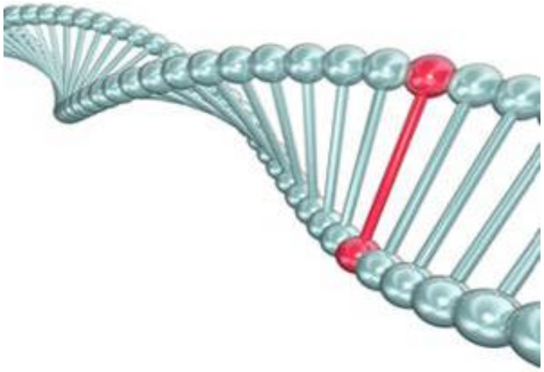
Source BioScience
Illumina Sequencing

Metagenomics Report



A:AK

Contents

Samples and groups

Table 1 below lists all samples included in the analysis as well as any sample group information provided, which can be used for filtering results by group (categorical data) or for correlations (numeric/continous data).

Sample ID	Well ID	Group (cat)	Subgroup (cat)	Time (num)
Sample1	A1	A	Type-1	0
Sample2	B1	A	Type-1	0
Sample3	C1	A	Type-1	0
Sample4	D1	A	Type-1	0
Sample5	E1	A	Type-1	0
Sample6	F1	A	Type-1	0
Sample7	G1	A	Type-1	0
Sample8	H1	A	Type-1	0
Sample9	A2	A	Type-1	0
Sample10	B2	A	Type-1	0
Sample11	C2	B	Type-1	0
Sample12	D2	B	Type-1	0
Sample13	E2	B	Type-1	0
Sample14	F2	B	Type-1	0
Sample15	G2	B	Type-1	0
Sample16	H2	B	Type-1	0
Sample17	A3	B	Type-1	0
Sample18	B3	B	Type-1	0
Sample19	C3	B	Type-1	0
Sample20	D3	B	Type-1	0
Sample21	E3	C	Type-1	0
Sample22	F3	C	Type-1	0
Sample23	G3	C	Type-1	0
Sample24	H3	C	Type-1	0
Sample25	A4	C	Type-1	30
Sample26	B4	C	Type-1	30
Sample27	C4	C	Type-1	30
Sample28	D4	C	Type-1	30
Sample29	E4	D	Type-2	30
Sample30	F4	D	Type-2	30
Sample31	G4	D	Type-2	30
Sample32	H4	D	Type-2	30
Sample33	A5	D	Type-2	30
Sample34	B5	D	Type-2	30
Sample35	C5	D	Type-2	30

Sample36	D5	E	Type-2	30
Sample37	E5	E	Type-2	30
Sample38	F5	E	Type-2	30
Sample39	G5	E	Type-2	30
Sample40	H5	E	Type-2	30
Sample41	A6	E	Type-2	30
Sample42	B6	E	Type-2	30
Sample43	C6	E	Type-2	30
Sample44	D6	E	Type-2	30
Sample45	E6	F	Type-2	30
Sample46	F6	F	Type-2	30
Sample47	G6	F	Type-2	30
Sample48	H6	F	Type-2	30
Sample49	A7	F	Type-2	60
Sample50	B7	F	Type-2	60
Sample51	C7	F	Type-2	60
Sample52	D7	F	Type-2	60
Sample53	E7	F	Type-2	60
Sample54	F7	F	Type-2	60
Sample55	G7	G	Type-2	60
Sample56	H7	G	Type-2	60
Sample57	A8	G	Type-2	60
Sample58	B8	G	Type-2	60
Sample59	C8	G	Type-2	60
Sample60	D8	G	Type-2	60
Sample61	E8	G	Type-2	60
Sample62	F8	G	Type-2	60
Sample63	G8	G	Type-2	60
Sample64	H8	G	Type-2	60
Sample65	A9	H	Type-2	60
Sample66	B9	H	Type-2	60
Sample67	C9	H	Type-2	60
Sample68	D9	H	Type-2	60
Sample69	E9	H	Type-2	60
Sample70	F9	H	Type-2	60
Sample71	G9	H	Type-2	60
Sample72	H9	H	Type-2	60
Sample73	A10	H	Type-2	90
Sample74	B10	H	Type-2	90
Sample75	C10	I	Type-2	90
Sample76	D10	I	Type-2	90

Sample77	E10	I	Type-2	90
Sample78	F10	I	Type-2	90
Sample79	G10	I	Type-2	90
Sample80	H10	I	Type-2	90
Sample81	A11	I	Type-2	90
Sample82	B11	I	Type-2	90
Sample83	C11	I	Type-2	90
Sample84	D11	I	Type-2	90
Sample85	E11	I	Type-2	90
Sample86	F11	J	Type-2	90
Sample87	G11	J	Type-2	90
Sample88	H11	J	Type-2	90
Sample89	A12	J	Type-2	90
Sample90	B12	J	Type-2	90
Sample91	C12	J	Type-2	90
Sample92	D12	J	Type-2	90
Sample93	E12	J	Type-2	90
Sample94	F12	J	Type-2	90
Sample95	G12	J	Type-2	90
Sample96	H12	J	Type-2	90

Table 1: Sample IDs used in this analysis and any associated group information. Any group information provided can be used to filter results for browser visualisation.

Advanced Bioinformatics Workflow

Ribosomal RNA from the 16S v3-v4 regions was amplified and processed using a customized bioinformatics pipeline based on the QIIME2 package. The analysis was composed of 4 key steps, with output files provided at each step (see below for details) and the key results summarised in this report.

1. Quality Filtering, Denoising and Merging of Reads
2. Taxonomic Assignment
3. Alpha and Beta Diversity
4. Differential Abundance

1. Quality Filtering, Denoising and Merging of Raw Reads

Single end reads were imported and primer-trimmed (GGACTACNNGGGTATCTAAT) using the *q2-cutadapt* tool. Reads were then quality filtered (p-trim-left: 0, p-trunc-len: 231), denoised and chimeras removed, using the *q2-DADA2* tool.

Output files:

1-denoising/stats/primer-trimmed-samples: various metrics tables, plots and an interactive quality plot providing detailed information about sample quality and quantity (sequence numbers)
1-denoising/stats/denoised-samples: metadata.tsv table summarising sequence numbers for each primer-trimmed sample following quality filtering, denoising, merging and chimera removal
1-denoising/rep-seqs: a list of representative sequences for each amplicon sequence variant (ASV) and stats
1-denoising/table: a feature table indicating how many reads of each ASV were observed per sample and various stats and plots

A summary of input and output read numbers for each sample and step is shown in table 2 below:

SampleID	Input	Filtered	Denoised	Non-chimeric
Sample1	10,581	9,731	9,674	9,205
Sample2	10,461	8,889	8,581	8,547
Sample3	17,961	15,467	15,099	15,099
Sample4	13,326	11,590	11,468	11,461
Sample5	18,590	15,497	15,343	13,367
Sample6	24,713	20,031	19,527	19,167
Sample7	7,377	4,236	3,985	3,389
Sample8	19,198	15,817	15,196	14,773
Sample9	19,208	12,846	11,864	11,294
Sample10	13,789	11,407	11,259	10,761
Sample11	7,344	5,982	5,745	5,713
Sample12	16,224	13,801	13,577	13,564
Sample13	16,833	13,405	13,284	13,217
Sample14	12,091	9,502	9,430	9,391
Sample15	17,902	7,685	6,488	6,184
Sample16	9,497	8,249	8,078	8,007
Sample17	9,859	8,868	8,777	8,777
Sample18	12,460	10,544	10,496	10,496
Sample19	6,972	6,038	5,769	5,769
Sample20	6,859	4,777	4,446	4,310
Sample21	4,141	3,879	3,806	3,743
Sample22	6,259	5,831	5,666	5,640
Sample23	7,418	6,737	6,640	6,328
Sample24	10,597	10,138	10,046	9,776
Sample25	9,173	8,563	8,434	8,434

Sample26	13,497	12,969	12,827	12,811
Sample27	5,267	4,976	4,861	4,753
Sample28	21,534	20,466	20,243	20,234
Sample29	9,287	8,671	8,602	8,564
Sample30	12,344	11,080	10,883	10,854
Sample31	15,079	8,523	8,153	8,137
Sample32	13,852	11,773	11,523	11,497
Sample33	11,040	10,012	9,867	9,846
Sample34	12,067	11,350	11,190	11,162
Sample35	9,471	8,660	8,499	8,485
Sample36	10,157	8,929	8,817	8,721
Sample37	17,069	15,839	15,706	15,323
Sample38	9,062	8,074	7,959	7,941
Sample39	19,583	17,757	17,583	17,559
Sample40	21,995	18,884	18,705	18,629
Sample41	19,881	17,247	17,031	16,435
Sample42	18,829	16,782	16,615	16,615
Sample43	18,853	8,686	8,490	8,350
Sample44	15,596	13,195	12,896	12,878
Sample45	7,179	6,298	6,146	6,146
Sample46	6,131	5,193	5,139	5,135
Sample47	10,572	9,324	9,000	9,000
Sample48	24,873	12,387	12,114	12,037
Sample49	10,218	8,893	8,708	8,702
Sample50	6,378	5,087	4,980	4,980
Sample51	6,147	5,145	4,959	4,947
Sample52	7,679	5,827	5,713	5,713
Sample53	13,213	12,357	12,173	11,992
Sample54	14,643	12,980	12,638	12,386
Sample55	23,125	21,787	21,570	21,384
Sample56	20,063	19,089	18,910	18,792
Sample57	14,522	12,985	12,798	12,744
Sample58	18,722	17,630	17,493	17,345
Sample59	17,936	16,810	16,614	16,573
Sample60	16,269	15,253	15,098	14,905
Sample61	22,092	19,940	19,663	19,353
Sample62	19,086	17,904	17,592	16,982
Sample63	16,958	15,978	15,739	15,474
Sample64	23,409	21,426	21,213	21,142
Sample65	16,536	14,790	14,343	14,078
Sample66	13,582	12,548	12,290	12,159

Sample67	15,026	13,931	13,749	13,526
Sample68	7,493	6,604	6,465	6,403
Sample69	13,561	10,954	10,567	10,444
Sample70	13,135	11,476	11,312	11,204
Sample71	14,725	13,543	13,309	13,291
Sample72	18,505	17,668	17,454	17,229
Sample73	31,282	28,035	27,560	26,576
Sample74	36,660	34,033	32,452	30,026
Sample75	6,634	6,162	6,112	5,860
Sample76	10,840	10,150	10,061	9,446
Sample77	2,744	2,277	2,218	2,218
Sample78	5,087	4,626	4,561	4,331
Sample79	9,568	8,810	8,756	8,358
Sample80	8,944	8,308	8,221	7,598
Sample81	13,438	12,381	12,323	11,827
Sample82	7,026	6,437	6,389	5,917
Sample83	6,093	5,554	5,475	5,264
Sample84	10,751	10,129	9,990	9,880
Sample85	8,481	8,060	7,886	7,837
Sample86	6,107	5,688	5,663	5,432
Sample87	4,141	3,531	3,488	3,233
Sample88	14,097	12,621	12,493	11,230
Sample89	12,038	11,195	11,134	10,273
Sample90	2,556	2,006	1,948	1,939
Sample91	10,259	9,030	8,934	8,489
Sample92	10,594	9,820	9,762	9,157
Sample93	13,656	12,130	12,045	11,563
Sample94	4,881	4,486	4,456	4,296
Sample95	3,745	3,368	3,350	3,304
Sample96	8,079	6,562	6,538	6,362

Table 2: Sequence numbers of primer-trimmed samples (Input) for each sample and following QC-filtering, denoising and removal of chimeric sequences. Source: 1-denoised/stats/denoised-samples/metadata.tsv

2. Taxonomic assignment

For taxonomic assignment, amplicon sequence variants (ASVs) (also known as Operational Taxonomic Units - OTUs) were compared to a reference database of known organisms at up to 7 levels:

Level 1 (D0) = Kingdom
Level 2 (D1) = Phylum
Level 3 (D2) = Class
Level 4 (D3) = Order
Level 5 (D4) = Family
Level 6 (D5) = Genus
Level 7 (D6) = Species

Taxonomy assignment was performed using a Naïve Bayes classifier that was pre-trained on the 16S rRNA SILVA (97% OTU) reference database (version 132), using the *q2-classify-skylearn* plugin.

