

Whole Lung Tissue vs. BAL Fluid 16S Sequencing Analysis (Trimmed Dataset)

Jennifer M. Baker

August 18, 2020 (updated)

This report was created with:

R version 4.0.2 (2020-06-22)

Platform: x86_64-apple-darwin17.0 (64-bit)

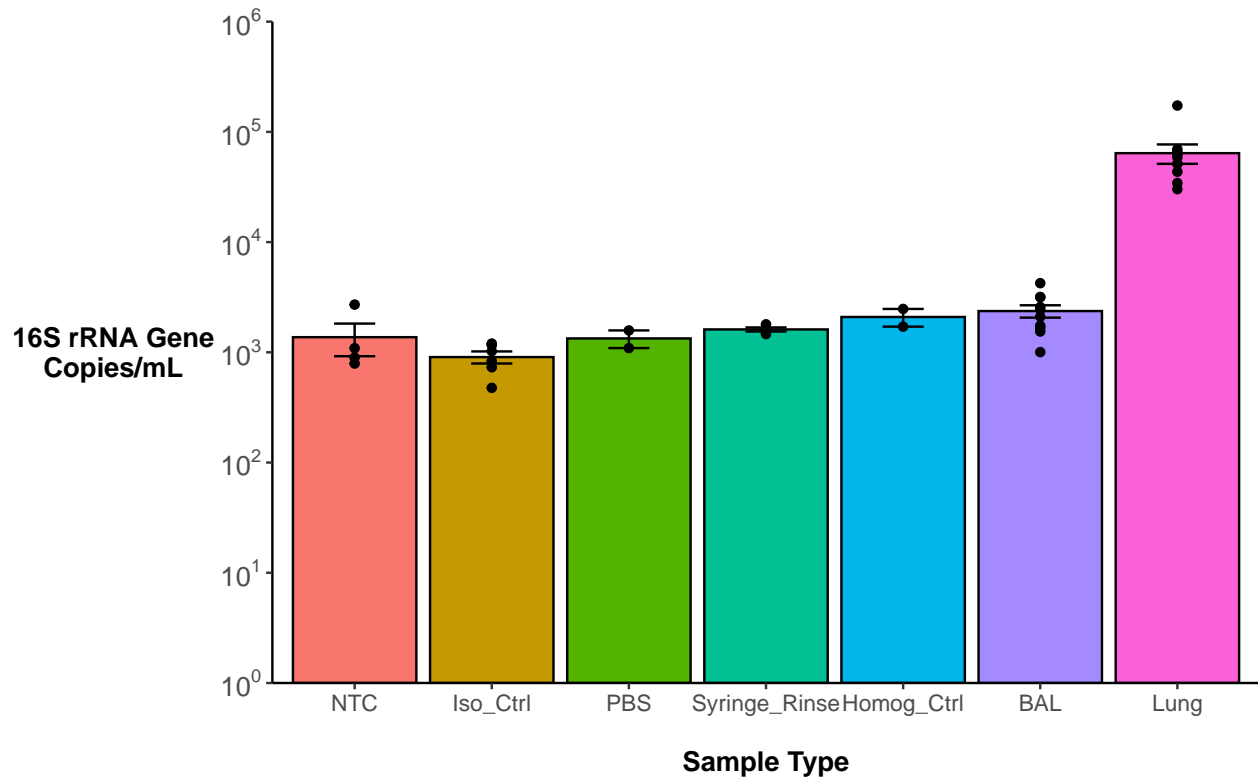
locale: en_US.UTF-8|en_US.UTF-8|en_US.UTF-8|C|en_US.UTF-8|en_US.UTF-8

attached base packages: *stats*, *graphics*, *grDevices*, *utils*, *datasets*, *methods* and *base*

other attached packages: *pander(v.0.6.3)*, *cbmbtools(v.0.0.0.9034)*, *vegan(v.2.5-6)*, *lattice(v.0.20-41)*, *permute(v.0.9-5)*, *ggpubr(v.0.4.0)*, *knitr(v.1.29)*, *scales(v.1.1.1)*, *gtools(v.3.8.2)*, *tidyselect(v.1.1.0)*, *forcats(v.0.5.0)*, *stringr(v.1.4.0)*, *dplyr(v.1.0.1)*, *purrr(v.0.3.4)*, *readr(v.1.3.1)*, *tidyr(v.1.1.1)*, *tibble(v.3.0.3)*, *ggplot2(v.3.3.2)* and *tidyverse(v.1.3.0)*

loaded via a namespace (and not attached): *httr(v.1.4.2)*, *jsonlite(v.1.7.0)*, *splines(v.4.0.2)*, *carData(v.3.0-4)*, *modelr(v.0.1.8)*, *assertthat(v.0.2.1)*, *blob(v.1.2.1)*, *cellranger(v.1.1.0)*, *yaml(v.2.2.1)*, *pillar(v.1.4.6)*, *backports(v.1.1.8)*, *glue(v.1.4.1)*, *digest(v.0.6.25)*, *ggsignif(v.0.6.0)*, *rvest(v.0.3.6)*, *colorspace(v.1.4-1)*, *Matrix(v.1.2-18)*, *htmltools(v.0.5.0)*, *pkgconfig(v.2.0.3)*, *broom(v.0.7.0)*, *haven(v.2.3.1)*, *openxlsx(v.4.1.5)*, *rio(v.0.5.16)*, *mgcv(v.1.8-31)*, *generics(v.0.0.2)*, *car(v.3.0-9)*, *ellipsis(v.0.3.1)*, *withr(v.2.2.0)*, *cli(v.2.0.2)*, *magrittr(v.1.5)*, *crayon(v.1.3.4)*, *readxl(v.1.3.1)*, *evaluate(v.0.14)*, *fs(v.1.5.0)*, *fansi(v.0.4.1)*, *nlme(v.3.1-148)*, *MASS(v.7.3-51.6)*, *rstatix(v.0.6.0)*, *xmll2(v.1.3.2)*, *foreign(v.0.8-80)*, *tools(v.4.0.2)*, *data.table(v.1.13.0)*, *hms(v.0.5.3)*, *lifecycle(v.0.2.0)*, *munsell(v.0.5.0)*, *reprex(v.0.3.0)*, *cluster(v.2.1.0)*, *zip(v.2.0.4)*, *compiler(v.4.0.2)*, *rlang(v.0.4.7)*, *grid(v.4.0.2)*, *rstudioapi(v.0.11)*, *rmarkdown(v.2.3)*, *codetools(v.0.2-16)*, *gtable(v.0.3.0)*, *abind(v.1.4-5)*, *DBI(v.1.1.0)*, *curl(v.4.3)*, *R6(v.2.4.1)*, *lubridate(v.1.7.9)*, *stringi(v.1.4.6)*, *parallel(v.4.0.2)*, *Rcpp(v.1.0.5)*, *vctrs(v.0.3.2)*, *dbplyr(v.1.4.4)* and *xfun(v.0.16)*

