

Commits

Commits: When?

- Finished something discrete
 - Got a code chunk working
 - Fixed a bug
- Safety Net
 - Working on a bug and want to try something drastic
 - Making a big change (should you branch?)
- Can you think of a simple commit message?
- “It is always a good time to commit” - Josh Granek

Good Commit Messages

- Like a comment
 - Summarize
 - Explain
 - Don't repeat the change
- What would your most important collaborators want to know about this commit in six months
- Look at some examples

commit 53499c116ede77792964f3060f95fe1bb4f55690

abundance_and_diversity_plots.Rmd: got rid of duplicate chunk labels

commit 1f75ea64fad5c3cf2321cb7c13d4bd8db5e31c10

abundance_and_diversity_plots.Rmd: cleaned up cruft and text

commit 96bb8ce654d49d81923c4859bfc54e1c585d401

abundance_and_diversity_plots.Rmd: new (based on challenge from last year)

commit 2f0742d1278a871af4d85c35875b53c78d70502f

added atacama 10pct notebooks

commit dff94e0fae45caac81083247883813b37941fedc

bit of cleanup for dada2 tutorial

commit c8790c014458d3c4c0292cc9654eeffd4c1526de

argonne_download_2019.Rmd: added md5sum for map file

commit 26835bf5288c823ad36f9d3703488701f3de1236

argonne_download_2019.Rmd: cleaned up cruft, added followup steps to download, and added documentation

Optimization

Premature Optimization

“... premature optimization is the root of all evil.”

Donald Knuth, "Structured Programming with go to Statements" 1974.

Is It Worth the Time?

HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?
(ACROSS FIVE YEARS)

		HOW OFTEN YOU DO THE TASK					
		50/DAY	5/DAY	DAILY	WEEKLY	MONTHLY	YEARLY
HOW MUCH TIME YOU SHAVE OFF	1 SECOND	1 DAY	2 HOURS	30 MINUTES	4 MINUTES	1 MINUTE	5 SECONDS
	5 SECONDS	5 DAYS	12 HOURS	2 HOURS	21 MINUTES	5 MINUTES	25 SECONDS
	30 SECONDS	4 WEEKS	3 DAYS	12 HOURS	2 HOURS	30 MINUTES	2 MINUTES
	1 MINUTE	8 WEEKS	6 DAYS	1 DAY	4 HOURS	1 HOUR	5 MINUTES
	5 MINUTES	9 MONTHS	4 WEEKS	6 DAYS	21 HOURS	5 HOURS	25 MINUTES
	30 MINUTES		6 MONTHS	5 WEEKS	5 DAYS	1 DAY	2 HOURS
	1 HOUR		10 MONTHS	2 MONTHS	10 DAYS	2 DAYS	5 HOURS
	6 HOURS				2 MONTHS	2 WEEKS	1 DAY
	1 DAY					8 WEEKS	5 DAYS

Project Organization

What	Where
Raw Data	/data
Intermediate Files	/tmp/scratch or /sharedspace/OUR_PROJECT/scratch
Shared Results	/sharedspace/OUR_PROJECT/results
Rmarkdown, etc	/home/guest/OUR_PROJECT_REPO and Github

