Overview

Here we present the use of Divisive Amplicon Denoising Algorithm 2 (DADA2) pipeline for 16s rRNA data analysis. This pipeline flow allows for the inference of true biological sequences from reads. The datasets used were 16S rRNA amplicon sequencing data from (input sample names).

Set up

To work on DADA2, the data needs to be demultiplexed and Non-biological nucleotides removed: primers, adapters, linkers. If not, preprocessing and filtering steps are required. For paired-end sequence data, the forward and reverse fastq files contain reads in matched order.

For this exercise, we received a total of 124 paired-end reads samples with an average read length of 238bp.

For the DADA2 pipeline, the following packages were installed in the HPC and used for the analysis.

No	Tool	Version
1.	DADA2	1.18.0
2.	phyloseq	1.34.0
3.	dplyr	1.0.3
4.	vegan	2.5.7
5.	phangorn	2.5.5
6.	ggplot2	3.3.3
7.	scales	1.1.1
8.	grid	4.0.2
9.	reshape2	1.4.4
10.	profvis	0.3.7
11.	DECIPHER	2.18.1
12.	Rcolorbrewer	1.1.2

Next, we created a working directory to contain all the output from each process.

Preprocessing

Regarding the number and lengths of the reads, reads ranged from 144 bp to 251. After using FastQC and MultiQC, we further used Dada2 built-in tools for quality check and trimming.

The reads' quality was analyzed by plotting quality profiles of random samples using an in-built feature provided by DADA2. The majority of the reads were of poor sequence quality, with the first 40 bases of most reads exhibiting a low Phred score which could have been attributed to the high percentage N-count at the start of the reads. Primer metadata also indicated the barcode sequences and reverse primers that were still present in the reads and could have contributed to the low quality reported. Moreover, high adapter content characterized the end of the reads.

Figure 1. Raw Quality Profiles for forward reads

Figure 2. Raw Quality Profiles for reverse reads

These details informed the trimming procedure performed on DADA2. We set the trimming parameters to retain ~ 230 bp forward reads and 200bp reverse reads. This is because forward reads maintain better quality throughout, with the quality dropping at the end around position 230, the reverse reads quality drops