Output files:

2-taxonomy/classified-rep-seqs: a list of representative sequences classified against the reference database
2-taxonomy/barplots: tables and interactive barplots of classified representative sequence results for each of the 7 taxonomic levels

Overall taxonomic assignment across all samples is summarised in table 4 below and top assignments at each level summarised in tables 4-10 and figures 1-7 below:

Level	Reads	%
Unassigned	5,384	0.52
1 (D0)	50,987	99.48
2 (D1)	2,778	94.53
3 (D2)	2,128	94.26
4 (D3)	11,272	94.05
5 (D4)	69,059	92.95
6 (D5)	410,933	86.25
7 (D6)	477,147	46.34
Total	1,029,688	

Table 3: Number of reads and percentage of total reads assigned to each taxonomic level across all samples. Source: 2-taxonomy-barplots/level-7.csv.

Level 1: Kingdom

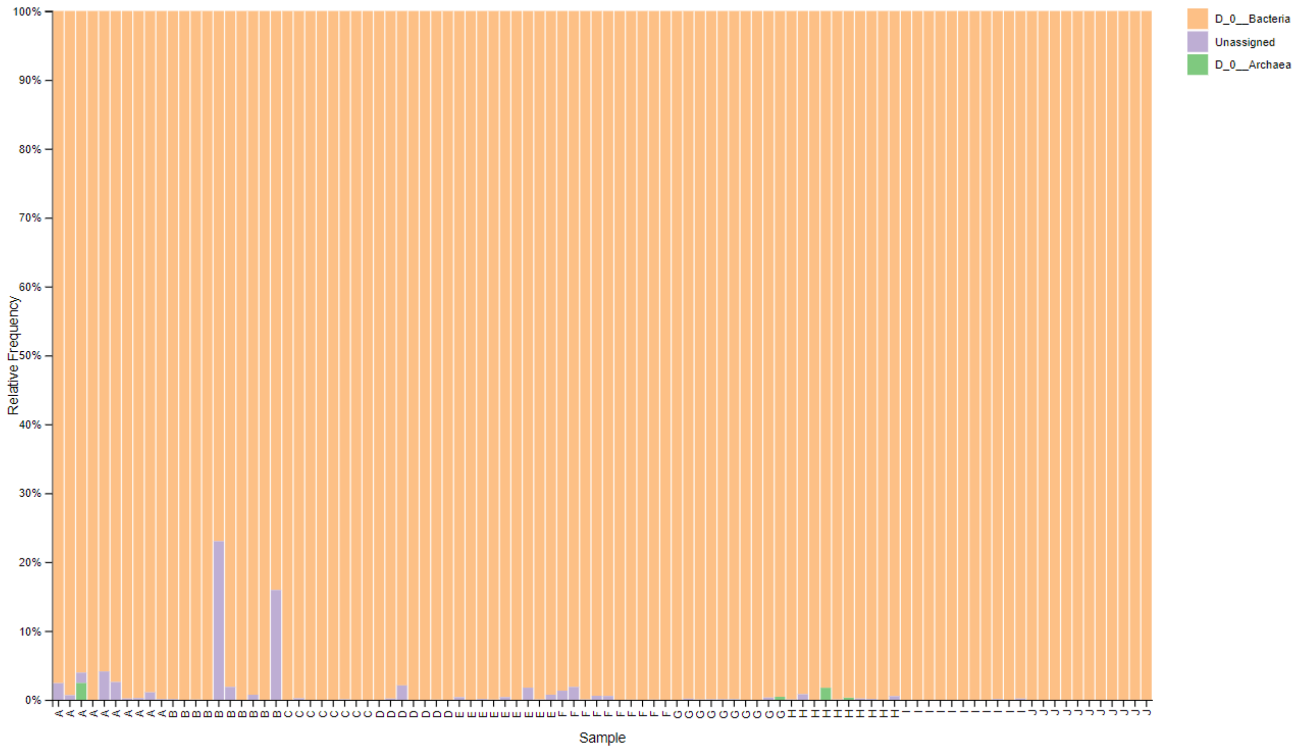


Figure 1: Barplots of taxonomic abundance for each sample at level 1 (kingdom). Source: 2-taxonomy-barplots/index.html

Group	A	B	C	D	E	F	G	H	I	J
D_0__Bacteria	98.35	97.33	99.97	99.72	99.56	99.66	99.87	99.48	99.97	100.00
Unassigned	1.46	2.67	0.03	0.28	0.44	0.34	0.09	0.13	0.03	0.00
D_0__Archaea	0.19	0.00	0.00	0.00	0.00	0.00	0.05	0.39	0.00	0.00

Table 4: Percentage taxonomic abundance for each group at level 1 (kingdom). Source: 2-taxonomy-barplots/level-1.csv.

Level 2: Phylum

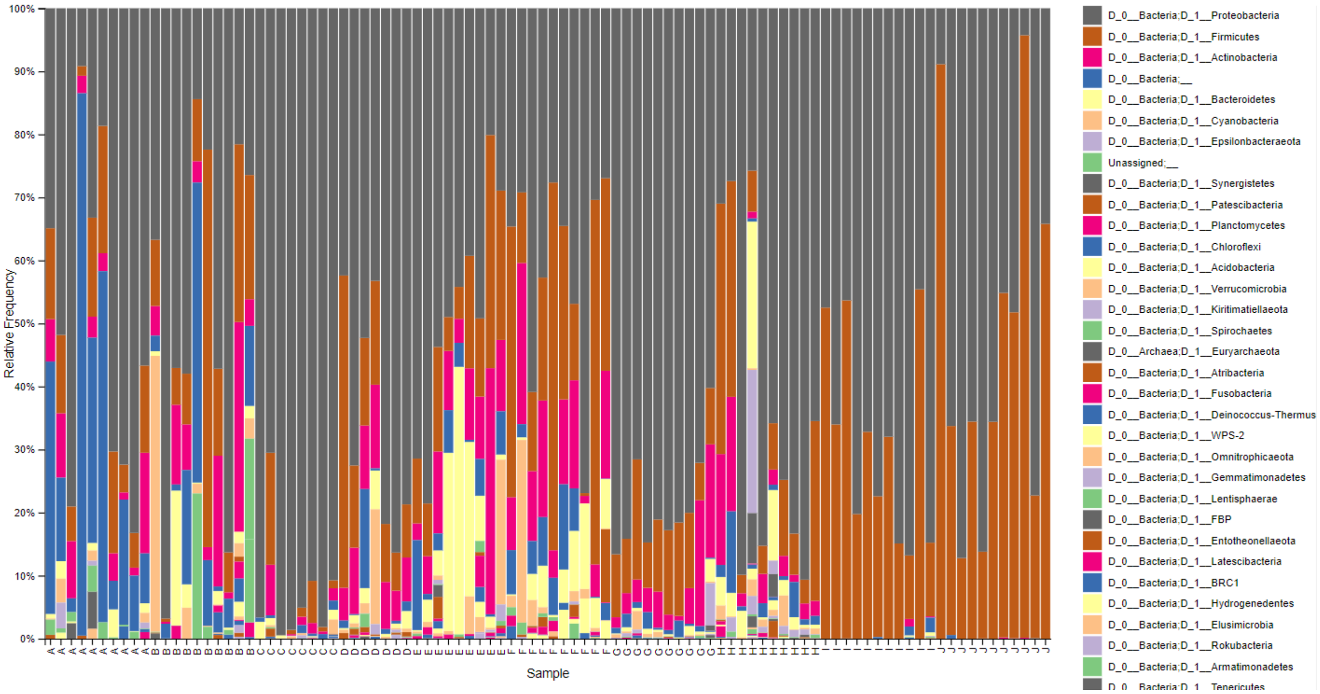


Figure 2: Barplots of taxonomic abundance for each sample at level 2 (phylum). Source: 2-taxonomy-barplots/index.html

Group	A	B	C	D	E	F	G	H	I	J
D_1__Proteobacteria	53.02	49.06	92.33	65.80	52.07	39.50	78.47	59.88	69.44	57.99
D_1__Firmicutes	11.73	14.45	3.25	16.01	13.85	29.91	10.91	13.16	29.80	41.94
D_1__Actinobacteria	6.08	10.16	1.71	8.46	10.12	12.06	6.80	4.79	0.17	0.04
D_0__Bacteria;__	22.95	7.93	0.75	2.46	5.06	4.29	1.14	2.02	0.43	0.02
D_1__Bacteroidetes	1.18	4.42	0.74	2.09	11.20	5.48	0.87	7.81	0.08	0.00
D_1__Cyanobacteria	1.00	8.01	0.78	3.56	3.13	2.92	0.57	0.87	0.01	0.00
D_1__Epsilonbacteraeota	0.86	0.00	0.04	0.36	0.39	0.07	0.76	5.25	0.01	0.00
Unassigned;__	1.46	2.67	0.03	0.28	0.44	0.34	0.09	0.13	0.03	0.00
D_1__Synergistetes	0.74	0.00	0.00	0.00	0.26	0.00	0.08	2.21	0.00	0.00
D_1__Patescibacteria	0.08	0.10	0.14	0.54	0.63	2.23	0.03	0.07	0.01	0.00

Table 5: Percentage taxonomic abundance for each sample group at level 2 (phylum) for the 10 most frequent results amongst all samples. Source: 2-taxonomy-barplots/level-2.csv.