Statistical Analysis of Amplicon Data

Part 1

February 7, 2019

Count Data

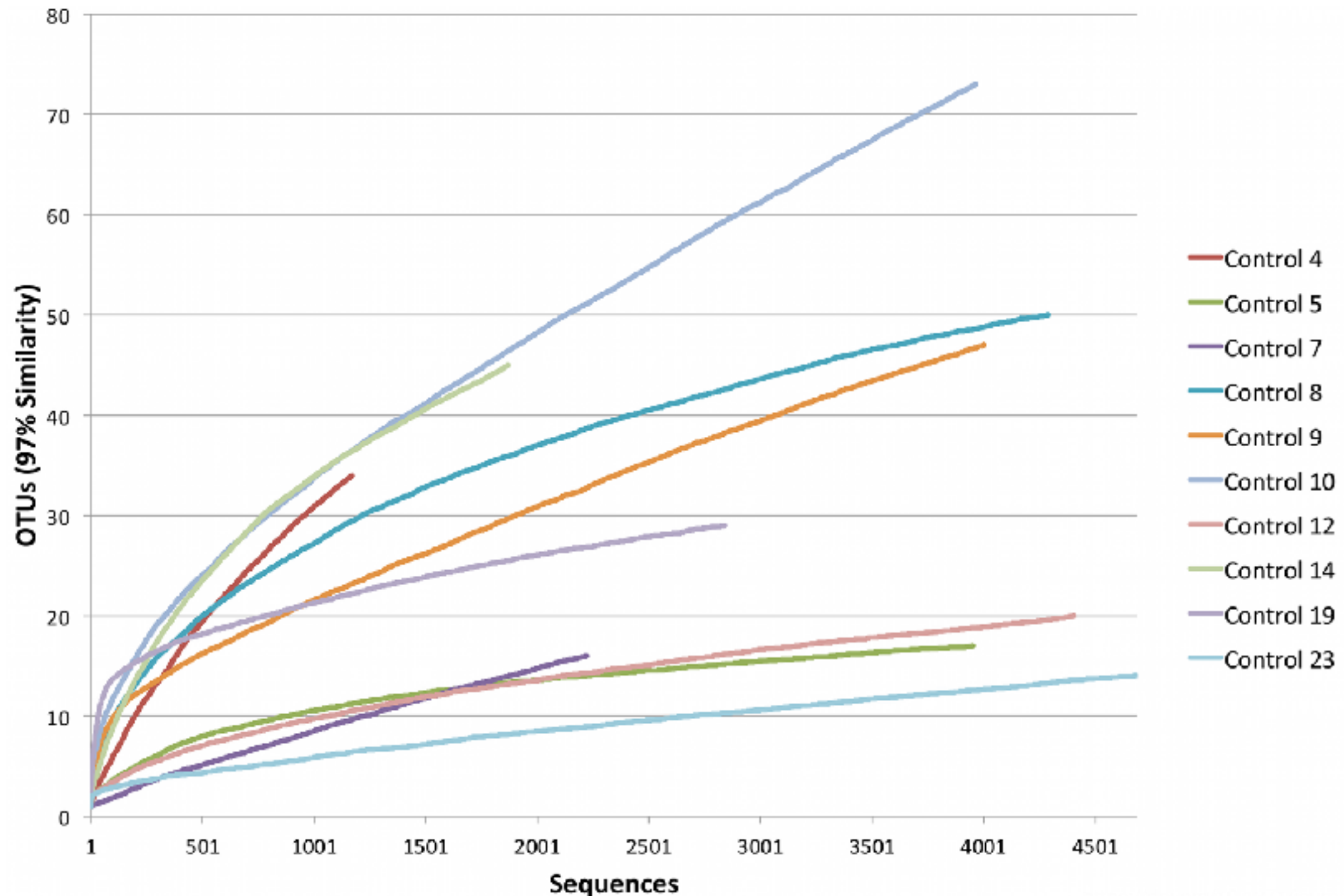
	OTU 1	OTU 2	...	Meta 1	Meta 2	...
Sample 1						
Sample 2						
...						
Sample N						