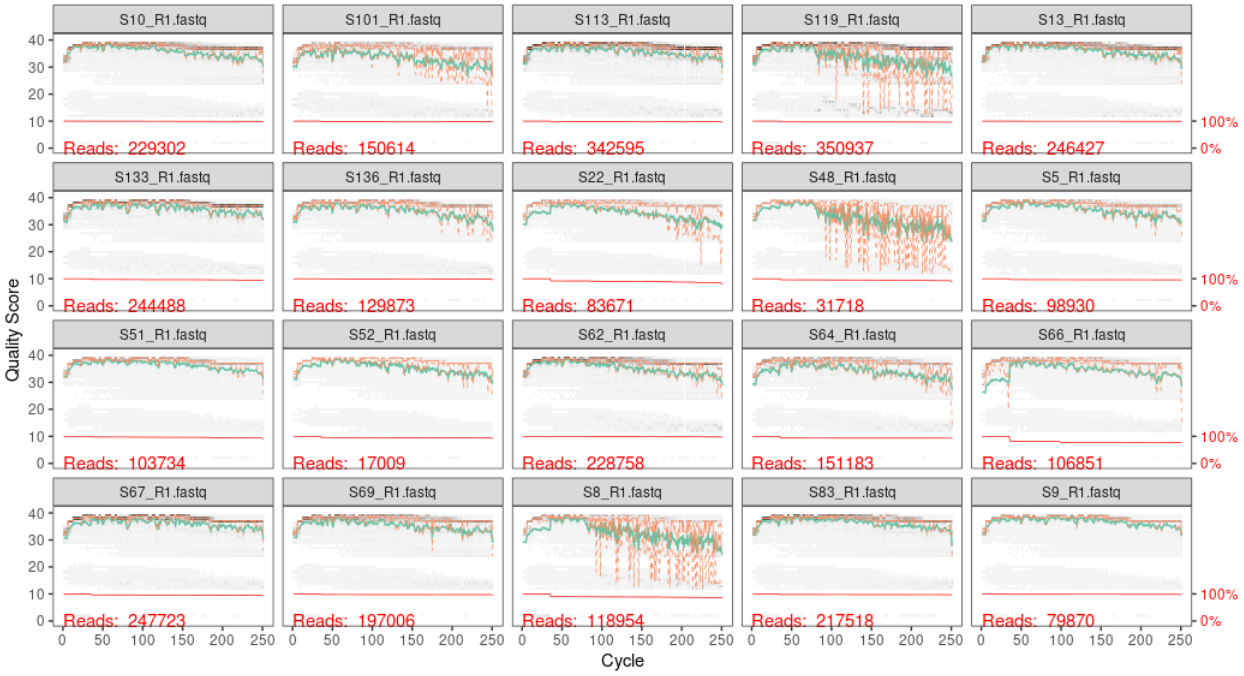

Figure 1: QualityProfileForward

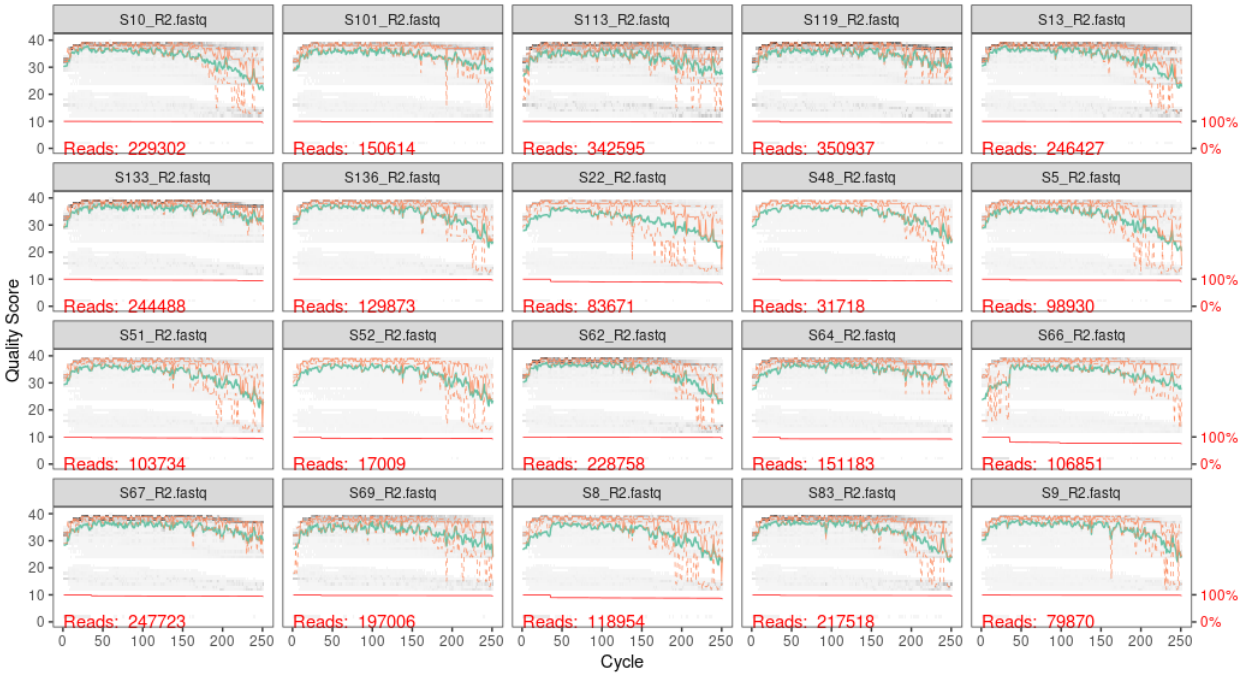

Figure 2: RawQualityProfileReverse

significantly at about position 200. Using the barcode and reverse primer metadata, we trimmed the first 25 and last 25 nucleotides to remain with the true reads. We set the maximum expect error at 3 for both forward and reverse reads.

```
Parameters
maxEE=c(3,3),
rm.phix=TRUE,
truncLen=c(230,200),
trimLeft = c(25,25),
multithread = TRUE
```

Approximately 11.6% of the reads were lost after trimming. Quality profiles of random samples were plotted, which confirmed a significant quality improvement; hence the reads proceeded to further downstream processing.

Figure 3. Quality Profiles for filtered forward reads

Figure 3: FilteredForwardPlot

Figure 4. Quality Profiles for filtered reverse reads

Learning Error Rates

DADA2 allows for error modeling using a machine-learning-based algorithm, which we utilized to establish sequencing error rates, including substitutions such as Single Nucleotide Polymorphisms. To verify that the error rates have been reasonably well-estimated, we inspected the fit between the observed error rates (black points) and the fitted error rates (black lines).