Level 3: Class

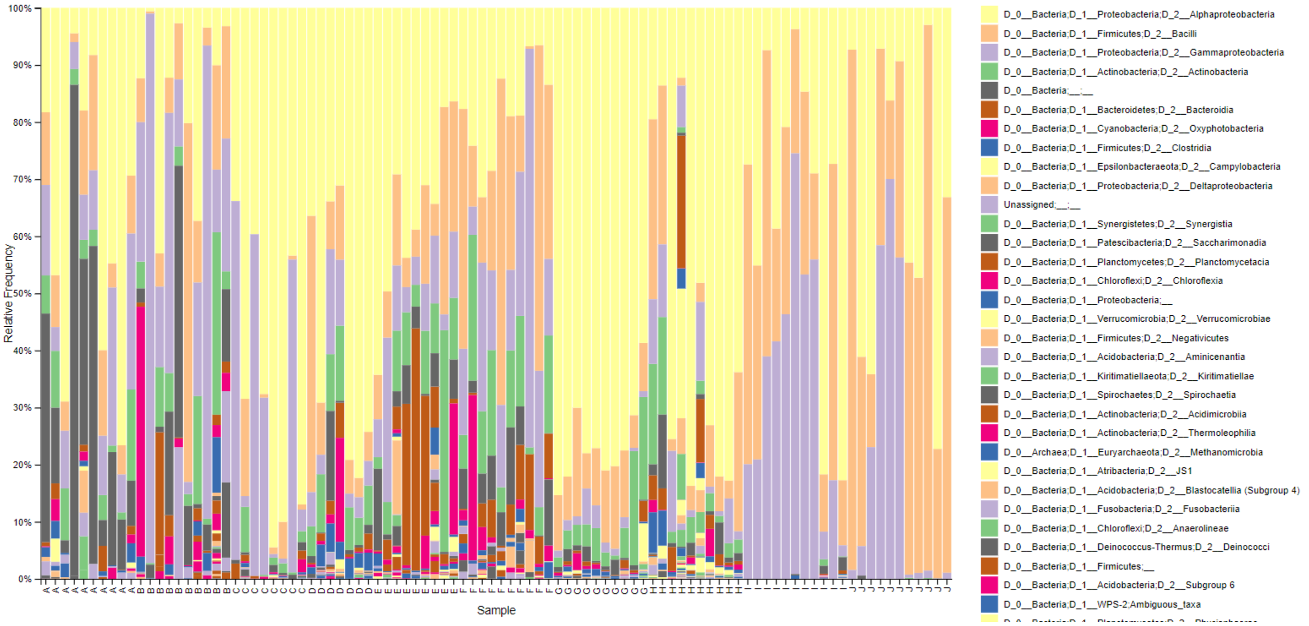


Figure 3: Barplots of taxonomic abundance for each sample at level 3 (class). Source: 2-taxonomy-barplots/index.html

Group	A	B	C	D	E	F	G	H	I	J
D_2__Alphaproteobacteria	39.62	17.56	74.47	58.70	38.23	17.22	75.93	49.86	37.24	34.17
D_2__Bacilli	10.34	12.17	3.02	12.71	12.23	28.93	10.30	9.99	29.73	41.94
D_2__Gammaproteobacteria	12.25	31.41	17.85	6.86	11.34	21.29	2.26	7.90	32.15	23.82
D_2__Actinobacteria	6.04	9.36	1.69	8.36	9.41	11.64	6.76	4.66	0.16	0.04
D_0__Bacteria;__;__	22.95	7.93	0.75	2.46	5.06	4.29	1.14	2.02	0.43	0.02
D_2__Bacteroidia	1.18	4.42	0.74	2.09	11.20	5.46	0.86	7.78	0.08	0.00
D_2__Oxyphotobacteria	1.00	8.01	0.78	3.56	3.11	2.92	0.55	0.83	0.01	0.00
D_2__Clostridia	1.30	2.17	0.22	3.13	1.18	0.67	0.51	2.44	0.06	0.00
D_2__Campylobacteria	0.86	0.00	0.04	0.36	0.39	0.07	0.76	5.25	0.01	0.00
D_2__Deltaproteobacteria	1.02	0.08	0.01	0.18	2.10	0.36	0.12	1.92	0.01	0.00

Table 6: Percentage taxonomic abundance for each sample group at level 3 (class) for the 10 most frequent results amongst all samples. Source: 2-taxonomy-barplots/level-3.csv

Level 4: Order

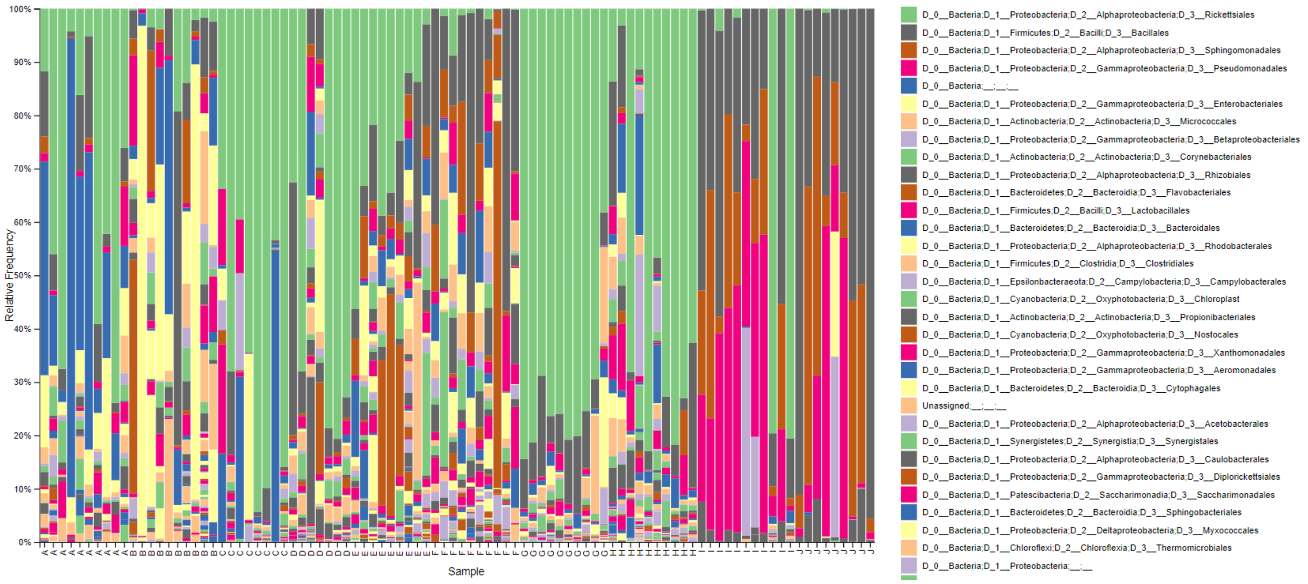


Figure 4: Barplots of taxonomic abundance for each sample at level 4 (order). Source: 2-taxonomy-barplots/index.html

Group	A	B	C	D	E	F	G	H	I	J
D_3__Rickettsiales	37.13	4.44	73.62	48.28	27.63	1.53	72.89	46.37	18.75	0.07
D_3__Bacillales	8.34	8.99	2.72	11.24	10.35	27.85	9.31	7.93	29.69	41.94
D_3__Sphingomonadales	0.62	7.19	0.05	0.79	4.88	6.14	0.08	1.11	16.57	30.20
D_3__Pseudomonadales	2.36	4.14	2.06	3.10	1.84	6.23	0.99	2.49	27.68	15.21
D_0__Bacteria;__;__;__	22.95	7.93	0.75	2.46	5.06	4.29	1.14	2.02	0.43	0.02
D_3__Enterobacteriales	7.62	24.32	4.68	1.04	4.73	5.84	0.70	0.74	0.07	3.52
D_3__Micrococcales	2.09	3.67	0.65	2.40	4.77	5.13	4.83	1.75	0.10	0.02
D_3__Betaproteobacteriales	0.66	1.73	2.42	2.04	1.79	2.04	0.30	2.71	4.38	5.05
D_3__Corynebacteriales	2.19	3.62	0.56	3.72	2.90	3.40	1.31	1.61	0.04	0.01
D_3__Rhizobiales	1.18	3.03	0.21	0.97	2.46	4.25	0.22	1.19	1.75	3.80

Table 7: Percentage taxonomic abundance for each sample group at level 4 (order) for the 10 most frequent results amongst all samples. Source: 2-taxonomy-barplots/level-4.csv.

Level 5: Family

Figure 5: Barplots of taxonomic abundance for each sample at level 5 (family). Source: 2-taxonomy-barplots/index.html

Group	A	B	C	D	E	F	G	H	I	J
D_4__Anaplasmataceae	36.93	2.80	73.51	48.26	27.08	1.35	72.87	46.29	18.24	0.07
D_4__Bacillaceae	1.05	1.21	0.04	0.21	3.95	24.07	7.39	6.05	15.19	28.24
D_4__Sphingomonadaceae	0.62	7.19	0.05	0.79	4.88	6.14	0.08	1.11	16.57	30.20
D_4__Moraxellaceae	1.61	3.92	1.97	3.02	1.69	4.72	0.94	2.33	26.80	14.70
D_4__Enterobacteriaceae	7.62	24.32	4.68	1.04	4.73	5.84	0.70	0.74	0.07	3.52
D_0__Bacteria;__;__;__;__	22.95	7.93	0.75	2.46	5.06	4.29	1.14	2.02	0.43	0.02
D_4__Staphylococcaceae	6.75	5.69	2.61	10.64	4.23	2.35	1.06	1.39	0.02	0.01
D_4__Planococcaceae	0.41	1.46	0.06	0.34	1.95	1.28	0.71	0.24	14.03	10.40
D_4__Micrococcaceae	1.49	2.07	0.27	1.83	3.73	2.82	4.18	1.30	0.07	0.02
D_4__Burkholderiaceae	0.64	1.36	2.39	1.56	1.55	1.85	0.19	1.68	4.35	5.05

Table 8: Percentage taxonomic abundance for each sample group at level 5 (family) for the 10 most frequent results amongst all samples. Source: 2-taxonomy-barplots/level-5.csv.

Level 6: Genus

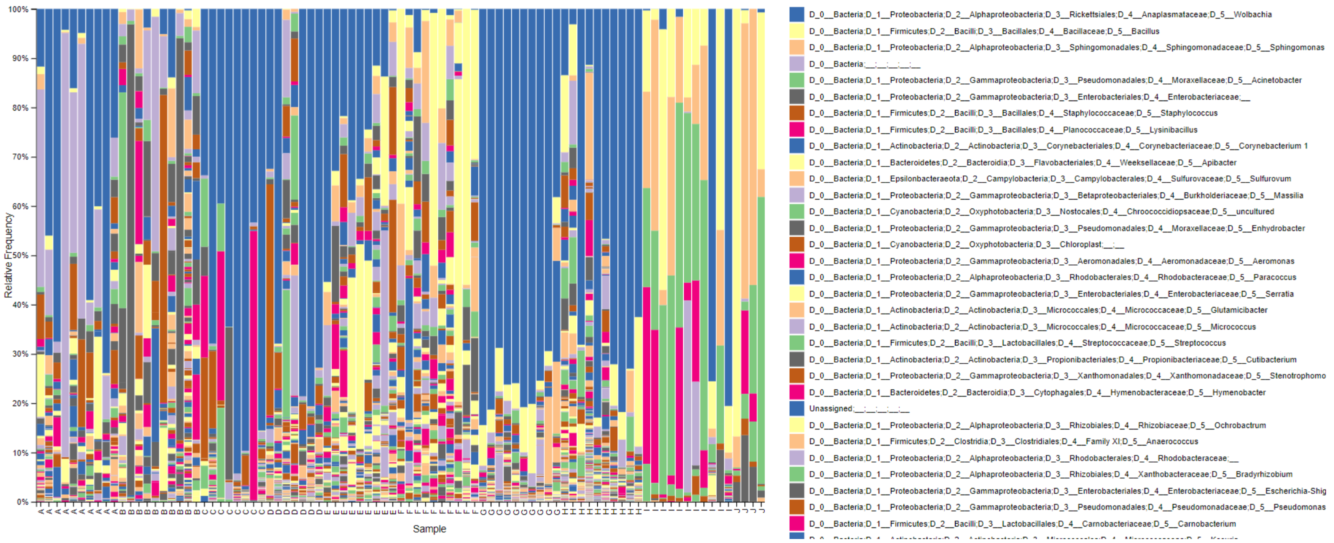


Figure 6: Barplots of taxonomic abundance for each sample at level 6 (genus). Source: 2-taxonomy-barplots/index.html