P

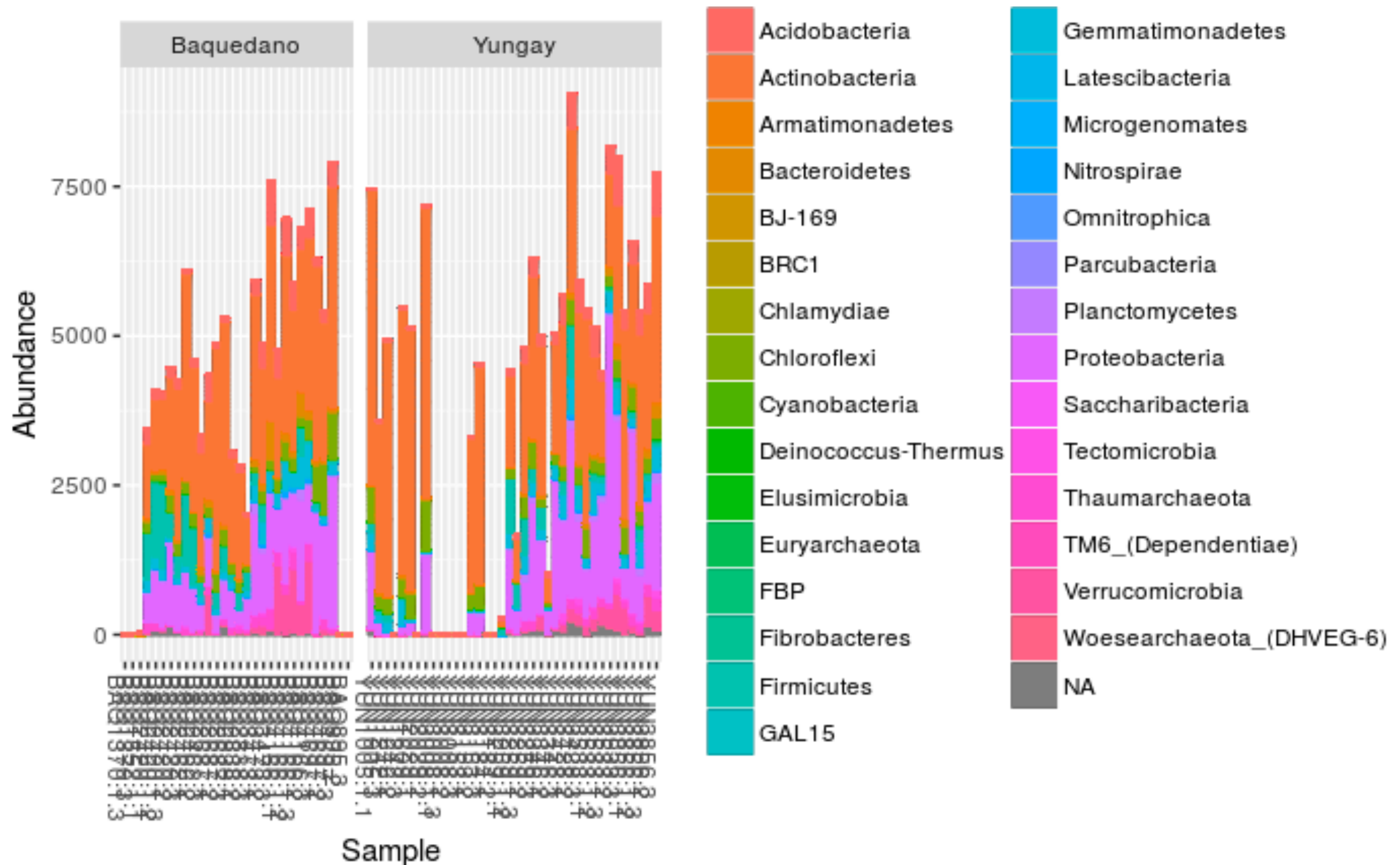
N

Exploratory Analysis and Quality Control