These figures show the frequencies of each type of transition as a function of the quality. Error rate plots revealed a decrease in error rates with an increase in sequence quality which was a satisfactory observation

Figure 4: ReverseFilteredPlot

that validated the estimated error rates; that is, the estimated error rate was similar to the observed error rate.

Figure 5. Error rate plot for forward reads

Figure 6. Error rate plot for reverse reads

Dereplication

Dereplication involves retrieving unique sequences from all the identical sequence reads, which reduces redundancy and computation time needed for analysis. New quality scores were assigned to the unique sequences, which is a functionality of the dereplication process.

Sample Inference

Sample inference was performed to obtain sequence variants from the dereplicated sequences using the core sample inference algorithm supported by DADA2. DADA2 provides two modes, "`pool=TRUE`"and "`pool=FALSE`". "`pool=TRUE`"improves the detection of rare variants observed just once or twice in an individual sample but often across all samples. However, it is a very computationally taxing step and can become intractable for datasets of tens of millions of reads. If a study does not need detection of rare variants then the Independent inference "`pool=FALSE`"is recommended. It has the advantage that computation time is linear in the number of samples, and memory requirements are flat with the number of samples. This allows scaling out to datasets of almost unlimited size. Therefore, we set the multithreading parameter to true since the process is computationally heavy.