Group	A	B	C	D	E	F	G	H	I	J
D_5__Wolbachia	36.89	2.50	73.51	48.26	27.08	1.35	72.87	46.29	18.24	0.07
D_5__Bacillus	1.00	0.91	0.04	0.08	3.44	23.67	5.99	5.63	15.19	28.24
D_5__Sphingomonas	0.51	4.28	0.03	0.58	3.79	5.49	0.06	1.04	16.54	30.20
D_0__Bacteria;__;__;__;__;__	22.95	7.93	0.75	2.46	5.06	4.29	1.14	2.02	0.43	0.02
D_5__Acinetobacter	0.59	1.40	1.83	1.26	0.72	1.97	0.58	1.86	26.72	14.62
D_4__Enterobacteriaceae;__	2.09	19.18	4.68	0.66	3.88	4.50	0.28	0.52	0.06	3.52
D_5__Staphylococcus	6.64	5.67	2.61	10.64	4.23	2.34	1.05	1.36	0.02	0.01
D_5__Lysinibacillus	0.41	1.17	0.06	0.34	1.75	1.21	0.64	0.23	14.03	10.40
D_5__Corynebacterium 1	1.04	2.23	0.39	2.90	1.91	1.57	0.85	0.97	0.01	0.00
D_5__Apibacter	0.00	0.00	0.00	0.00	9.17	0.93	0.00	0.00	0.00	0.00

Table 9: Percentage taxonomic abundance for each sample group at level 6 (genus) for the 10 most frequent results amongst all samples. Source: 2-taxonomy-barplots/level-6.csv

Level 7: Species

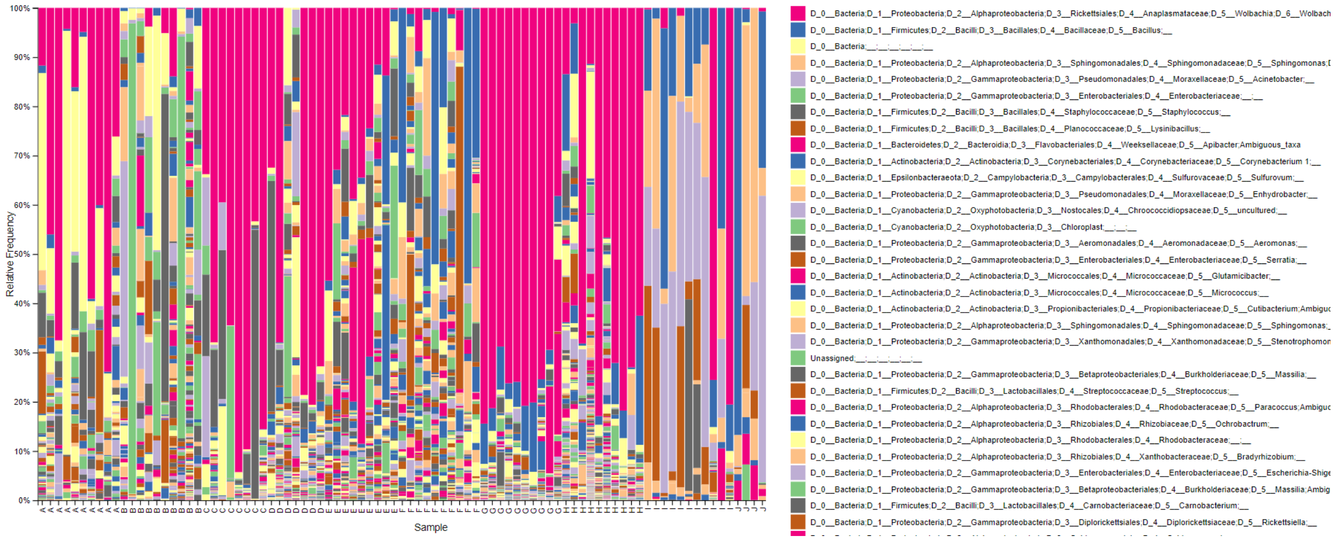


Figure 7: Barplots of taxonomic abundance for each sample at level 7 (species). Source: 2-taxonomy-barplots/index.html

Group	A	B	C	D	E	F	G	H	I	J
D_6__Wolbachia sp.	36.89	2.50	73.51	48.26	27.08	1.35	72.87	46.29	18.24	0.07
D_5__Bacillus;__	1.00	0.91	0.04	0.04	3.44	23.66	5.99	5.62	15.19	28.24
D_0__Bacteria;__;__;__;__;__;__	22.95	7.93	0.75	2.46	5.06	4.29	1.14	2.02	0.43	0.02
D_6__Sphingomonas paucimobilis	0.37	0.60	0.00	0.02	3.04	4.39	0.03	1.02	16.42	30.11
D_5__Acinetobacter;__	0.59	1.40	1.83	1.26	0.72	1.97	0.58	1.86	26.72	14.62
D_4__Enterobacteriaceae;__;__	2.09	19.18	4.68	0.66	3.88	4.50	0.28	0.52	0.06	3.52
D_5__Staphylococcus;__	6.64	5.67	2.61	10.64	4.23	2.34	1.05	1.36	0.02	0.01
D_5__Lysinibacillus;__	0.41	1.17	0.06	0.34	1.75	1.21	0.64	0.22	14.03	10.40
D_5__Apibacter;Ambiguous_taxa	0.00	0.00	0.00	0.00	9.17	0.93	0.00	0.00	0.00	0.00
D_5__Corynebacterium 1;__	1.04	1.35	0.37	2.80	1.90	1.54	0.80	0.94	0.01	0.00

Table 10: Percentage taxonomic abundance for each sample group at level 7 (species) for the 10 most frequent results amongst all samples. Source: 2-taxonomy-barplots/level-7.csv

3. Alpha and Beta diversity

Alpha rarefaction

Prior to performing diversity analyses, alpha rarefaction curves were generated to assess species richness from the results of sampling. Rarefaction curves grow rapidly at first, as the most common species are found, but plateau as only the rarest species remain to be sampled. Most accurate alpha and beta diversity analysis results will be obtained if a sampling depth is used where rarefaction curves have levelled off. After this point, increasing the sampling depth will add no benefit to the diversity analyses.

Output files:

3-alpha-beta-diversity/alpha-rarefaction: tables and html file for visualisation of sample depth against the number of observed OTUs, Faith's phylogenetic diversity and Shannon's index

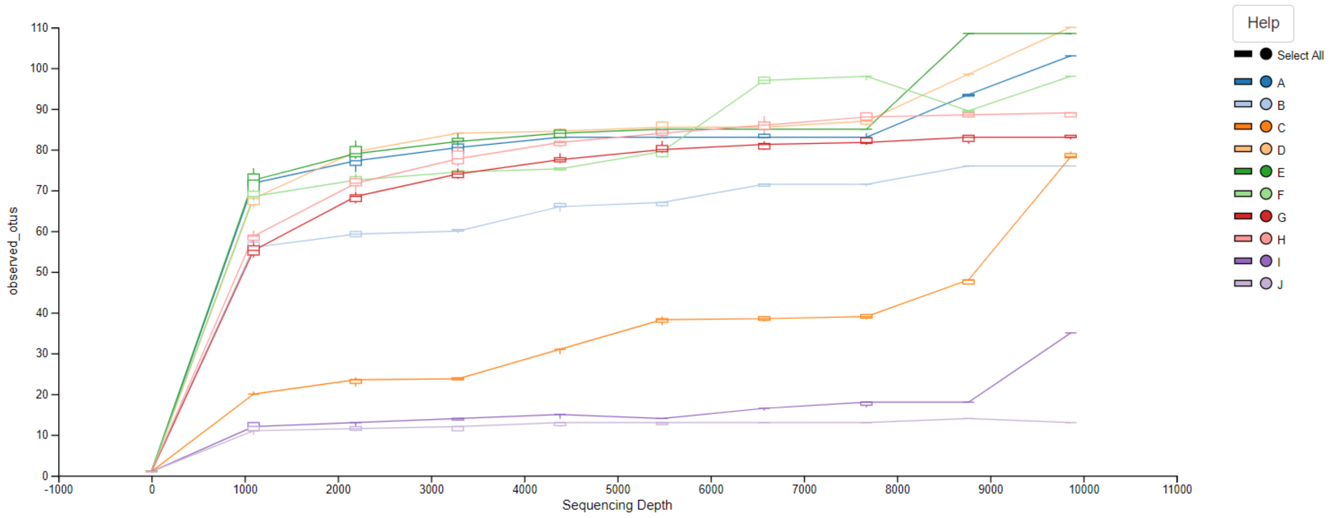


Figure 8: Alpha rarefaction plot of observed OTUs against sampling depth for each sample group. Source: 3-alpha-beta-diversity/alpha-rarefaction/index.html

Alpha diversity

Alpha diversity is a metric used to assess species richness, i.e. the diversity *within* samples. The analysis was performed using the *mafft, FastTre*e and *diversity* qiime2 plugins with a sampling depth of 1939 reads, to allow for the inclusion of all samples.

Categorical groups (“Group” and “Subgroup”) were subjected to group-wise comparisons using a Kruskal-Wallis test with false discovery rate (FDR) correction for the following four metrics: Observed OTUs (a taxonomy-based metric), Faith’s phylogenetic diversity (a phylogeny-based metric), Shannon’s index and Pielou’s eveness (metrics for exploring species abundance and evenness).

Numeric groups (“Time”) were subjected Spearman correlation analysis for Faith’s phylogenetic diversity and Pielou’s eveness metrics.