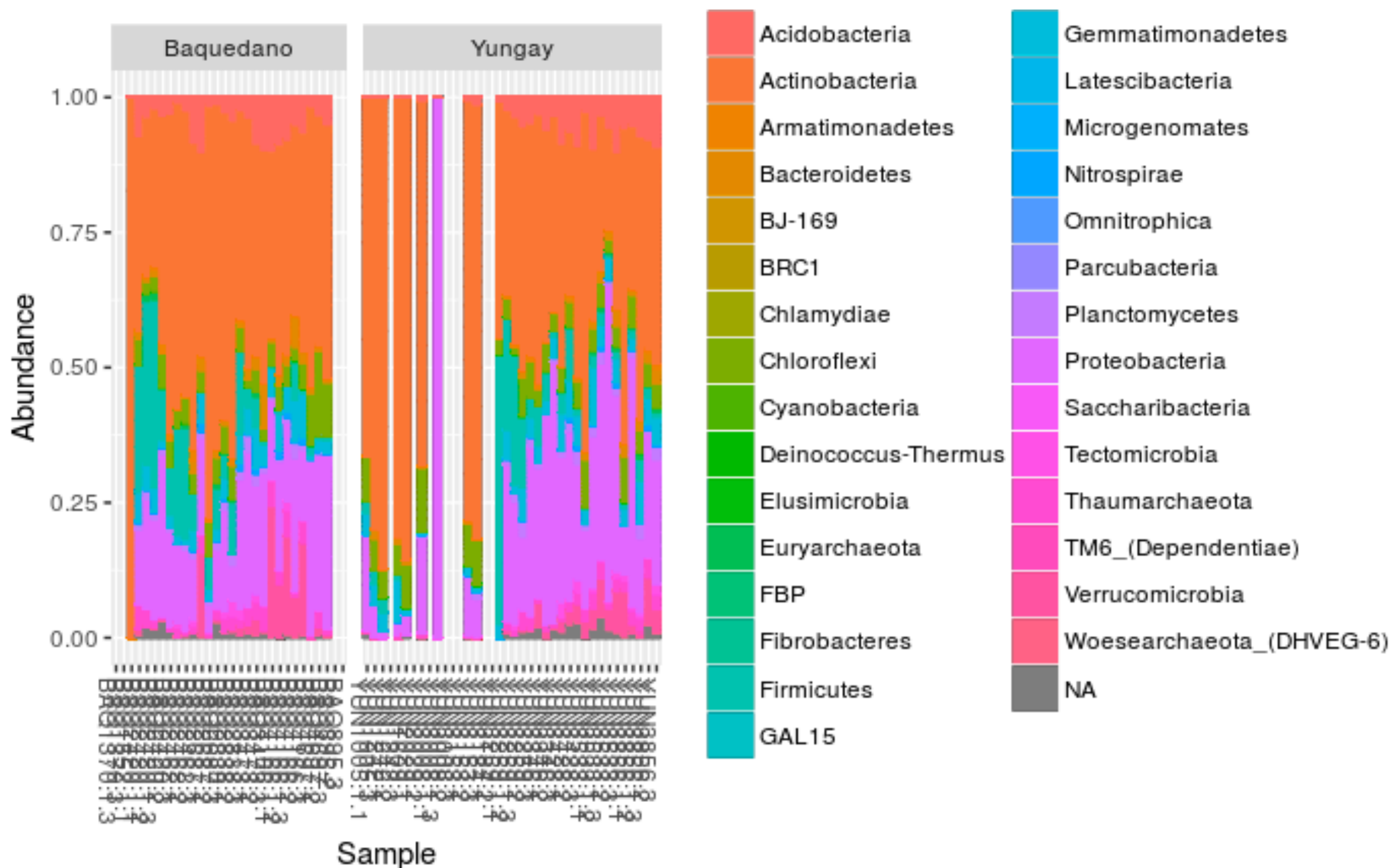
Collector's Curves (aka rarefaction)