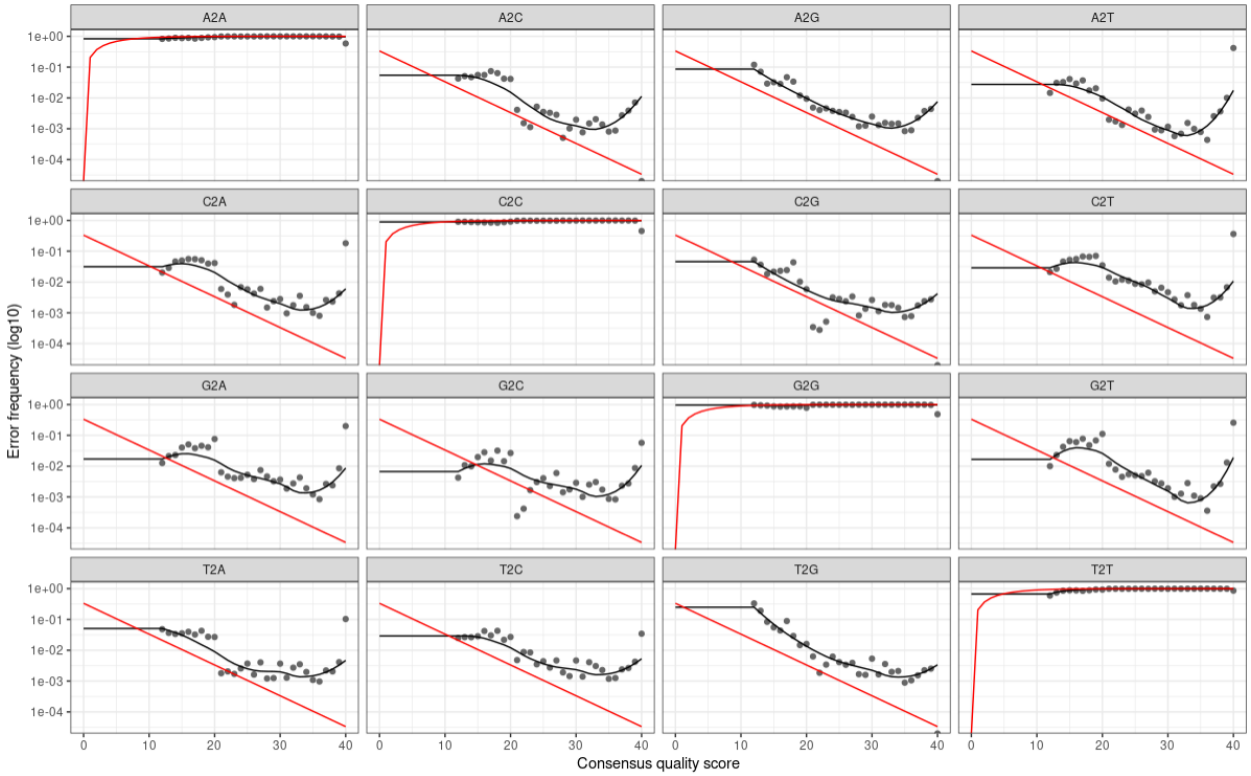

Figure 5: forward_error_plot

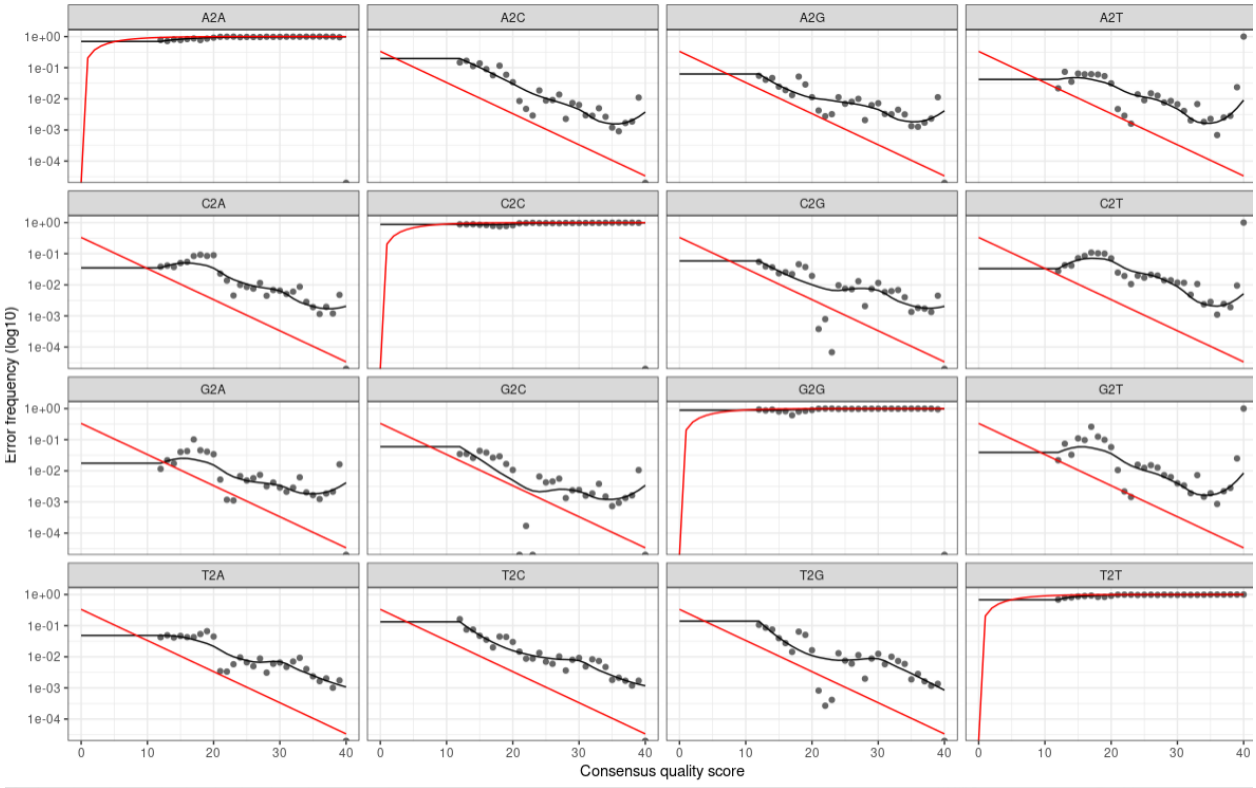

Figure 6: reverse_error_plot

Merging

Merging of the forward and reverse paired reads was carried out using the default `minOverlap` of 20 and setting the `trimOverhang` parameter to true since we did not trim overhangs earlier in the pipeline. We chose the parameters to facilitate optimal merging without a decrease in quality. From there, we observed that most of the reads merged, with only 1.83% of the reads not merged.

Constructing sequence table

The sequence table is a sample by sequence feature table, which is valued by the number of times each sequence was observed in each sample. From the sequence table, we observed 3879 ASVs. The majority of the merged sequences had similar lengths, although there was a significant change in some samples.