Output files:

3-alpha-beta-diversity/alpha-diversity/observed-otus: tables, pairwise boxplots and html file for visualization of pairwise comparison (Kruskal-Wallis tests) of observed OTUs for categorical groups (“Group” and “Subgroup”)

3-alpha-beta-diversity/alpha-diversity/faiths-phylogenetic-diversity: tables, pairwise boxplots and html file for visualization of pairwise comparison (Kruskal-Wallis tests) of the Faith’s phylogenetic diversity metric for categorical groups (“Group” and “Subgroup”)

3-alpha-beta-diversity/alpha-diversity/shannon-index: tables, pairwise boxplots and html file for visualization of pairwise comparison (Kruskal-Wallis tests) of the Shannon’s index metric for categorical groups (“Group” and “Subgroup”)

3-alpha-beta-diversity/alpha-diversity/pielous-eveness: tables, pairwise boxplots and html file for visualization of pairwise comparison (Kruskal-Wallis tests) of the Pielou’s eveness metric for categorical groups (“Group” and “Subgroup”)

3-alpha-beta-diversity/alpha-diversity/correlations/faiths-phylogenetic-diversity: table and html file for visualization of Spearman correlation analysis of the Faith’s phylogenetic diversity metric for numeric groups (“Time”)

3-alpha-beta-diversity/alpha-diversity/correlations/pielous-eveness: table and html file for visualization of Spearman correlation analysis of the Pielou’s eveness metric for numeric groups (“Time”)

A summary of alpha diversity comparisons between “Groups” is shown below:

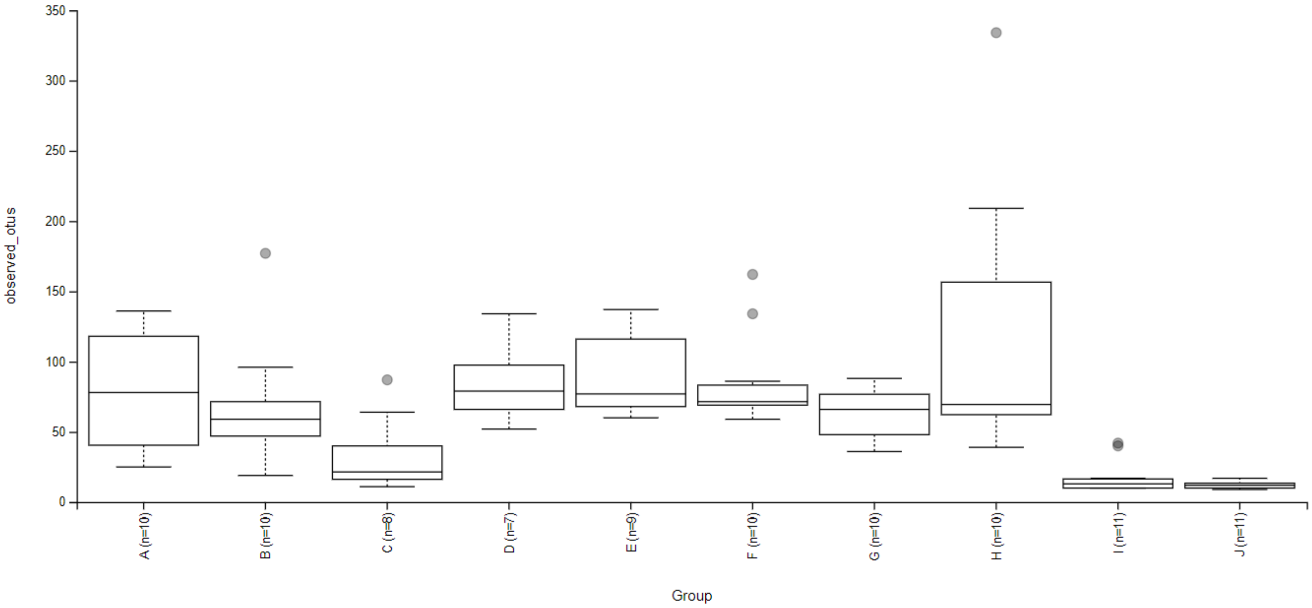


Figure 9: Box and whisker plot of the number of observed OTUs, i.e. the number of distinct features (ASVs or OTUs) in a sample or group of samples to allow comparison of species richness (Ref: DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K. Huber, T., Davis, D., Hu, P., Andersen, G.L. (2006). "Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB". Applied and Environmental Microbiology (72): 5069–5072.) Source: 3-alpha-beta-diversity/alpha-diversity/observed-otus/index.html

Group 1	Group 2	H	p-value	q-value
F (n=10)	**I (n=11)**	15.11	1.015E-04	0.001
A (n=10)	**J (n=11)**	15.06	1.042E-04	0.001
F (n=10)	**J (n=11)**	15.06	1.042E-04	0.001
G (n=10)	**J (n=11)**	15.06	1.042E-04	0.001
B (n=10)	**J (n=11)**	15.05	1.048E-04	0.001
H (n=10)	**J (n=11)**	15.05	1.048E-04	0.001
E (n=9)	**I (n=11)**	14.25	1.601E-04	0.001
E (n=9)	**J (n=11)**	14.20	1.647E-04	0.001
G (n=10)	**I (n=11)**	14.03	1.800E-04	0.001
H (n=10)	**I (n=11)**	14.02	1.809E-04	0.001

Table 11: Pairwise Kruskal-Wallis test of the 10 most significant differences in the number of observed OTUs between groups. Source: 3-alpha-beta-diversity/alpha-diversity/observed-otus/Kruskal-wallis-pairwise.csv.

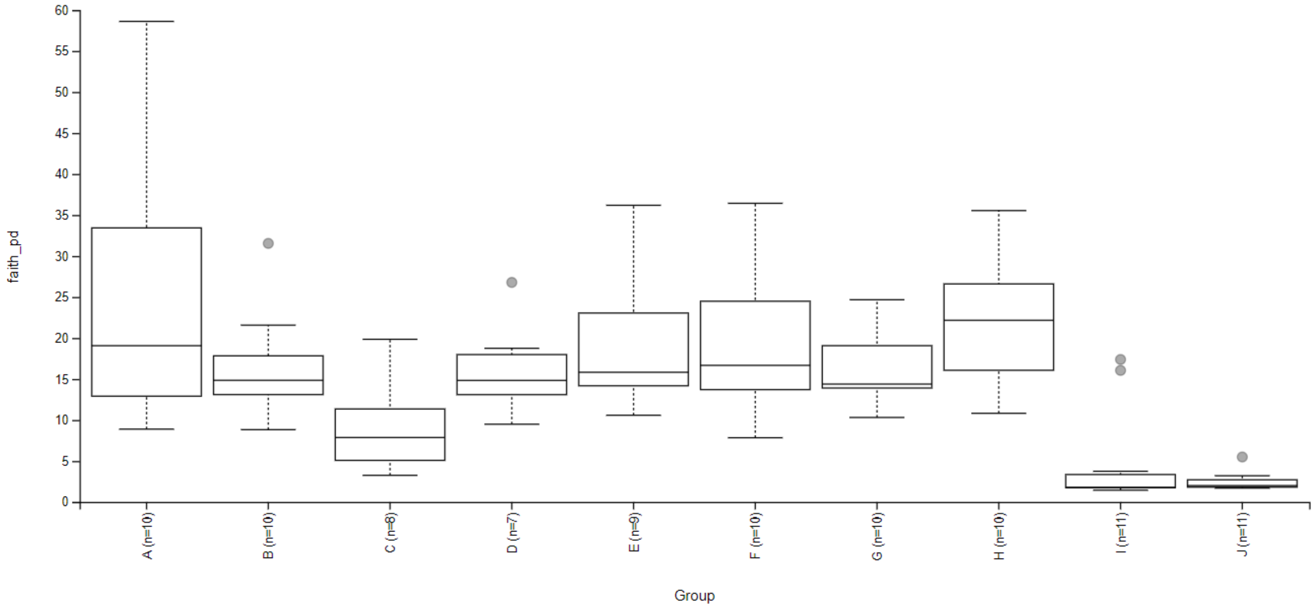


Figure 10: Box and whisker plot of the Faith's phylogenetic diversity metric, a qualitative measure of biodiversity that incorporates phylogenetic relationships/difference between species (ASVs or OTUs) by assessing the sum of the length of branches. (Ref: Faith. D.P. (1992). "Conservation evaluation and phylogenetic diversity". Biological Conservation. (61) 1-10. Source: 3-alpha-beta-diversity/alpha-diversity/faiths-phylogenetic-diversity/index.html.

Group 1	Group 2	H	p-value	q-value
A (n=10)	**J (n=11)**	15.00	1.075E-04	0.001
B (n=10)	**J (n=11)**	15.00	1.075E-04	0.001
F (n=10)	**J (n=11)**	15.00	1.075E-04	0.001
G (n=10)	**J (n=11)**	15.00	1.075E-04	0.001
H (n=10)	**J (n=11)**	15.00	1.075E-04	0.001
E (n=9)	**J (n=11)**	14.14	1.694E-04	0.001
D (n=7)	**J (n=11)**	12.16	4.888E-04	0.003
C (n=8)	**J (n=11)**	11.46	0.001	0.004
H (n=10)	**I (n=11)**	11.42	0.001	0.004
A (n=10)	**I (n=11)**	10.95	0.001	0.004

Table 12: Pairwise Kruskal-Wallis test of the 10 most significant differences in the Faith's phylogenetic diversity metric between groups. Source: 3-alpha-beta-diversity/alpha-diversity/faith-phylogenetic-diversity/Kruskal-wallis-pairwise.csv

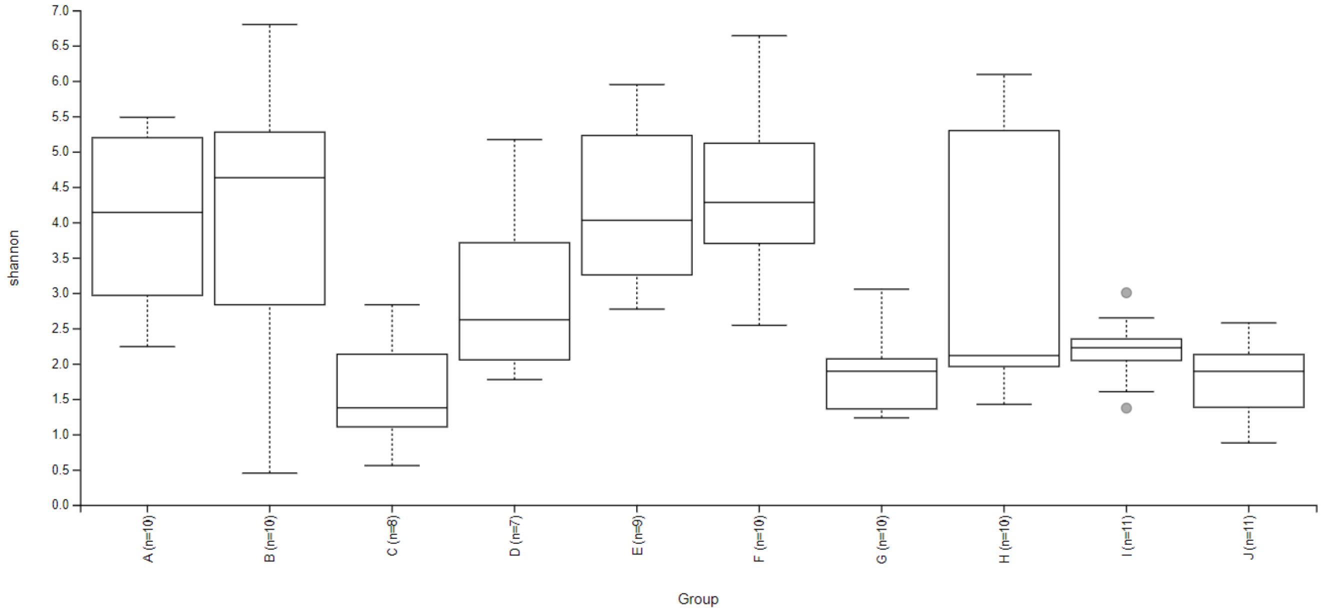


Figure 11: Box and whisker plot of the Shannon's index metric, a quantitative measure of community richness and diversity using a natural logarithm. It accounts for both abundance and evenness of the taxa present (Ref: Shannon, C.E. and Weaver, W. (1949). "The mathematical theory of communication". University of Illonois Press, Champaign, Illonois). Source: 3-alpha-beta-diversity/alpha-diversity/shannon-index/index.html.