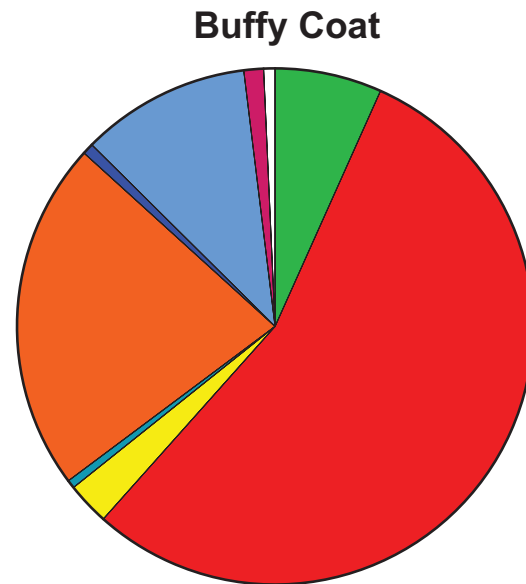
Absolute Abundance



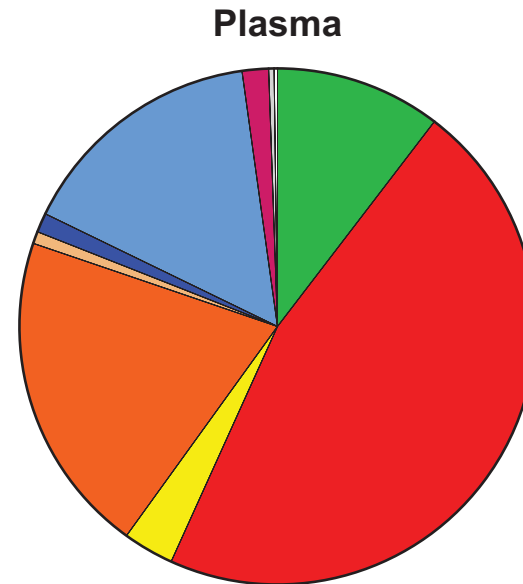
Relative Abundance



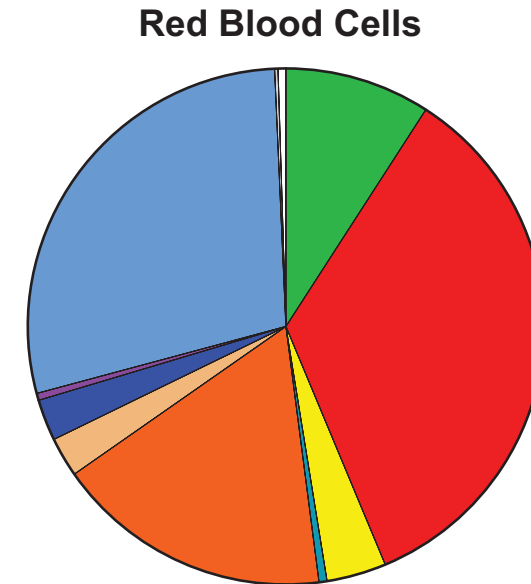
Bad Figures: Pie Charts



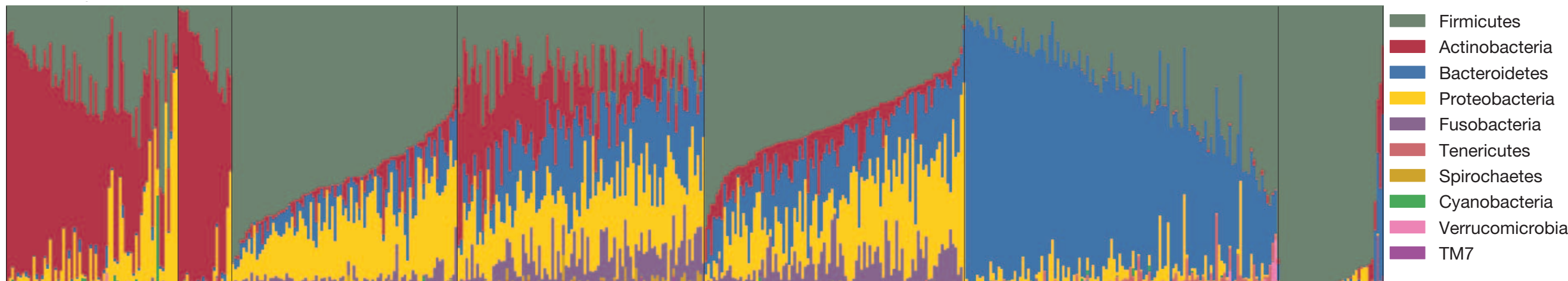
6.70% Actinobacteria
54.89% Alphaproteobacteria
2.68% Bacilli
0.52% Bacteroidia
21.94% Betaproteobacteria
0.70% Flavobacteria
10.63% Gammaproteobacteria
1.23% Sphingobacteria
0.71% Other (<0.40%)