Removing chimeras

Chimeric sequences are identified if they can be exactly reconstructed by combining a left-segment and a right-segment from two more abundant "parent" sequences. We used the `removeBimeraDenovo` function, where sequence variants identified as bimeric are removed and return a bimera free collection of unique sequences. We set multithreading true to minimize on time taken and optimize the compute resources.

After removing the chimeras, 95.8% of the reads were retained. Chimera detection identified 7928 bimeras out of 11807 input sequences, therefore retaining 3879 ASVs.

Tracking reads through the pipeline

A mean of 79.68% of the reads was retained across all the processing steps of the pipeline.

Table 1. Summary table for reads tracking

	Input	Filtered	dada_forward	dada_reverse	Merged	Non chimera	final_perc_reads_retained
S1	108494	87977	84153	82263	80386	79551	73.3
S10	229302	212582	209823	207974	203574	197078	85.9
S100	126661	101181	99153	98396	91306	87519	69.1
S101	150614	132925	131426	131246	129713	126862	84.2
S102	146772	133672	132374	132079	130082	124447	84.8
S103	366244	340616	338735	338147	334604	312962	85.5
S104	339681	308417	305406	305055	298815	283637	83.5
S105	176537	150667	147410	145776	143371	140069	79.3
S106	409963	381879	380950	379950	373275	328478	80.1
S107	162844	123467	118933	116446	114732	114449	70.3
S110	461750	401964	392765	388100	381198	372196	80.6
S112	49460	43272	42736	42495	42024	40839	82.6
S113	342595	312178	307734	306039	301222	292414	85.4
S114	331129	304547	302856	302420	297833	280947	84.8
S115	109200	96798	95981	95738	94561	91153	83.5
S117	254884	235364	234703	234264	230560	206942	81.2
S119	350937	307307	306048	304755	300706	292361	83.3
S122	353419	323560	319317	318550	315258	313813	88.8
S123	225713	203143	200784	200230	197486	189080	83.8
S124	140088	125925	125216	125087	123542	118573	84.6

	Input	Filtered	dada_forward	dada_reverse	Merged	Non chimera	final_perc_reads_retained
S125	105919	84419	83345	83220	81814	80575	76.1
S126	380866	341856	338601	336647	331971	318407	83.6
S127	536496	487825	485620	485347	479980	460847	85.9
S128	210601	191488	189766	189357	187537	181359	86.1
S129	90546	76435	74881	74325	72968	71634	79.1
S13	246427	230047	229335	229078	226521	214413	87
S130	387131	353136	351304	350105	343847	323098	83.5
S131	217614	193390	190225	188520	186153	183682	84.4
S132	133395	109974	108459	108261	106615	104278	78.2
S133	244488	219885	218874	217946	215083	199598	81.6
S134	141947	112473	110765	110255	108974	107780	75.9
S135	303167	279799	278117	277540	274165	258633	85.3
S136	129873	119730	117891	117516	115797	114458	88.1
S138	110492	97111	96271	96017	94765	88886	80.4
S139	361408	334807	333354	333017	329740	298380	82.6
S14	215017	195747	193561	190793	188288	180629	84
S140	33079	24939	23391	22976	22446	21905	66.2
S141	161665	145047	143941	143502	141759	136722	84.6
S142	45924	40471	39167	38875	38213	37739	82.2
S144	372434	340876	339322	338790	334300	317234	85.2
S145	106428	93940	91888	91582	89546	87550	82.3
S15	207191	186555	184435	182166	178427	169667	81.9
S16	51951	45907	45120	44887	43985	42953	82.7
S17	226827	210349	209141	208010	204829	192657	84.9
S18	70709	54401	51122	49599	47850	46216	65.4
S19	190605	175335	174396	174215	172294	162886	85.5
S2	258746	233259	231586	229612	224111	208690	80.7
S20	78533	68818	67211	65813	64454	61709	78.6
S21	14155	12624	12256	12167	11831	11714	82.8
S22	83671	61199	57786	54533	52377	51471	61.5
S23	165500	138095	135665	133949	132263	130274	78.7
S24	104447	88550	86464	85584	84507	82958	79.4
S25	154810	139356	138037	137353	135267	130278	84.2
S27	69708	58645	55895	54213	52854	50849	72.9
S29	92226	83184	81792	81397	79792	78497	85.1
S3	29696	25983	25625	25472	25212	24905	83.9
S31	98386	92007	91707	91634	90792	86067	87.5
S32	160676	140773	137571	132308	129687	123581	76.9
S33	30434	26654	25490	25180	24654	24298	79.8
S34	231852	216706	216187	215856	213233	201141	86.8
S36	130881	122889	122612	122478	121265	115042	87.9
S38	155810	123597	119565	115482	112909	107959	69.3
S39	114609	99138	97085	94708	92819	86848	75.8
S40	102954	88259	86095	85418	83391	81917	79.6
S41	80777	69026	66708	65674	64355	61590	76.2
S42	161516	151018	150758	150606	149465	140224	86.8
S43	26710	24707	24314	24031	23753	23297	87.2
S44	89349	74266	72337	71051	69551	66138	74
S46	96397	78264	74135	72574	70555	68345	70.9
S47	188461	175068	173888	173232	171127	164719	87.4
S48	31718	26722	25485	25089	24668	24621	77.6
S49	121644	107021	104233	101289	98840	93308	76.7