I. 16S rRNA gene amplicon quantification



Hypothesis testing - 16S rRNA gene amplicon quantification

Kruskal-Wallis rank sum test

data: Gene_16S_copies_per_mL by Sample_Type
 Kruskal-Wallis chi-squared = 29.276, df = 6, p-value = 5.394e-05

Pairwise comparisons using Wilcoxon rank sum exact test

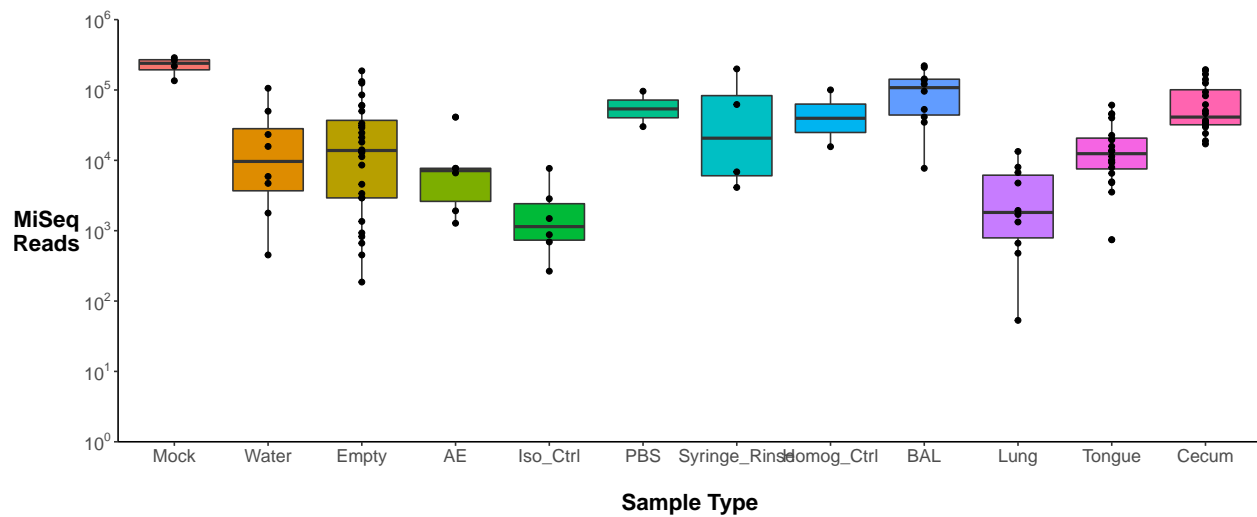
data: ddpcr\$Gene_16S_copies_per_mL and ddpcr\$Sample_Type

	NTC	Iso_Ctrl	PBS	Syringe_Rinse	Homog_Ctrl	BAL
Iso_Ctrl	0.64000	-	-	-	-	-
PBS	0.58947	0.40000	-	-	-	-
Syringe_Rinse	0.42353	0.03333	0.40000	-	-	-
Homog_Ctrl	0.58947	0.16667	0.42353	0.40000	-	-
BAL	0.22238	0.00839	0.31818	0.27082	0.90909	-
Lung	0.00839	0.00262	0.07955	0.00839	0.07955	0.00023