10.40% Actinobacteria
46.36% Alphaproteobacteria
3.19% Bacilli
20.27% Betaproteobacteria
0.76% Clostridia
1.21% Flavobacteria
15.61% Gammaproteobacteria
1.65% Sphingobacteria
0.34% unclassified
0.21% Other (<0.40%)

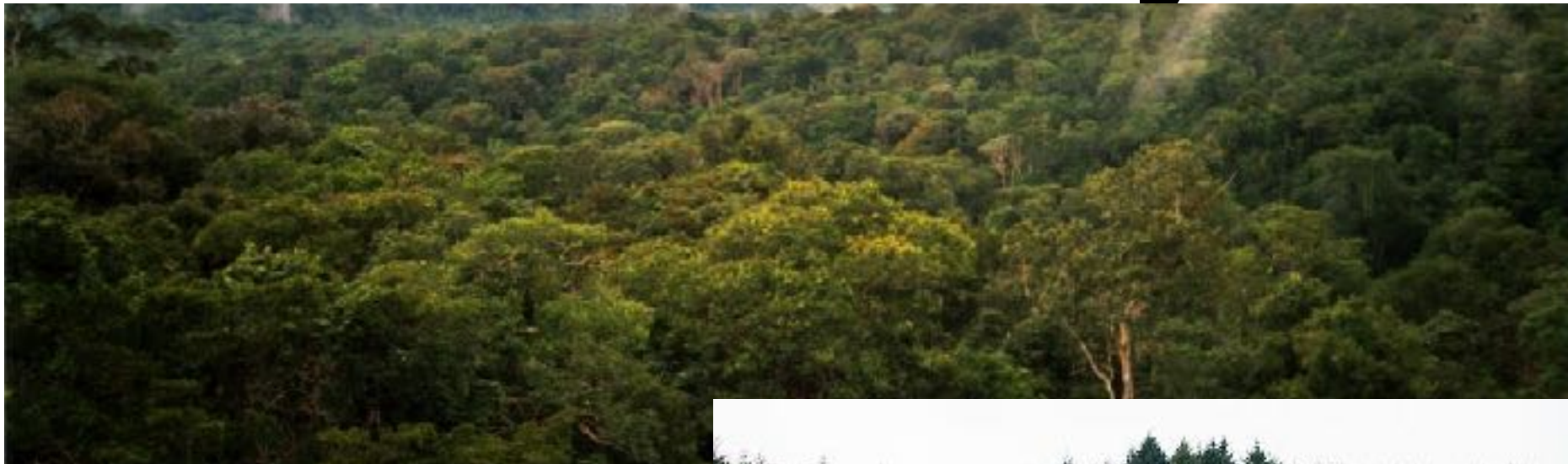


9.12% Actinobacteria
34.59% Alphaproteobacteria
3.74% Bacilli
0.50% Bacteroidia
17.35% Betaproteobacteria
2.49% Clostridia
2.59% Flavobacteria
0.44% Fusobacteria
28.48% Gammaproteobacteria
0.21% unclassified
0.49% Other (<0.40%)