	Input	Filtered	dada_forward	dada_reverse	Merged	Non chimera	final_perc_reads_retained
S5	98930	88018	87412	87159	86028	81316	82.2
S51	103734	90436	87618	85805	83688	78840	76
S52	17009	14712	14497	14346	14175	13701	80.6
S54	193042	182405	181816	181584	179740	169673	87.9
S55	183286	161893	157615	154337	151230	144831	79
S56	133249	111665	108594	107411	106055	103369	77.6
S57	229468	197421	195521	194530	192139	189676	82.7
S58	107133	98625	97946	97695	96564	92318	86.2
S59	203605	191888	191243	190994	188681	177743	87.3
S6	211514	199936	199409	199216	196872	188397	89.1
S60	138892	127837	127007	126674	125153	120442	86.7
S61	172758	159699	158677	158305	156435	149695	86.7
S62	228758	210207	207312	206306	203149	195533	85.5
S63	131697	103216	102022	101739	100883	97638	74.1
S64	151183	129644	128995	128924	127695	124539	82.4
S65	228293	202697	200846	200518	198387	191977	84.1
S66	106851	74125	73152	72554	70826	66795	62.5
S67	247723	222604	218792	216798	214402	211504	85.4
S68	70210	64757	64370	64144	63427	61282	87.3
S69	197006	179837	178471	177704	175155	159872	81.2
S7	117051	102597	100988	99314	96158	94705	80.9
S70	248320	217931	213805	212790	208912	204465	82.3
S71	167183	148736	147634	147123	144847	140248	83.9
S72	160242	131726	125085	123445	119058	114677	71.6
S73	90548	80779	79982	79823	78936	76815	84.8
S74	156178	119274	117189	116495	114730	112282	71.9
S75	207264	181524	178274	177179	174315	168514	81.3
S76	30263	17206	16351	16195	15779	15650	51.7
S77	117901	99016	97272	96178	94219	90881	77.1
S78	231741	207803	204514	203448	198465	192950	83.3
S79	140625	118789	118190	118131	116887	113580	80.8
S8	118954	91396	88504	87103	85653	84526	71.1
S80	96400	39171	38549	38066	37523	37050	38.4
S81	46643	26045	22335	21233	19962	19843	42.5
S82	312725	278569	277548	277101	273461	264009	84.4
S83	217518	201386	199600	198788	196526	187857	86.4
S84	270208	247874	243758	242763	240048	238061	88.1
S85	342197	301317	296809	295302	286931	268323	78.4
S86	80376	59108	57086	56517	54978	53291	66.3
S87	244014	222004	217205	215329	211941	207377	85
S89	176154	156779	154622	154171	152066	147645	83.8
S9	79870	74727	74153	73963	73483	72441	90.7
S90	66713	49895	47491	46753	45625	45403	68.1
S91	218723	191772	190949	190284	188505	183670	84
S92	74331	56050	53897	53405	52459	52151	70.2
S93	268130	233773	231517	231059	228755	219959	82
S94	93202	50389	49881	49763	49318	48734	52.3
S95	31443	16995	16301	16144	15762	15672	49.8
S96	227880	206676	205726	205373	203753	195309	85.7
S97	317252	290181	287580	286673	283081	270532	85.3
S98	396734	358134	355994	355400	350853	328345	82.8
S99	193089	175185	174442	174214	172224	165513	85.7