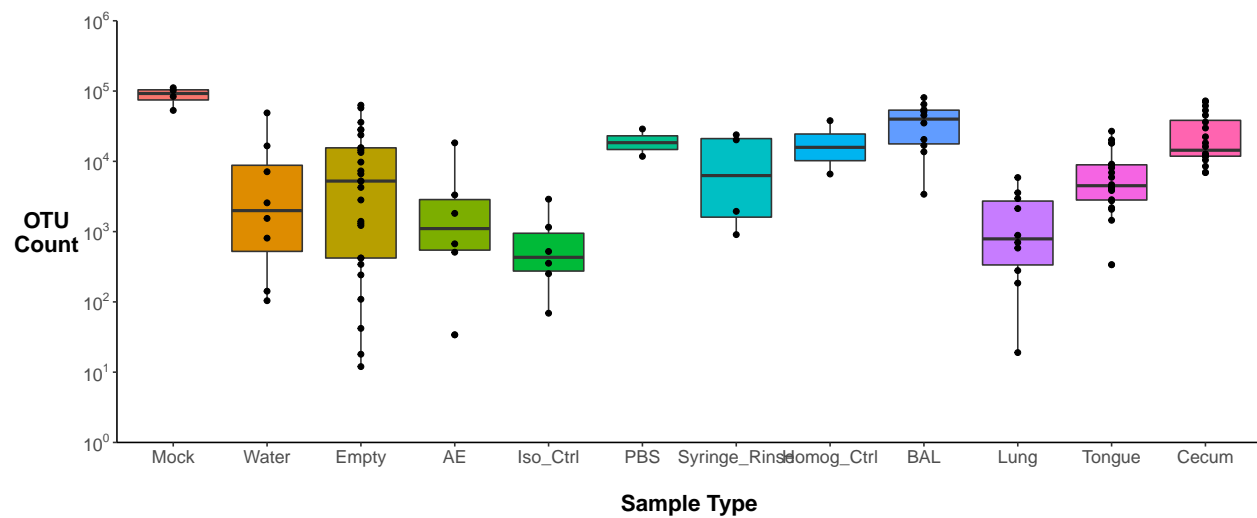
P value adjustment method: BH

II. Quality Checks - Confirmation of Sufficient Reads

2.1 MiSeq Reads by Sample Type

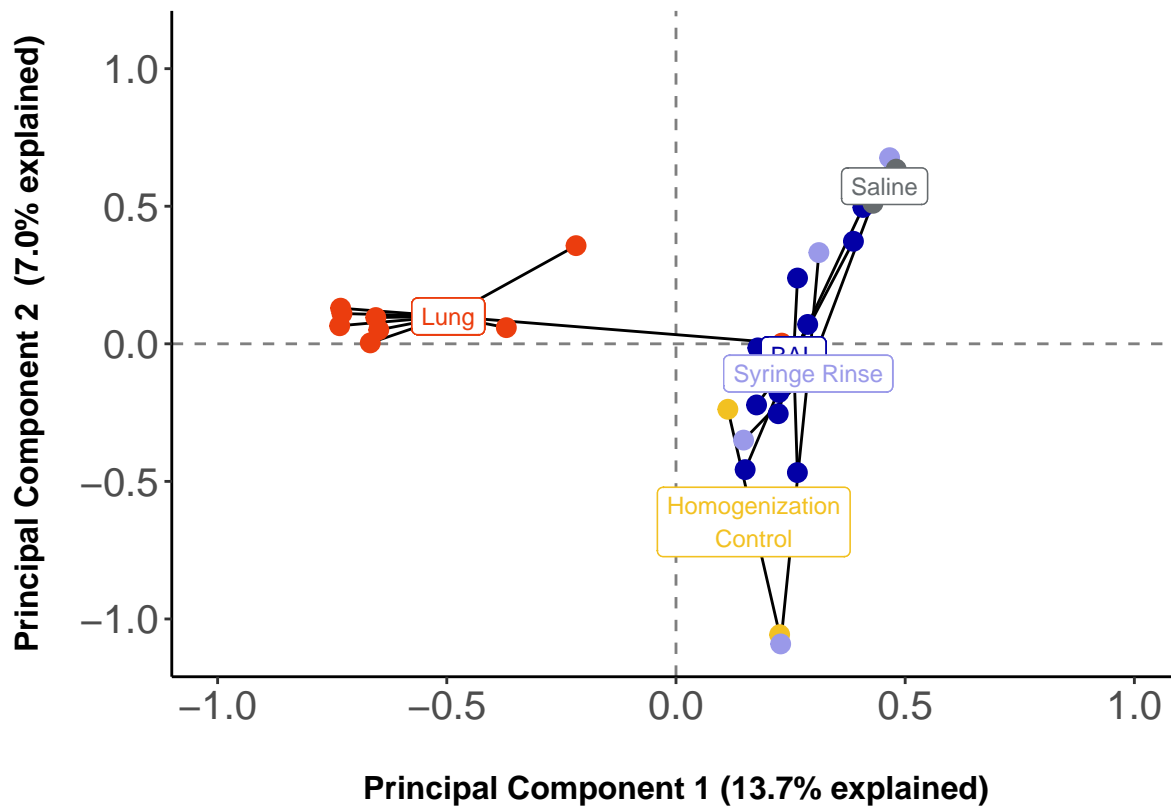


2.2 OTU Counts by Sample Type



III. Principal Component Analysis

3.1 Lung Samples vs. Sampling Controls



Call:

```
adonis(formula = adonis.hel.df.wbn ~ otu.df.lungsamp.sampctrls$RA_Groups, permutations = 10000, me
```

Permutation: free

Number of permutations: 10000

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2
otu.df.lungsamp.sampctrls\$RA_Groups	2	3.9786	1.98931	2.4485	0.16379
Residuals	25	20.3117	0.81247		0.83621
Total	27	24.2903			1.00000

Pr(>F)

```
otu.df.lungsamp.sampctrls$RA_Groups 9.999e-05 ***
```

Residuals

Total

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:

```
adonis(formula = adonis.hel.bl ~ adonis.otudf.bl$Sample_Type, permutations = 10000, method = "eucl
```

Permutation: free

Number of permutations: 10000

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
adonis.otudf.bl\$Sample_Type	1	2.4653	2.46530	3.1702	0.14975	9.999e-05 ***
Residuals	18	13.9976	0.77764		0.85025	
Total	19	16.4629			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:

adonis(formula = adonis.hel.wn ~ adonis.otudf.wn\$RA_Groups, permutations = 10000, method = "euclidean")

Permutation: free

Number of permutations: 10000

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
adonis.otudf.wn\$RA_Groups	1	2.5349	2.5349	3.403	0.17539	2e-04 ***
Residuals	16	11.9183	0.7449		0.82461	
Total	17	14.4532			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:

adonis(formula = adonis.hel.bn ~ adonis.otudf.bn\$Organ, permutations = 10000, method = "euclidean")

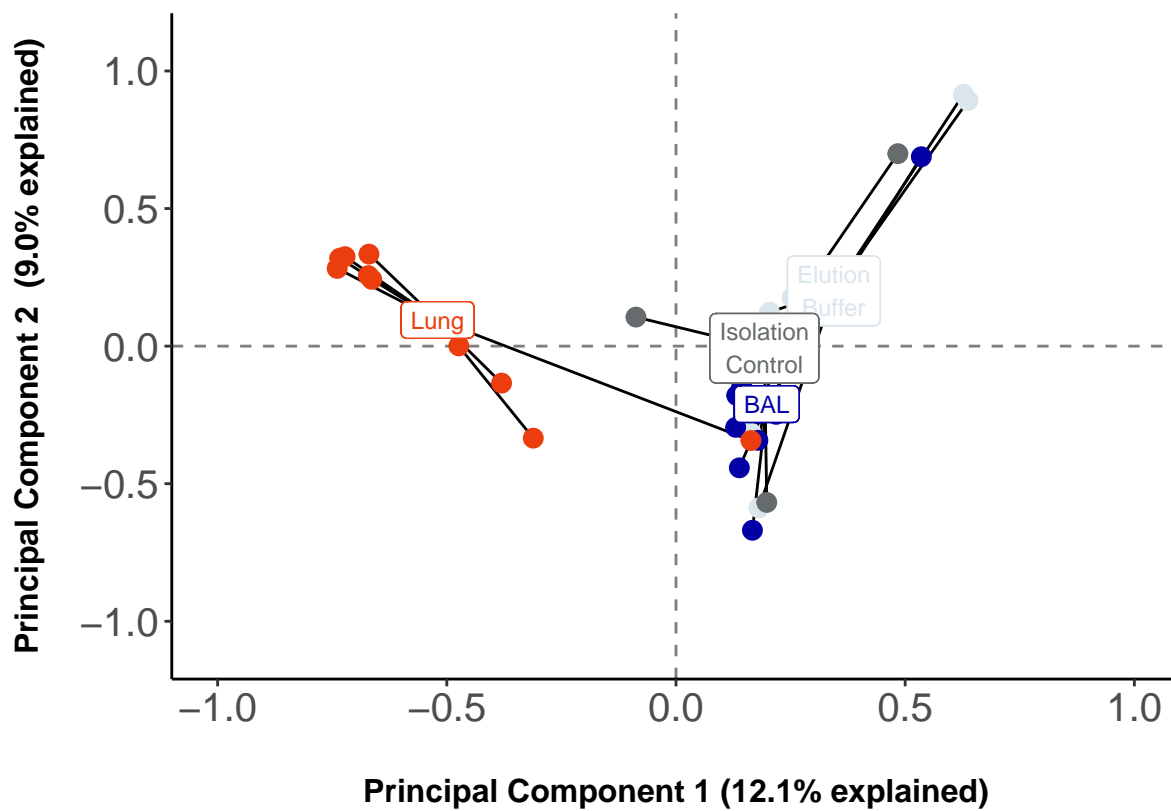
Permutation: free

Number of permutations: 10000

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
adonis.otudf.bn\$Organ	1	0.9148	0.91483	0.99522	0.05856	0.464
Residuals	16	14.7076	0.91922		0.94144	
Total	17	15.6224			1.00000	

3.2 Lung Samples v. Isolation Controls



Call:

```
adonis(formula = otu.good.lungsamp.isoctrls_hel ~ otu.df.lungsamp.isoctrls$Sample_Type,          permutation
```

Permutation: free

Number of permutations: 10000

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2
otu.df.lungsamp.isoctrls\$Sample_Type	3	5.1542	1.71806	2.0821	0.18239
Residuals	28	23.1047	0.82517		0.81761
Total	31	28.2589			1.00000

Pr(>F)

```
otu.df.lungsamp.isoctrls$Sample_Type 9.999e-05 ***
```

Residuals

Total

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:

```
adonis(formula = adonis.hel.il ~ adonis.otudf.il$RA_Groups, permutations = 10000,          method = "euclid
```

