Morista-Horn Index of Similarity is

$$C_H = \frac{2\sum_{i=1}^{D_{12}} \frac{X_i}{n}\frac{Y_i}{m}}{\sum_{i=1}^{D_1}\left(\frac{X_i}{n}\right)^2 + \sum_{i=1}^{D_2}\left(\frac{Y_i}{m}\right)^2}$$

Where

- D_1=number of species in community 1
- D_2=number of species in community 2
- D_{12}=number of species in shared in both communities
- X_i=number of individuals of species *i* in community 1
- Y_i=number of individuals of species *i* in community 2
- n=total number of individuals in community 1
- m=total number of individuals in community 2

So that X_i/n and Y_i/m are proportion of individuals of species *i* in each of the communities.

Using the grand totals for each treatment, calculate the proportion of each species in each community. Using the Excel trick that $ before a column or row prevents Excel from iterating when copying a formula makes this easy. The sums of these proportions squared are used to calculate the denominator.

To calculate the numerator, we need to know which species are present in both communities. We can use the following formula for a particular taxa:

=IF(AND(B2>0,B3>0),B2*B3,0)

Where B2 is the cell with the proportion of individuals of the taxa for one community and B3 is the cell with the proportion of individuals of the same taxa for the other community. The formula first checks that the number of individuals is greater than zero for both samples. If this is true, it finds the product. If not, it returns a value of 0. The formula can be copied for all of the taxa and then SUM can be used to add up the values to calculate numerator. The SUM should be multiplied by 2 to calculate the numerator.