Group 1	Group 2	H	p-value	q-value
F (n=10)	**J (n=11)**	14.46	1.432E-04	0.002
E (n=9)	**J (n=11)**	14.14	1.694E-04	0.002
A (n=10)	**J (n=11)**	13.93	1.899E-04	0.002
F (n=10)	**I (n=11)**	13.93	1.899E-04	0.002
F (n=10)	**G (n=10)**	13.72	2.122E-04	0.002
E (n=9)	**I (n=11)**	13.02	3.077E-04	0.002
A (n=10)	**G (n=10)**	12.62	3.811E-04	0.002
E (n=9)	**G (n=10)**	12.33	4.465E-04	0.003
C (n=8)	**F (n=10)**	12.01	0.001	0.003
C (n=8)	**E (n=9)**	11.34	0.001	0.003

Table 13: Pairwise Kruskal-Wallis test of the 10 most significant differences in the Shannon's Index metric between groups. Source: 3-alpha-beta-diversity/alpha-diversity/shannon-index/Kruskal-wallis-pairwise.csv

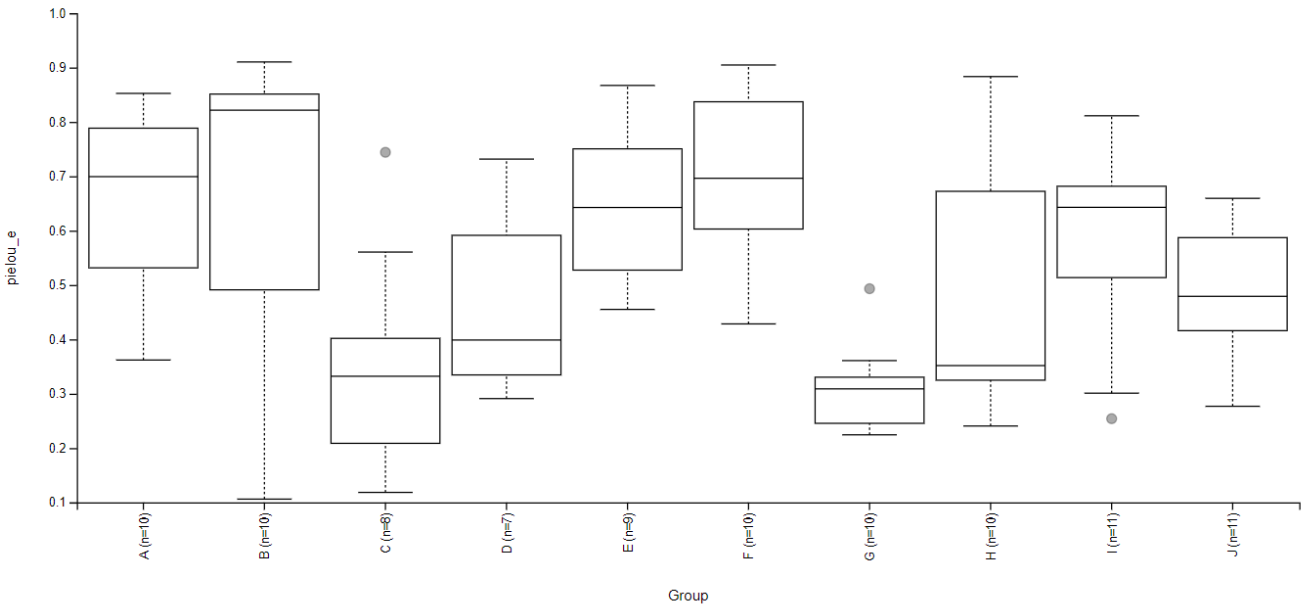


Figure 12: Box and whisker plot of the Pielou's evenness metric, a measure of relative evenness of species richness, to identify whether species (ASVs or OTUs) have similar abundance levels within a sample or group of samples, or whether some species dominate (Ref: Pielou, E. (1966). "The measurement of diversity in different types of biological collections". J. Theor. Biol. (13): 131-144). Source: 3-alpha-beta-diversity/alpha-diversity/pielous-eveness/index.html.

Group 1	Group 2	H	p-value	q-value
F (n=10)	G (n=10)	13.72	2.122E-04	0.005
A (n=10)	G (n=10)	13.17	2.851E-04	0.005
E (n=9)	G (n=10)	12.91	3.274E-04	0.005
G (n=10)	J (n=11)	9.60	0.002	0.022
G (n=10)	I (n=11)	9.17	0.002	0.022
A (n=10)	C (n=8)	8.08	0.004	0.029
C (n=8)	F (n=10)	8.08	0.004	0.029
F (n=10)	J (n=11)	7.16	0.007	0.041
B (n=10)	G (n=10)	7.00	0.008	0.041
C (n=8)	E (n=9)	6.26	0.012	0.056

Table 14: Pairwise Kruskal-Wallis test of the 10 most significant differences in the Pielou's evenness metric between groups. Source: 3-alpha-beta-diversity/alpha-diversity/pielous-evenness/Kruskal-wallis-pairwise.csv

Beta diversity

Beta diversity is a metric used to compare species diversity and abundance *between* samples. The analysis was performed using the *mafft, FastTree*, *diversity* and *emperor* qiime2 plugins with a sampling depth of 1939 reads, to allow for the inclusion of all samples.

Categorical groups ("Group" and "Subgroup") were plotted on principle coordinate analysis (PCoA) plots and tested for significance using a pairwise permanova test: Jaccard distance (a qualitative measure of community dissimilarity), Bray-Curtis dissimilarity (a quantitative measure of community dissimilarity), unweighted and weighted UniFrac distances (qualitative and quantitative measures of community dissimilarity that incorporate phylogenetic relationships between features).

Numeric groups ("Time") were subjected Spearman correlation analysis for Bray-Curtis dissimilarity and unweighted UniFrac distance metrics.

Output files:

3-alpha-beta-diversity/beta-diversity/jaccard-distance: html file for interactive visualisation of a principle coordinates analysis (PCoA) plot using Emperor and pairwise tests (permanova) of the Jaccard distance between categorical groups ("Group" and "Subgroup")
3-alpha-beta-diversity/beta-diversity/bray-curtis-dissimilarity: html file for interactive visualisation of a principle coordinates analysis (PCoA) plot using Emperor and pairwise tests (permanova) of the Bray-Curtis dissimilarity between categorical groups ("Group" and "Subgroup")
3-alpha-beta-diversity/beta-diversity/unweighted-unifrac-distance: html file for interactive visualisation of a principle coordinates analysis (PCoA) plot using Emperor and pairwise tests (permanova) of the unweighted UniFrac distance between categorical groups ("Group" and "Subgroup")
3-alpha-beta-diversity/beta-diversity/weighted-unifrac-distance: html file for interactive visualisation of a principle coordinates analysis (PCoA) plot using Emperor and pairwise tests (permanova) of the weighted UniFrac distance between categorical groups ("Group" and "Subgroup")
3-alpha-beta-diversity/correlations/ bray-curtis-dissimilarity: html file for interactive visualisation of a principle coordinates analysis (PCoA) plot using Emperor of the Bray-Curtis dissimilarity for numeric groups ("Time")
3-alpha-beta-diversity/correlations/unweighted-unifrac-distance: html file for interactive visualisation of a principle coordinates analysis (PCoA) plot using Emperor of the unweighted UniFrac distance for numeric groups ("Time").

A summary of beta diversity comparisons between "Groups" is shown below:

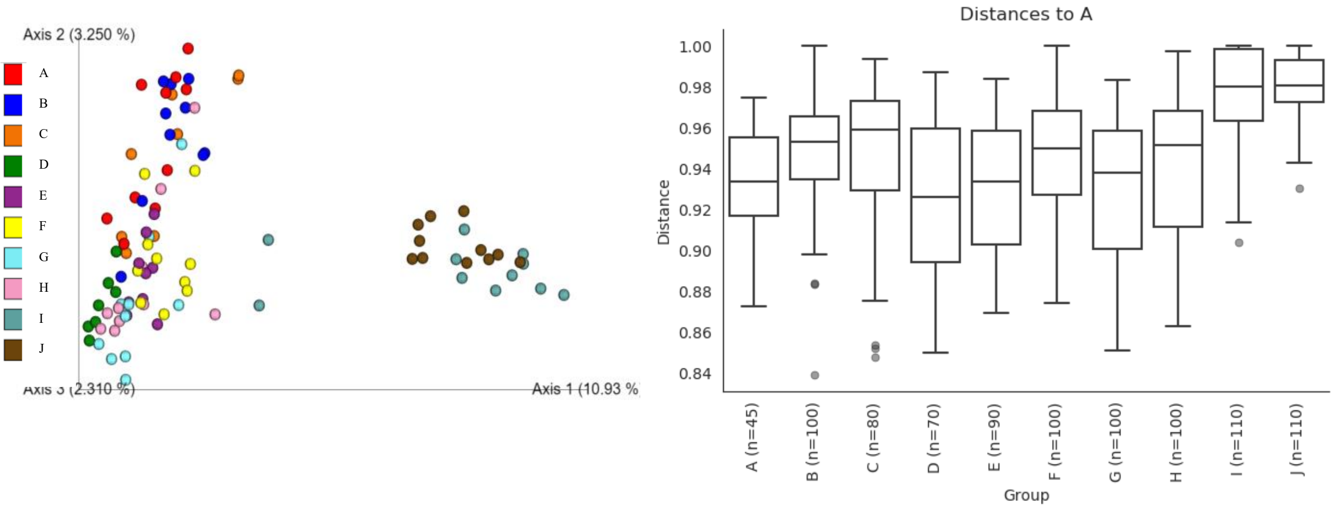


Figure 13: PCoA emperor plot of all samples, and box and whisker plots of pairwise comparisons (PERMANOVA) between groups (distances to Group A shown here), using the Jaccard distance metric. Jaccard distance is a measure of differences in microbial *composition* between two samples, regardless of abundance. Values from 0 to 1 are plotted, where 0 = both samples share the same species; 1 = both samples have no species in common. (Ref: Jaccard, P. (1908). "Nouvelles recherches sur la distribution florale." Bull. Soc. V and. Sci. Nat., (44):223-270.). Source: 3-alpha-beta-diversity/beta-diversity/jaccard-distance/PCoA/emperor.html & permanova/A-boxplots.png.