Permutation: free

Number of permutations: 10000

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
adonis.otudf.il\$RA_Groups	1	2.7451	2.74511	3.5179	0.14958	9.999e-05 ***
Residuals	20	15.6065	0.78032		0.85042	
Total	21	18.3516			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:

adonis(formula = adonis.hel.ib ~ adonis.otudf.ib\$RA_Groups, permutations = 10000, method = "euclid

Permutation: free

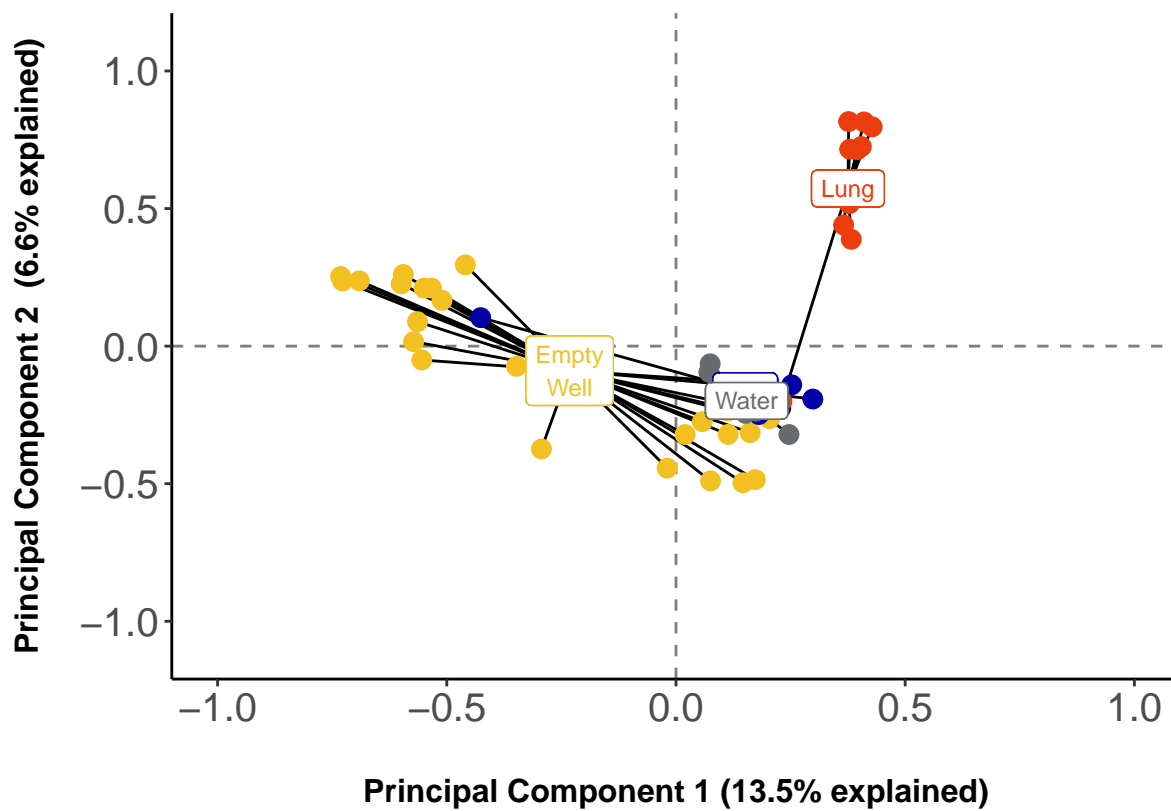
Number of permutations: 10000

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
adonis.otudf.ib\$RA_Groups	1	1.2087	1.20866	1.3141	0.06165	0.07029 .
Residuals	20	18.3957	0.91978		0.93835	
Total	21	19.6044			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

3.3 Lung Samples vs. Sequencing Controls



Call:

```
adonis(formula = otu.good.lungsamp.seqctrls_hel ~ otu.df.lungsamp.seqctrls$Sample_Type,          permutation = 10000)
```

Permutation: free

Number of permutations: 10000

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2
otu.df.lungsamp.seqctrls\$Sample_Type	3	8.141	2.71353	3.3839	0.16334
Residuals	52	41.698	0.80188		0.83666
Total	55	49.838			1.00000

Pr(>F)

```
otu.df.lungsamp.seqctrls$Sample_Type 9.999e-05 ***
```

Residuals

Total

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:

```
adonis(formula = adonis.hel.sl ~ adonis.otudf.sl$RA_Groups, permutations = 10000,          method = "euclidean")
```

Permutation: free

Number of permutations: 10000

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
adonis.otudf.sl\$RA_Groups	1	4.395	4.3950	5.4853	0.11085	9.999e-05 ***
Residuals	44	35.255	0.8012		0.88915	
Total	45	39.650			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:

adonis(formula = adonis.hel.sb ~ adonis.otudf.sb\$RA_Groups, permutations = 10000, method = "euclid