	Input	Filtered	dada_forward	dada_reverse	Merged	Non chimera	final_perc_reads_retained
Mean	174890.0483870968	170745.4838709677	170261.5161290323	169897.7983870968	170368.9032258065	169809.7983870968	79.6822580645161
	Input	Filtered	dada_forward	dada_reverse	Merged	Non chimera	final_perc_reads_retained

Assigning Taxonomy

In this step, the input sequences to be classified are from the sequence table without chimeras, while the training set of reference sequences with known taxonomy used was from the silva database, and taxonomy was assigned up to the species level. An alternative training set from the RDP database was used but was found to have more NAs than silva; hence silva was chosen for downstream analysis. Taxonomy was assigned utilizing a `minBootstrap` confidence of 50, which is the default parameter for the DADA2 algorithm.

Table 2. Taxonomic assignments of the top 50 ASVs

	Kingdom	Phylum	Class	Order	Family	Genus	Species
ASV1	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	iners
ASV2	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	NA
ASV3	Bacteria	Firmicutes	Negativicutes	Veillonellales-Selenomonadales	Veillonellaceae	Megasphaera	NA
ASV4	Bacteria	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Shuttleworthia	NA
ASV5	Bacteria	Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Gardnerella	vaginalis
ASV6	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	amnii
ASV7	Bacteria	Fusobacteriota	Fusobacteriia	Fusobacteriales	Leptotrichiaceae	Sneathia	NA
ASV8	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	NA
ASV9	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	crispatus
ASV10	Bacteria	Fusobacteriota	Fusobacteriia	Fusobacteriales	Leptotrichiaceae	Sneathia	sanguinegens
ASV11	Bacteria	Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Gardnerella	vaginalis
ASV12	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	NA
ASV13	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	timonensis
ASV14	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	amnii
ASV15	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	NA
ASV16	Bacteria	Firmicutes	Negativicutes	Veillonellales-Selenomonadales	Veillonellaceae	Dialister	NA
ASV17	Bacteria	Actinobacteriota	Coriobacteriia	Coriobacteriales	Atopobiaceae	Atopobium	vaginae
ASV18	Bacteria	Firmicutes	Clostridia	Oscillospirales	Hungateiclostridiaceae	Fastidiosipila	NA
ASV19	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella_7	melaninogenica
ASV20	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	NA
ASV21	Bacteria	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Family_XI	Finegoldia	magna
ASV22	Bacteria	Firmicutes	Negativicutes	Veillonellales-Selenomonadales	Veillonellaceae	Megasphaera	NA
ASV23	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	NA
ASV24	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	disiens
ASV25	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	NA
ASV26	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	crispatus
ASV27	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella_7	NA
ASV28	Bacteria	Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Gardnerella	vaginalis
ASV29	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	bivia
ASV30	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	NA
ASV31	Bacteria	Fusobacteriota	Fusobacteriia	Fusobacteriales	Leptotrichiaceae	Sneathia	NA

	Kingdom	Phylum	Class	Order	Family	Genus	Species
ASV32	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Escherichia-Shigella	NA
ASV33	Bacteria	Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Gardnerella	vaginalis
ASV34	Bacteria	Fusobacteriota	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	nucleatum
ASV35	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	NA
ASV36	Bacteria	Patescibacteria	Saccharimonadia	Saccharimonadales	NA	NA	NA
ASV37	Bacteria	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Shuttleworthia	NA
ASV38	Bacteria	Firmicutes	Negativicutes	Veillonellales-Selenomonadales	Veillonellaceae	Veillonella	montpellierensis
ASV39	Bacteria	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Family_XI	Parvimonas	NA
ASV40	Bacteria	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Family_XI	Fenollaria	NA
ASV41	Bacteria	Firmicutes	Negativicutes	Veillonellales-Selenomonadales	Veillonellaceae	Dialister	NA
ASV42	Bacteria	Actinobacteriota	Actinobacteria	Corynebacteriales	Corynebacteriaceae	Corynebacterium	glucuronolyticum
ASV43	Bacteria	Actinobacteriota	Coriobacteriia	Coriobacteriales	Eggerthellaceae	DNF00809	NA
ASV44	Bacteria	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	Aerococcus	christensenii
ASV45	Bacteria	Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Gardnerella	NA
ASV46	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	NA
ASV47	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas	uenonis
ASV48	Bacteria	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Family_XI	Peptoniphilus	NA
ASV49	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	corporis
ASV50	Bacteria	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Shuttleworthia	NA

Taxonomy rank statistics

Rank	Total assigned	% of ASVs assigned	No. of unique without NA
Kingdom	3724	96.0	3
Phylum	2829	72.93	18
Class	2629	67.78	29
Order	2572	66.31	63
Family	2397	61.8	116
Genus	1579	40.71	305
Species	320	8.25	205

From the table, we observe that 40.71% and 8.25 ASVs were assigned to genus and species level, respectively. 51% could only be assigned to rank higher than genus.