Diversity

Diversity

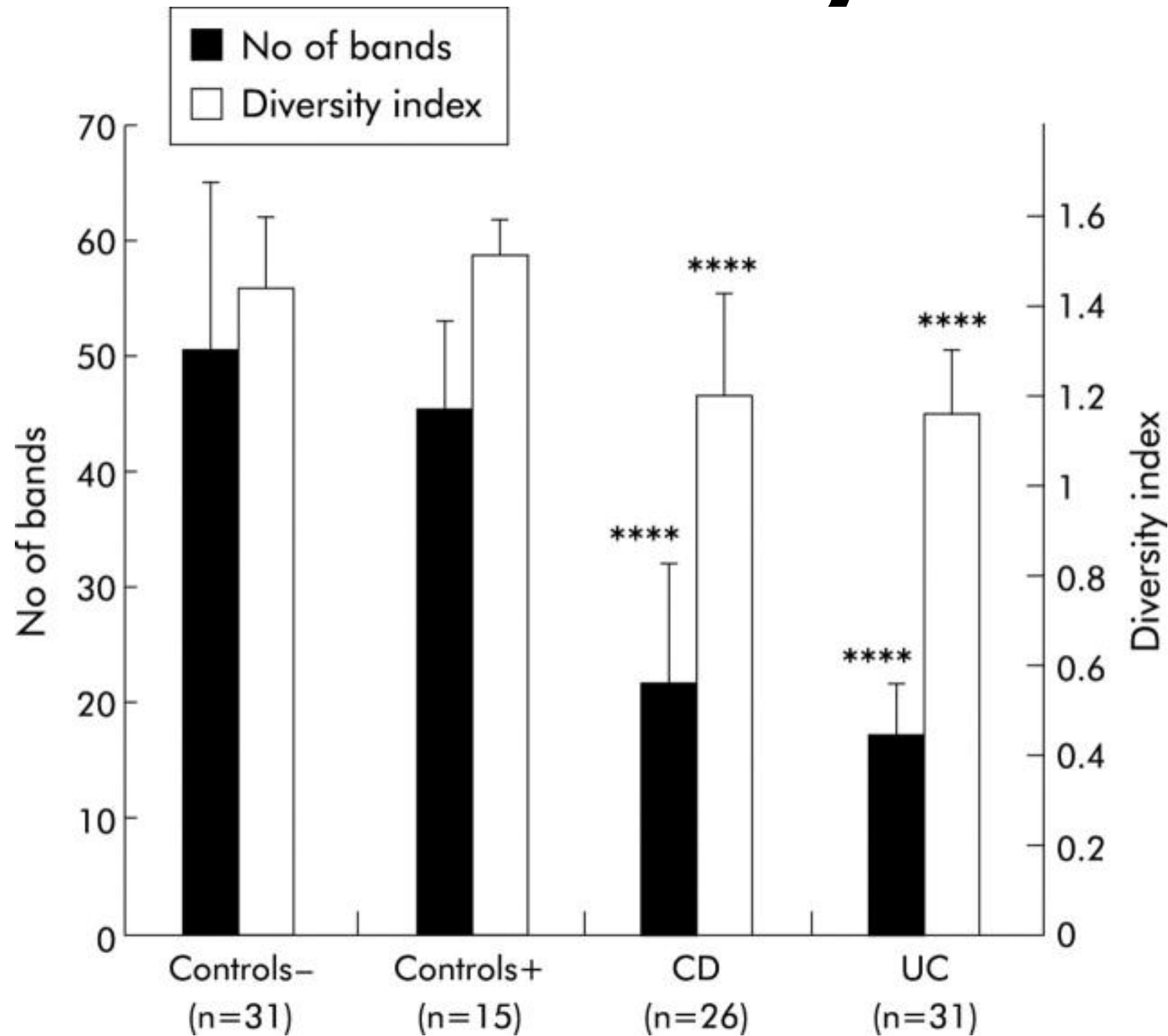


https://upload.wikimedia.org/wikipedia/commons/4/4b/Amazon_Manaus_forest.jpg

<https://en.wikipedia.org/wiki/File:Clearcutting-Oregon.jpg>

https://upload.wikimedia.org/wikipedia/commons/a/a0/Tractors_in_Potato_Field.jpg

Diversity



Alpha Diversity

- Diversity within a sample
 - Richness: number of different species
 - Evenness: distribution of species (i.e. relative abundance)