Permutation: free

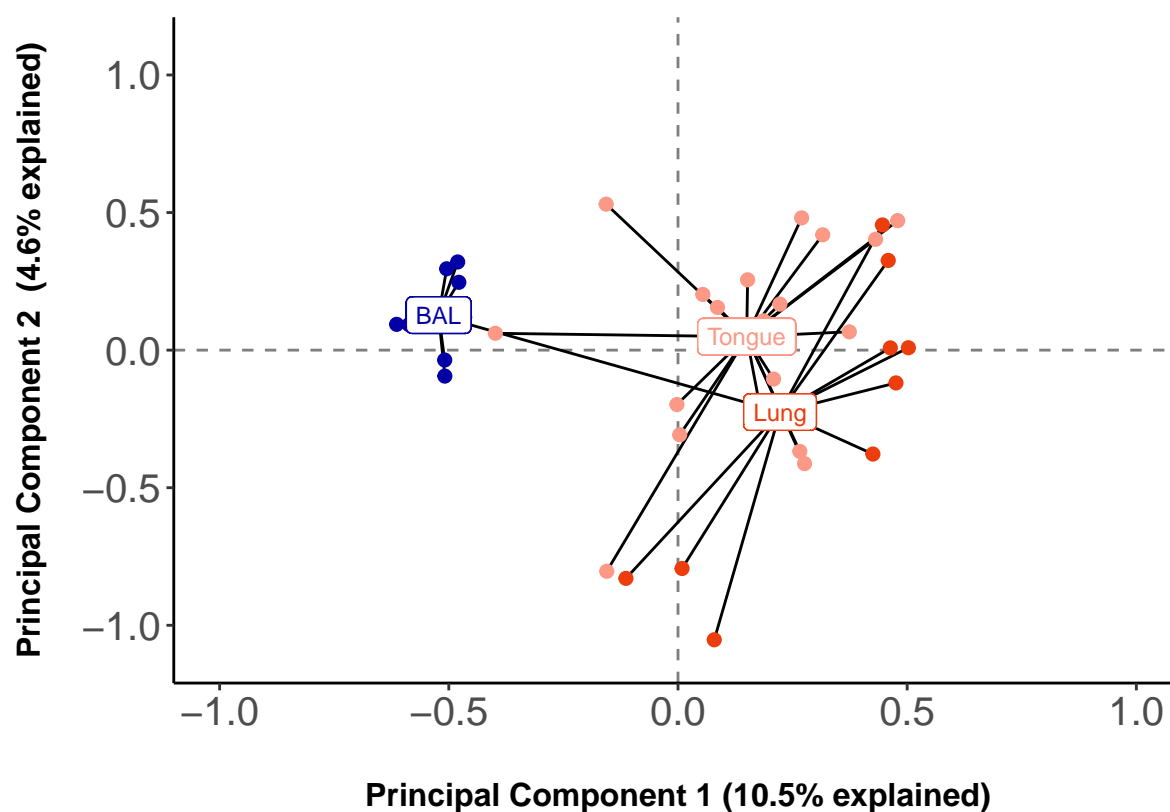
Number of permutations: 10000

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
adonis.otudf.sb\$RA_Groups	1	2.069	2.06931	2.3933	0.05159	4e-04 ***
Residuals	44	38.044	0.86463		0.94841	
Total	45	40.113			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

3.4 Lung Samples vs. Tongue



```
Call:
adonis(formula = otu.good.tong.lungsamp_hel ~ otu.df.tong.lungsamp$Sample_Type,      permutations = 10000)
```

```
Permutation: free
Number of permutations: 10000
```

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
otu.df.tong.lungsamp\$Sample_Type	2	3.981	1.99073	2.6492	0.12526	9.999e-05
Residuals	37	27.803	0.75144		0.87474	
Total	39	31.785			1.00000	

```
otu.df.tong.lungsamp$Sample_Type ***
```

```
Residuals
```

```
Total
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Call:
adonis(formula = adonis.hel.tl ~ adonis.otudf.tl$Sample_Type,      permutations = 10000, method = "eucl")
```

```
Permutation: free
Number of permutations: 10000
```

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
adonis.otudf.tl\$Sample_Type	1	1.0639	1.0639	1.5348	0.05197	0.009199 **
Residuals	28	19.4097	0.6932		0.94803	
Total	29	20.4736			1.00000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:

adonis(formula = adonis.hel.tb ~ adonis.otudf.tb\$Sample_Type, permutations = 10000, method = "eucl

Permutation: free

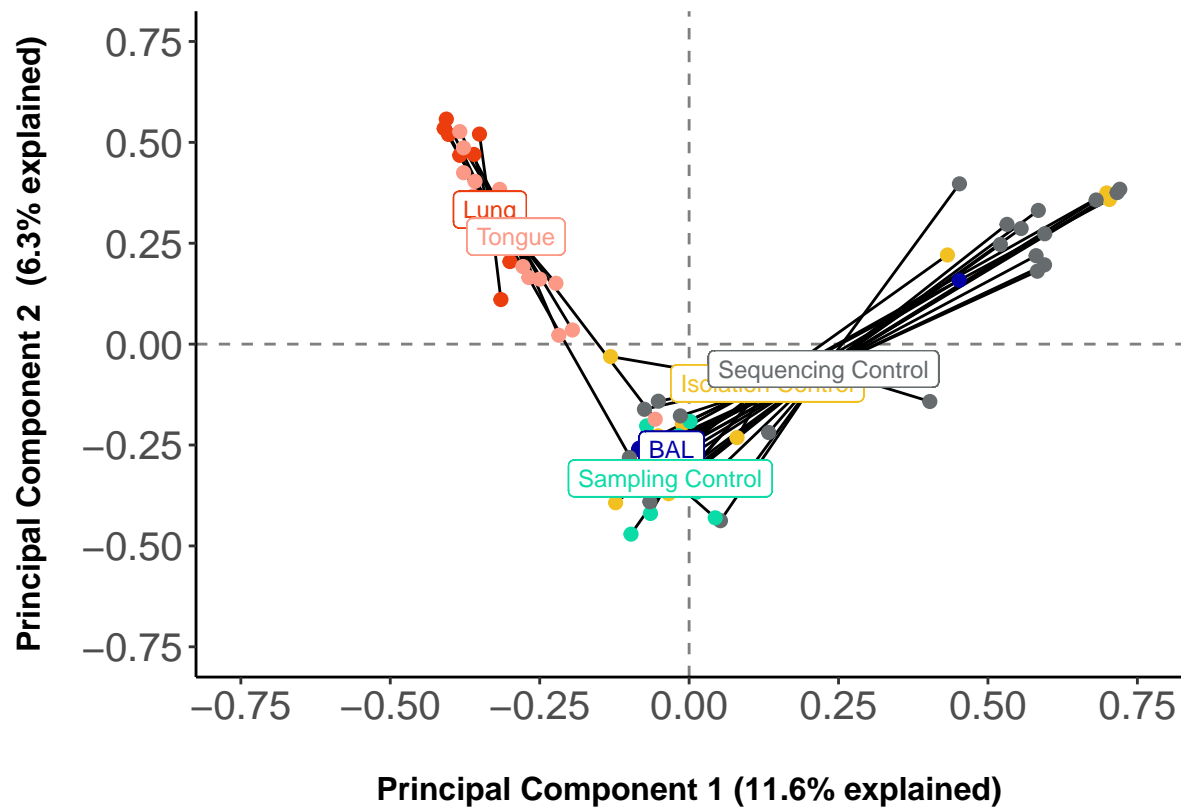
Number of permutations: 10000

Terms added sequentially (first to last)

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
adonis.otudf.tb\$Sample_Type	1	2.6012	2.60116	3.2809	0.10489	9.999e-05 ***
Residuals	28	22.1989	0.79282		0.89511	
Total	29	24.8001			1.00000	

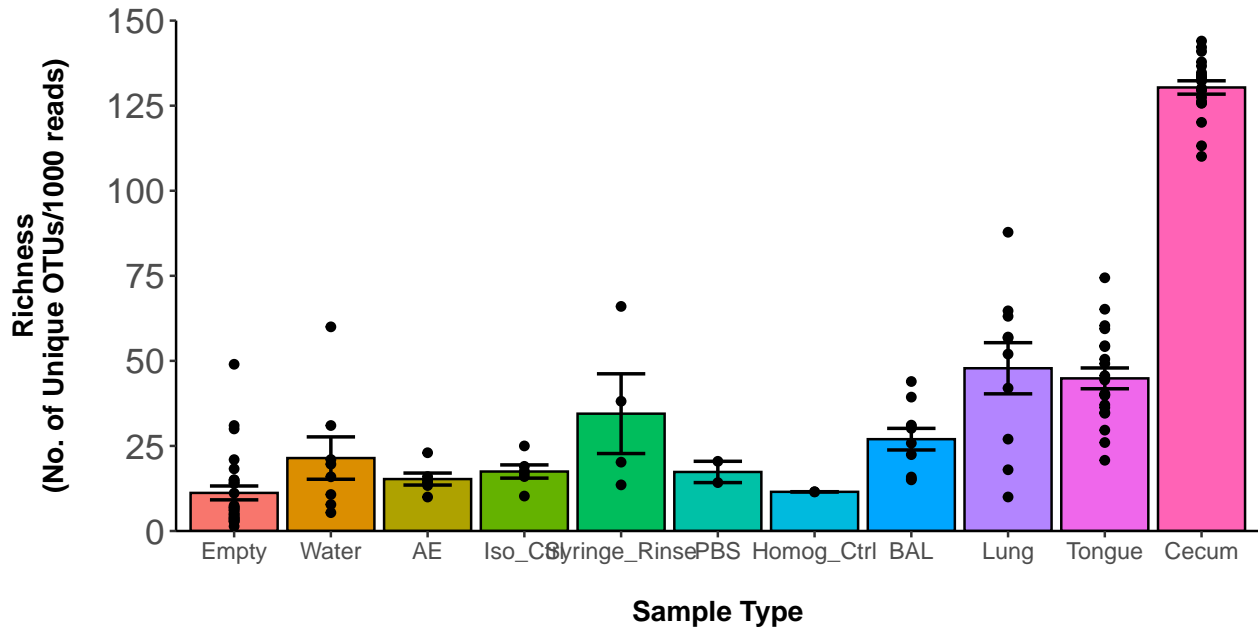
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