Group 1	Group 2	Sample size	Permutations	pseudo-F	p-value	q-value
A	C	18	999	1.46	0.001	0.001
A	D	17	999	1.69	0.001	0.001
A	E	19	999	1.45	0.001	0.001
A	F	20	999	1.42	0.001	0.001
A	G	20	999	1.72	0.001	0.001
A	H	20	999	1.49	0.001	0.001
A	I	21	999	4.81	0.001	0.001
A	J	21	999	4.55	0.001	0.001
B	C	18	999	1.48	0.001	0.001
B	D	17	999	1.78	0.001	0.001

Table 15: Pairwise PERMANOVA test of the 10 most significant differences in Jaccard distance between groups. Source: 3-alpha-beta-diversity/beta-diversity/jaccard-distance/permanova/permanova-pairwise.csv

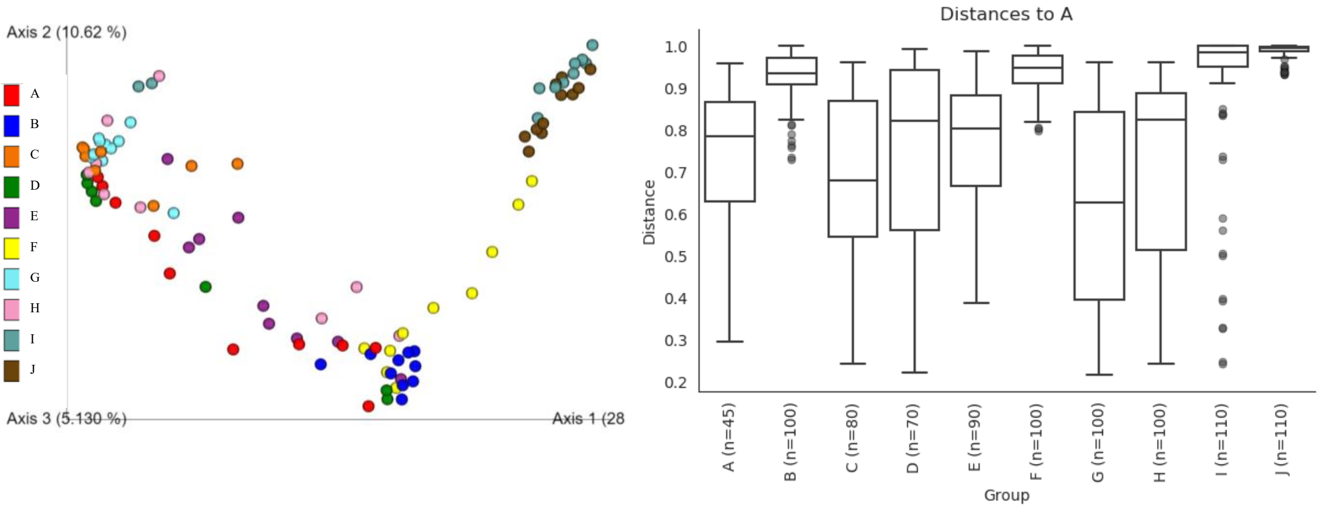


Figure 14: PCoA emperor plot of all samples, and box and whisker plots of pairwise comparisons (PERMANOVA) between groups (distances to Group A shown here), using the Bray-Curtis dissimilarity metric. Bray-Curtis dissimilarity is a measure of differences in microbial *abundance* between two samples. Values from 0 to 1 are plotted, where 0 = both samples share the same species at exactly the same abundances; 1 = both samples have completely different species abundances. (Ref: Sorenson, T. (1948) "A method of establishing groups of equal amplitude in plant sociology based on similarity of species content." Kongelige Danske Vide- nskabernes Selskab 5.1-34: 4-7.). Source: 3-alpha-beta-diversity/beta-diversity/bray-curtis-dissimilarity/PCoA/emperor.html & permanova/A-boxplots.png.

Group 1	Group 2	Sample size	Permutations	pseudo-F	p-value	q-value
A	B	20	999	2.89	0.001	0.001
A	F	20	999	3.96	0.001	0.001
A	G	20	999	4.80	0.001	0.001
A	I	21	999	9.45	0.001	0.001
A	J	21	999	11.12	0.001	0.001
B	C	18	999	6.50	0.001	0.001
B	D	17	999	3.51	0.001	0.001
B	E	19	999	3.04	0.001	0.001
B	F	20	999	1.92	0.001	0.001
B	G	20	999	9.98	0.001	0.001

Table 16: Pairwise PERMANOVA test of the 10 most significant differences in Bray-Curtis dissimilarity between groups. Source: 3-alpha-beta-diversity/beta-diversity/bray-curtis-dissimilarity/permanova/permanova-pairwise.csv

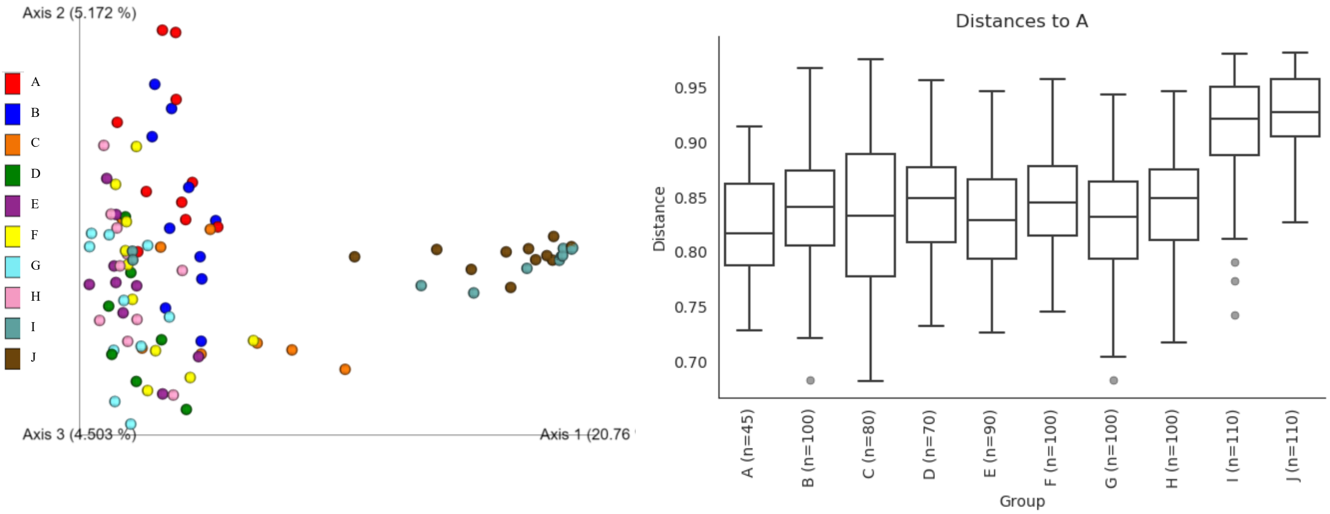


Figure 15: PCoA emperor plot of all samples, and box and whisker plots of pairwise comparisons (PERMANOVA) between groups (distances to Group A shown here), using the unweighted UniFrac distance metric. Unweighted UniFrac distance is a measure of sequence distances between two samples, regardless of abundance. The unique branch length fraction is measured, i.e. the fraction of branch length that is shared between two samples or unique to one or the other sample. (Ref: Lozupone, C. and Knight, R. (2005). "UniFrac: a new phylogenetic method for comparing microbial communities." Applied and environmental microbiology 71 (12): 8228-8235.) Source: 3-alpha-beta-diversity/beta-diversity/unweighted-unifrac-distance/PCoA/emperor.html & permanova/A-boxplots.png.

Group 1	Group 2	Sample size	Permutations	pseudo-F	p-value	q-value
A	D	17	999	2.46	0.001	0.002
A	E	19	999	2.22	0.001	0.002
A	F	20	999	2.17	0.001	0.002
A	G	20	999	2.74	0.001	0.002
A	H	20	999	2.24	0.001	0.002
A	I	21	999	7.03	0.001	0.002
A	J	21	999	8.82	0.001	0.002
B	D	17	999	2.06	0.001	0.002
B	E	19	999	2.00	0.001	0.002
B	G	20	999	2.53	0.001	0.002

Table 17: Pairwise PERMANOVA test of the 10 most significant differences in the unweighted UniFrac distance metric between groups. Source: 3-alpha-beta-diversity/beta-diversity/unweighted-unifrac-distance/permanova/permanova-pairwise.csv

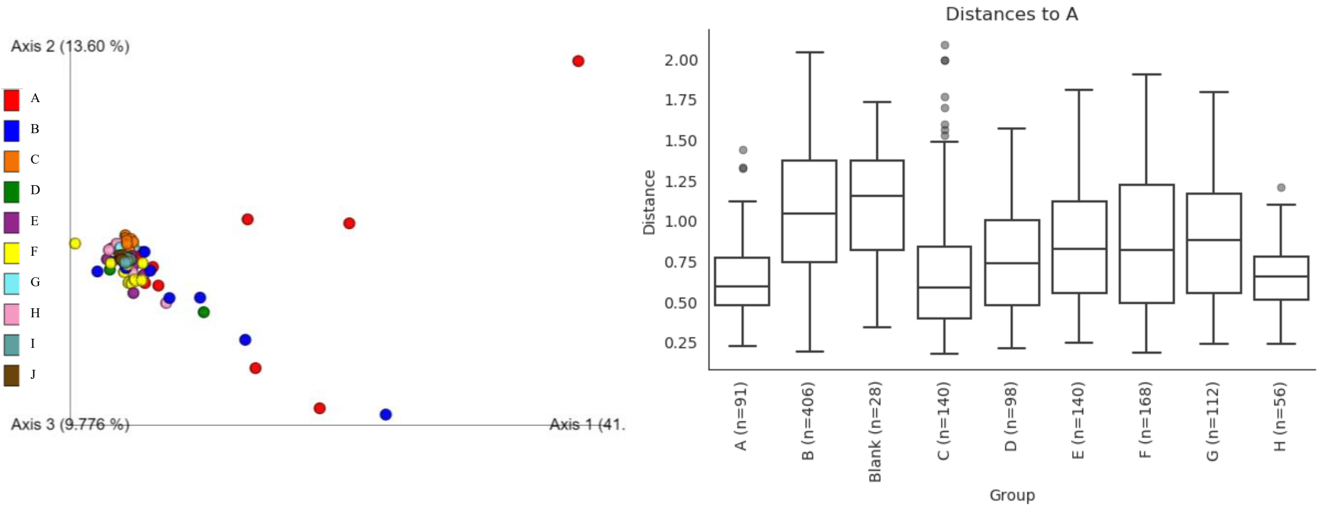


Figure 16: PCoA emperor plot of all samples, and box and whisker plots of pairwise comparisons (PERMANOVA) between groups (distances to Group A shown here), using the weighted UniFrac distance metric. Weighted UniFrac distance is a measure of sequence distances between two samples, weighted by relative abuncances. The same measurement as for unweighted UniFrac distance but branch length fractions are weighted by relative abundances. (Ref: Lozupone, C. A., Hamady, M., Kelley, S. T., Knight, R. (2007). "Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities". Applied and Environmental Microbiology. 73(5): 1576–85.) Source: 3-alpha-beta-diversity/beta-diversity/weighted-unifrac-distance/PCoA/emperor.html & permanova/A-boxplots.png.

Group 1	Group 2	Sample size	Permutations	pseudo-F	p-value	q-value
A	F	20	999	4.83	0.001	0.002
A	G	20	999	4.43	0.001	0.002
A	I	21	999	5.75	0.001	0.002
A	J	21	999	6.53	0.001	0.002
B	G	20	999	4.38	0.001	0.002
B	I	21	999	3.27	0.001	0.002
B	J	21	999	3.90	0.001	0.002
C	E	17	999	4.06	0.001	0.002
C	F	18	999	6.81	0.001	0.002
C	I	19	999	11.40	0.001	0.002

Table 18: Pairwise PERMANOVA test of the 10 most significant differences in the unweighted UniFrac distance metric between groups. Source: 3-alpha-beta-diversity/beta-diversity/weighted-unifrac-distance/permanova/permanova-pairwise.csv

4. Differential abundance

Differential abundance analysis was performed using ANCOM (*q2-composition* tool) to identify features that are differentially abundant across sample groups. Analysis of composition of microbiomes (ANCOM) is a statistical framework for the comparison of the composition of microbiomes in two or more populations. The relative abundance of a taxon between two ecosystems is compared by computing Aitchison's log-ratio of abundance of each taxon relative to the abundance of all remaining taxa one at a time. For each taxon, ANCOM obtains a count random variable *W* that represents the number of nulls among the tests that are rejected. It has to be noted that ANCOM is only suitable for subtle differences between comparative groups – i.e. less than about 25% of features. ANCOM makes no distributional assumptions and can be implemented in a linear model framework to adjust for covariates as well as model longitudinal data.