Richness



Richness: 5



Richness: 10

Evenness

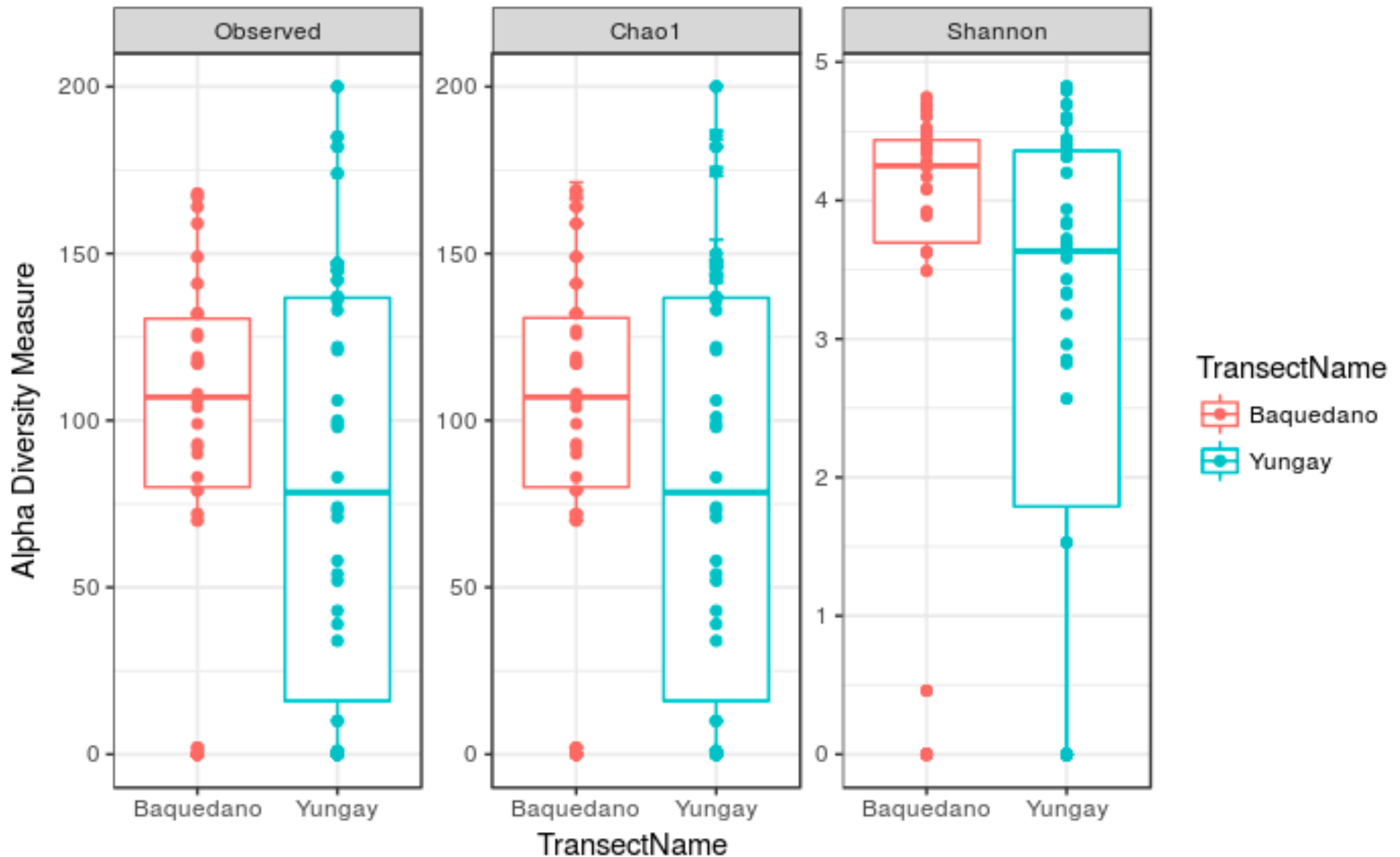


Richness: 5



Richness: 5

Alpha Diversity Metrics



Alpha Diversity Metrics

- Observed Richness

Counting

- Shannon (entropy)
- Simpson

Gambling

- Chao1
- ACE (abundance-based coverage estimators)

Weirdos

Gambling Metrics

- Jar with 8 balls
- Shannon: How much would you bet that a randomly selected ball is red?
- Simpson: How much would you bet that two randomly selected balls are the same color?

	Jar 1	Jar 2	Jar 3	Jar 4
Red	8	5	2	1
Yellow	0	1	2	2
Green	0	1	2	2
Blue	0	1	2	3
Total	8	8	8	8

Weirdo Metrics

- Chao1: How many species are present, and how many are observed only once or twice?
- ACE: How many species are present, and how many are observed less than 10 times?