3.5 Lung vs. Tongue vs. Negative Controls



IV. Diversity Indices

4.1 Alpha Diversity - Rarified Richness



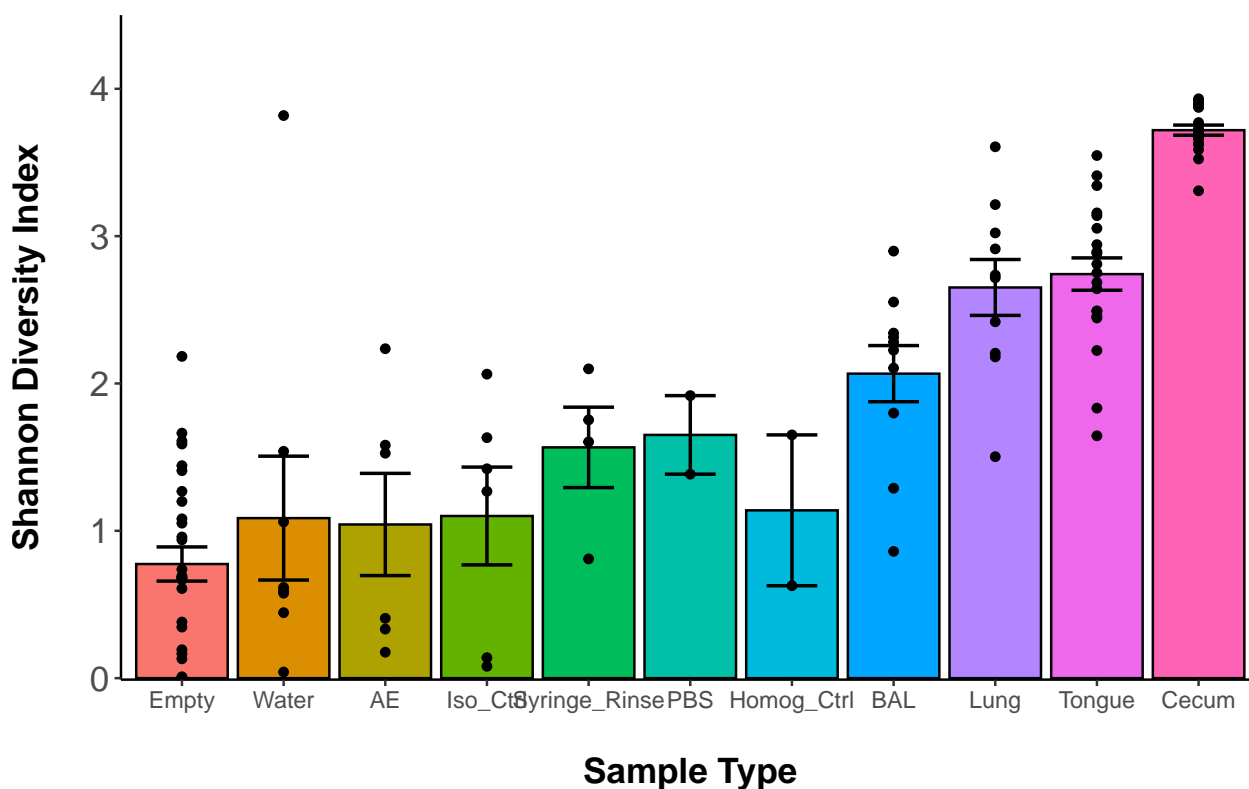
Tukey multiple comparisons of means
95% family-wise confidence level

```
Fit: aov(formula = tukey_otu_df[, "Unique_Otus_per_1k_reads"] ~ tukey_otu_df[, "Alpha_Div_Groups"])
```

```
$`tukey_otu_df[, "Alpha_Div_Groups"]`
```

	diff	lwr	upr
Isolation Control-BAL	-10.612291	-27.844847	6.6202640
Lung-BAL	20.846543	2.847725	38.8453597
Sampling Control-BAL	-2.528455	-21.619083	16.5621736
Sequencing Control-BAL	-13.515401	-27.901951	0.8711478
Lung-Isolation Control	31.458834	14.226279	48.6913895
Sampling Control-Isolation Control	8.083837	-10.286129	26.4538024
Sequencing Control-Isolation Control	-2.903110	-16.318636	10.5124164
Sampling Control-Lung	-23.374997	-42.465626	-4.2843690
Sequencing Control-Lung	-34.361944	-48.748493	-19.9753948
Sequencing Control-Sampling Control	-10.986946	-26.718045	4.7441525
	p adj		
Isolation Control-BAL	0.4262449		
Lung-BAL	0.0150895		
Sampling Control-BAL	0.9958863		
Sequencing Control-BAL	0.0757056		
Lung-Isolation Control	0.0000253		
Sampling Control-Isolation Control	0.7329503		
Sequencing Control-Isolation Control	0.9737499		
Sampling Control-Lung	0.0087438		
Sequencing Control-Lung	0.0000000		
Sequencing Control-Sampling Control	0.2986417		