Output files:

4-differential-abundance/features: pairwise comparison of features between sample groups for categorical Groups ("Group" and "Subgroup"); tables of statistical results and percentage abundance and html file for visualization of volcano plot and tables

4-differential-abundance/taxonomic-levels: pairwise comparison of groups for categorical Groups ("Group" and "Subgroup") at every taxonomic level (levels 1 – 7); tables of statistical results and percentage abundance and html file for visualisation of volcano plot and tables

A summary of differential abundance results between "Groups" is shown below:

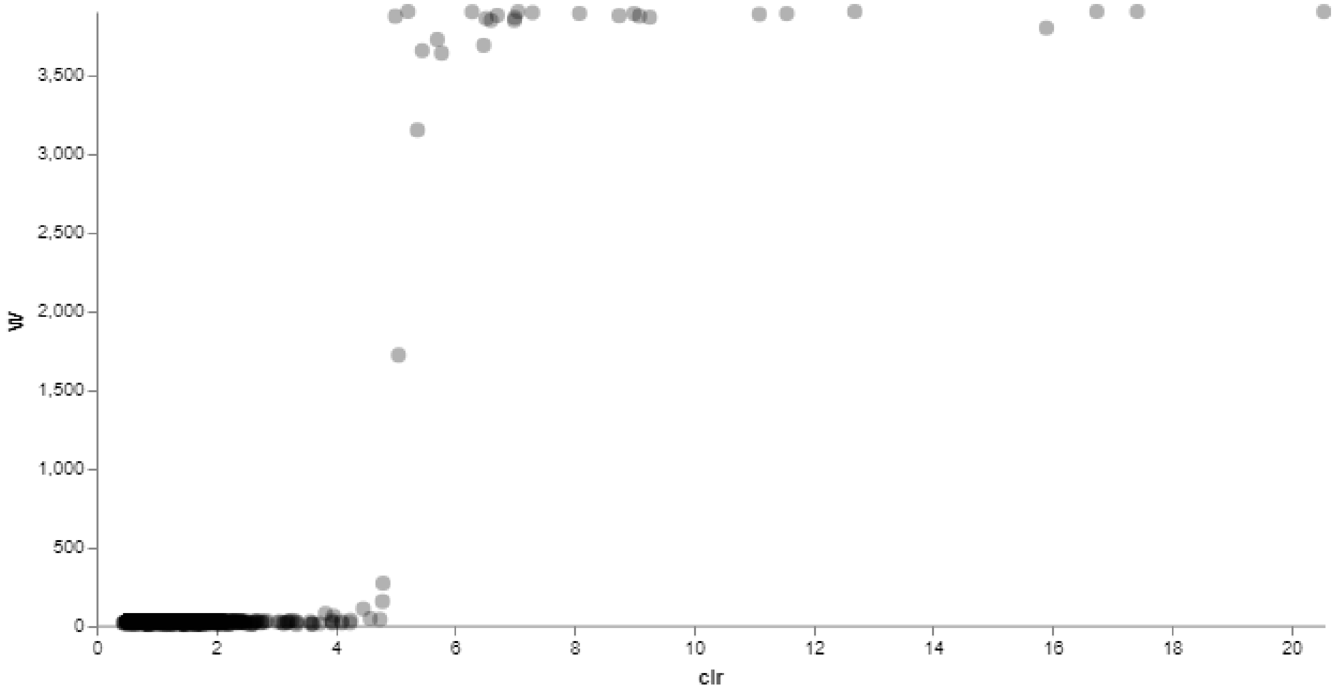


Figure 16: ANCOM – features: Volcano plot of the ANCOM W statistic (= significance) plotted against the f-score (= effect size). Features with high W-statistic and high F-score values are most differentially abundant across groups. (Ref: Mandal, S., Van Treuren W., White, R.A., Eggesbol, M. Knight, R., Peddada, S. D. (2015). "Analysis of composition of microbiomes: a novel method for studying microbial composition". Microb Ecology in Health and Disease 26(10): 27663.) Source 4-differential-abundance/features /Group.

Group	*W*	A	B	C	D	E	F	G	H	I	J
e7d135eb298d798d4cb5fe781da8578a	*3899*	1311	1135	5044	1594	4730	32	115	1453	133	35
c3fb8f26024d2f79566a6165d13460d1	*3899*	500	360	3652	1564	2396	453	129	447	450	58
958d933fa3e25e0cd575e83d5edb16b7	*3899*	365	4590	4982	4687	5521	2176	1071	428	99	1
bf411ff113fa3a86b04836f91ff34b01	*3899*	263	323	475	19	1	128	318	492	1424	339
4ba23fda04b59b853aa1887ba29c4f5e	*3898*	33	32	444	103	1160	1	1	191	37	1
c7cfc4dd7352b68faa5324d3ca4abfbd	*3898*	1	34	848	729	1206	66	25	241	1	17
906b303d8c1b412b57fe0ba7ea35bb17	*3897*	127	18	2923	2473	4571	413	1	958	341	581
2a4cb7bfbe0e5f42a869107ab7a83d25	*3892*	43	97	277	9	1	21	22	118	783	395
8fd8aa091da56e0b26c5444d237e6432	*3887*	1	1	1	1	1	6691	1707	3529	86	1
cfbf2ad616fc614bbcd6045998d6e32a	*3886*	3837	775	107	1	1	928	307	794	486	83

Table 17: ANCOM – features: Table of top 10 features (OTUs/ASVs) for which the null hypothesis was rejected and thus show significant differential abundance across groups. The W-statistic is shown for each feature, as well as the 100th percentile abundances for each group. For full list of features and other percentiles see files in: Source: 4-differential-abundance/features/Group.

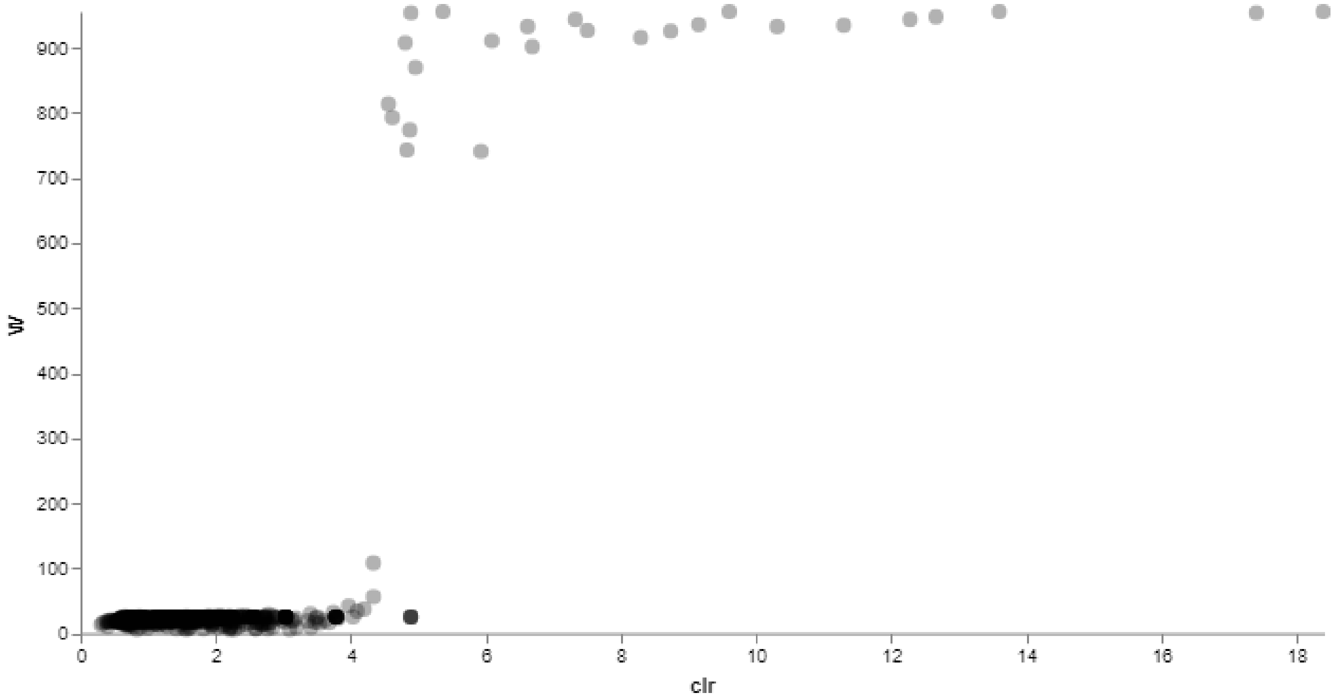


Figure 17: ANCOM – taxonomic level 6: Volcano plot of the ANCOM W statistic (= significance) plotted against the f-score (= effect size) at the taxonomic level 6 (genus) of the Silva database. Features with high W-statistic and high F-score values are most differentially abundant across groups. (Ref: Mandal, S., Van Treuren W., White, R.A., Eggesbol, M. Knight, R., Peddada, S. D. (2015). "Analysis of composition of microbiomes: a novel method for studying microbial composition". Microb Ecology in Health and Disease 26(10): 27663.) Source: 4-differential-abundance/taxonomic-levels/Group/level6 (data for other levels provided too but not shown here).

Group	*W*	A	B	C	D	E	F	G	H	I	J
D_5__Bradyrhizobium	*956*	33	32	444	103	1160	6	5	191	37	1
D_5__Sphingomonas	*956*	1360	1181	5064	1616	4757	87	115	1457	1582	136
D_5__Paenibacillus	*956*	1	4	52	263	736	9	196	98	8	52
D_5__Bacillus	*956*	553	4616	4999	4714	5588	6716	2101	3549	236	160
D_0__Bacteria;__;__;__;__;__	*954*	2508	812	960	183	4	555	1350	5352	6293	548
D_5__Lysinibacillus	*954*	510	360	3676	1564	2435	453	129	447	450	58
D_5__Staphylococcus	*948*	4014	864	214	13	10	928	431	1773	3568	886
D_5__Escherichia-Shigella	*944*	43	97	277	9	1	21	22	118	783	395
D_5__Corynebacterium 1	*944*	782	435	247	12	1	370	298	628	437	597
D_4__Enterobacteriaceae;__	*936*	2873	1149	475	28	2633	806	437	7650	1436	4707

Table 18: ANCOM – taxonomic level 6: Table of top 10 taxonomic assignments (OTUs/ASVs) for which the null hypothesis was rejected and thus show significant differential abundance across groups. The W-statistic is shown for each feature, as well as the 100th percentile abundances for each group. For full list of features and other percentiles for all taxonomic levels, see files in: Source: 4-differential-abundance/taxonomic-levels/Group.