4.2 Alpha Diversity - Shannon Diversity Index



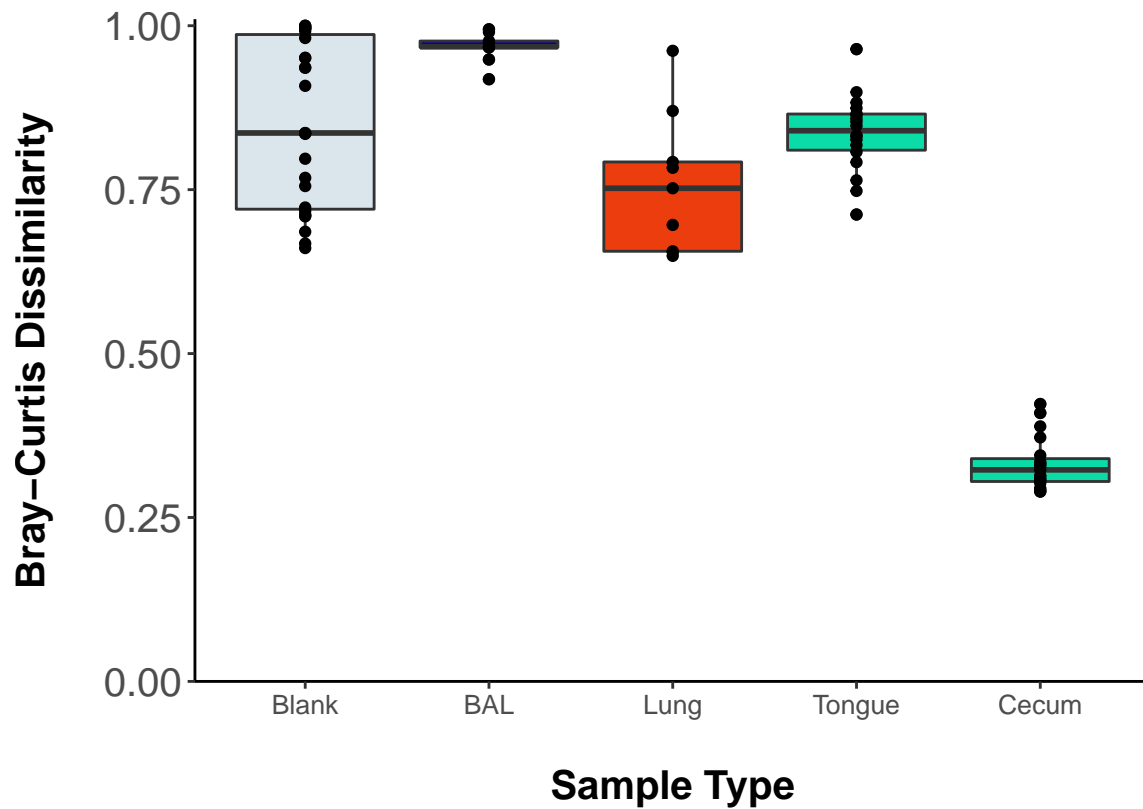
Tukey multiple comparisons of means
95% family-wise confidence level

```
Fit: aov(formula = tukey_otu_df[, "Shannon"] ~ tukey_otu_df[, "Alpha_Div_Groups"])
```

```
$`tukey_otu_df[, "Alpha_Div_Groups"]`
```

	diff	lwr	upr	p adj
Isolation Control-BAL	-0.9947478	-1.8477479	-0.1417477	0.0141500
Lung-BAL	0.5850532	-0.3058764	1.4759827	0.3602822
Sampling Control-BAL	-0.5859416	-1.5309150	0.3590319	0.4191854
Sequencing Control-BAL	-1.2226705	-1.9347952	-0.5105458	0.0000800
Lung-Isolation Control	1.5798009	0.7268008	2.4328010	0.0000190
Sampling Control-Isolation Control	0.4088062	-0.5004949	1.3181074	0.7171772
Sequencing Control-Isolation Control	-0.2279227	-0.8919824	0.4361369	0.8716711
Sampling Control-Lung	-1.1709947	-2.1159682	-0.2260213	0.0077170
Sequencing Control-Lung	-1.8077237	-2.5198483	-1.0955990	0.0000000
Sequencing Control-Sampling Control	-0.6367289	-1.4154079	0.1419501	0.1606002

4.3 Beta Diversity - Bray-Curtis dissimilarity Index within Sample Type



[1] 88 2

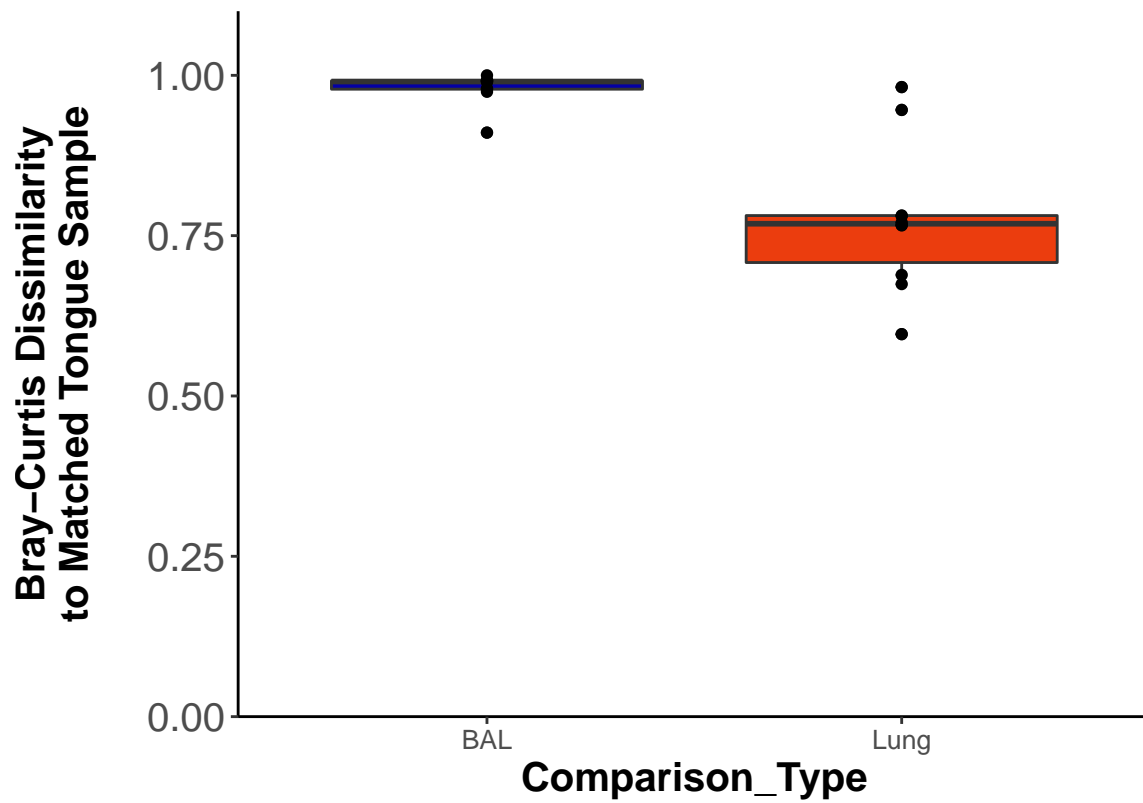
Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data: bray_dist_lbtce_long_mut_filt_uniq\$BC_Index and bray_dist_lbtce_long_mut_filt_uniq\$Sample_Type

	BAL	Blank	Cecum	Lung
Blank	0.27300	-	-	-
Cecum	< 2e-16	< 2e-16	-	-
Lung	1.3e-10	5.2e-07	< 2e-16	-
Tongue	< 2e-16	2.2e-11	< 2e-16	0.00026

P value adjustment method: BH

4.4 Bray-Curtis dissimilarity - Lung Samples vs. Matched Tongue Samples



Wilcoxon rank sum exact test

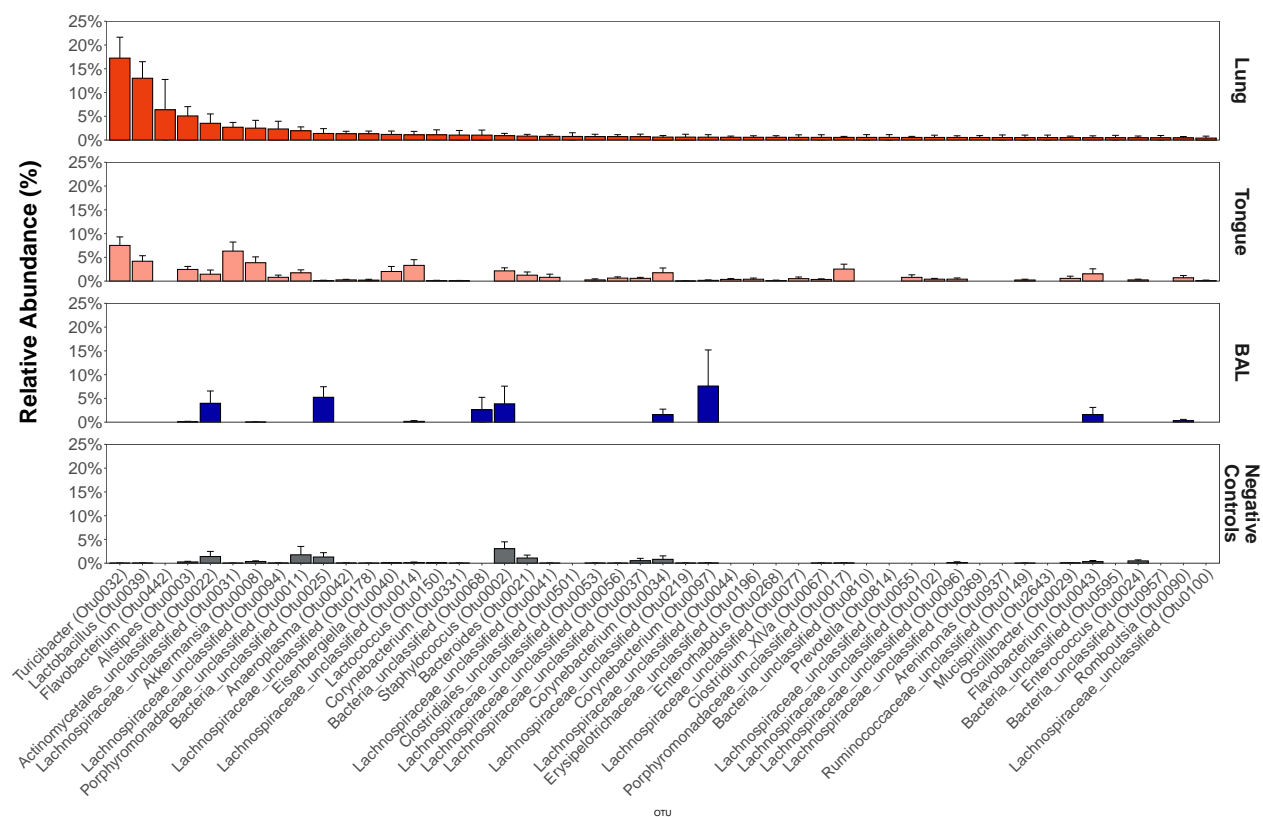
data: BC_Index by Comparison_Type

W = 95, p-value = 0.0002057

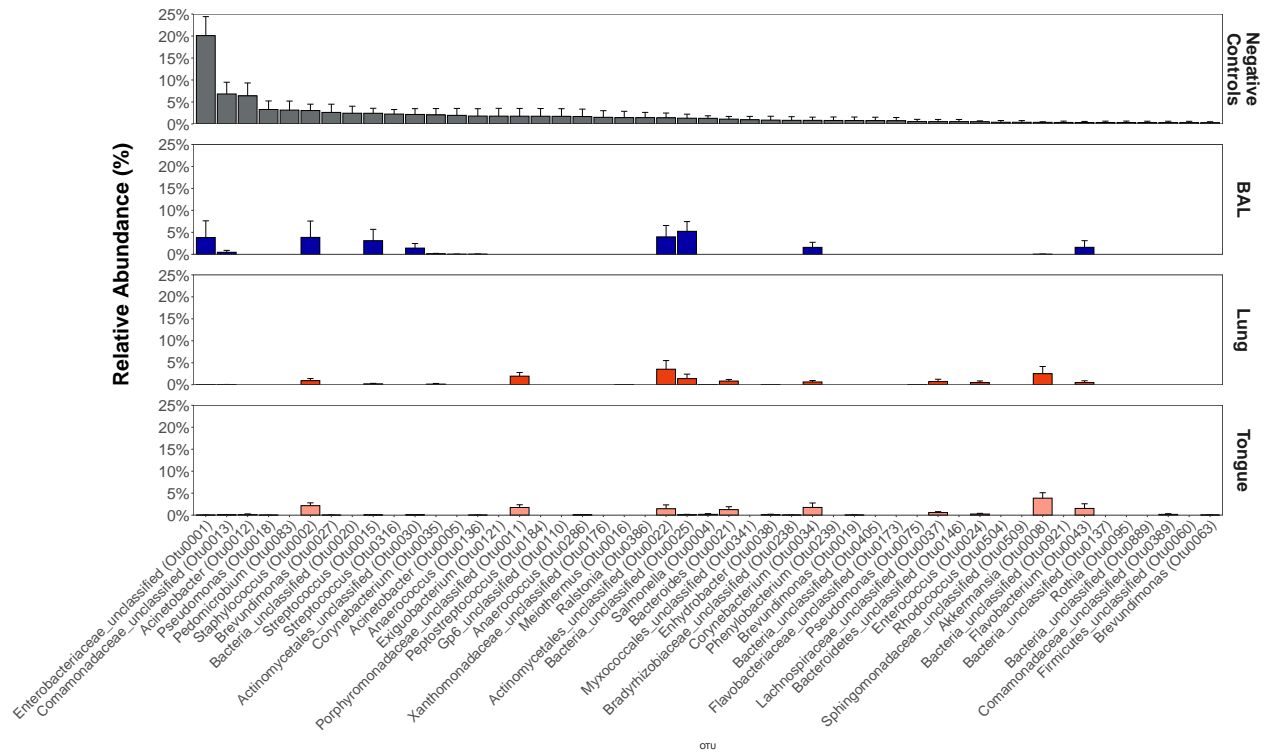
alternative hypothesis: true location shift is not equal to 0

V. Relative Abundance

5.1 Ordered by Lung OTUs



5.2 Ordered by Pooled Negative Control OTUs



5.3 Ordered by Tongue OTUs