Phylogeny

Phylogenetic relatedness is commonly used to inform downstream analyses, especially calculating phylogeny-aware distances between microbial communities. The DADA2 sequence inference method is reference-free, so we constructed the phylogenetic tree relating the inferred sequence variants de novo.

Using the DECIPHER R package, we carried out Phylogenetic analysis by firstly performing multiple sequence alignment, after which a distance matrix was assigned for phylogenetic tree construction. We used the `phangorn` R package with the Neighbor-Joining algorithm as our clustering method for phylogenetic

inference. The Generalized Time Reversible Model (GTR) was used as the substitution model, and stochastic rearrangement was set, which allowed for random permutation in the phylogenetic tree.

Here are the parameters:

```
Parameters used
maxEE=c(3,3),
m.phix=TRUE,
truncLen=c(230,200),
trimLeft = c(25,25),
multithread = TRUE
```

Alpha diversity

Alpha diversity is the average species diversity in our samples generated from summary metrics that describe individual samples.

Richness and diversity estimates

Plotting was done using Chao1 richness estimates and Shannon diversity values. Chao1 is a richness estimator, "richness" being the total number of distinct ASVs in the samples, while Shannon's diversity index is a metric of diversity. The term diversity includes "richness" (the total number of your distinct units) and "evenness" (the relative proportions of all of your distinct units). We used the `phyloseq` package, and specifically, the "`plot_richness()`"function.

Figure 7. Richness Barplot for top 30 ASVS by abundance by age using inflammation as fill as BV as facet wrap

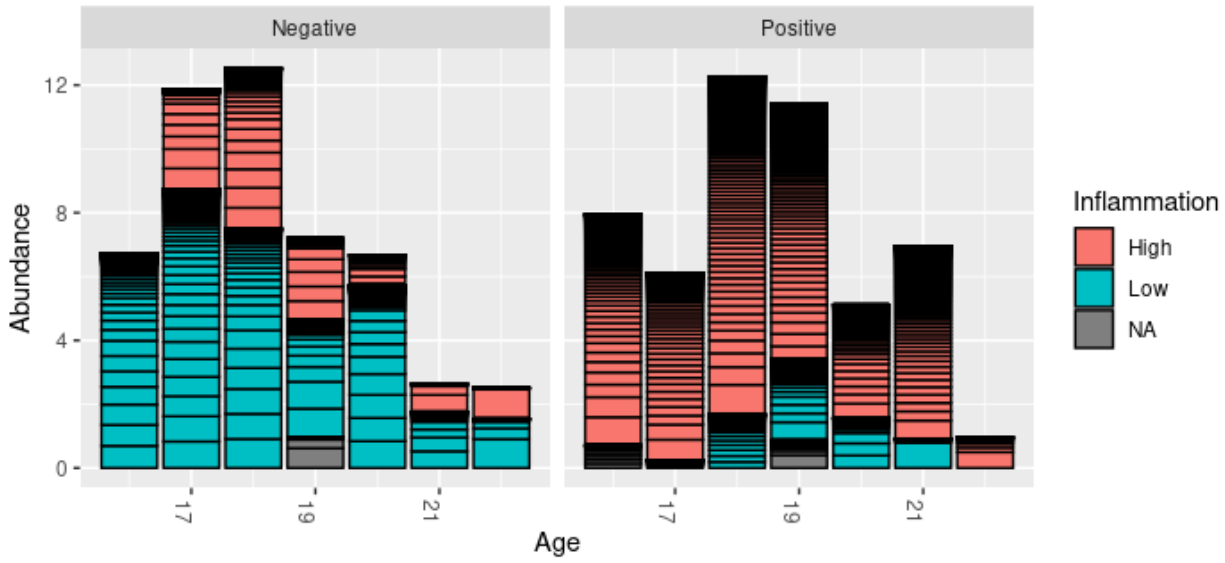

Figure 7: barplot

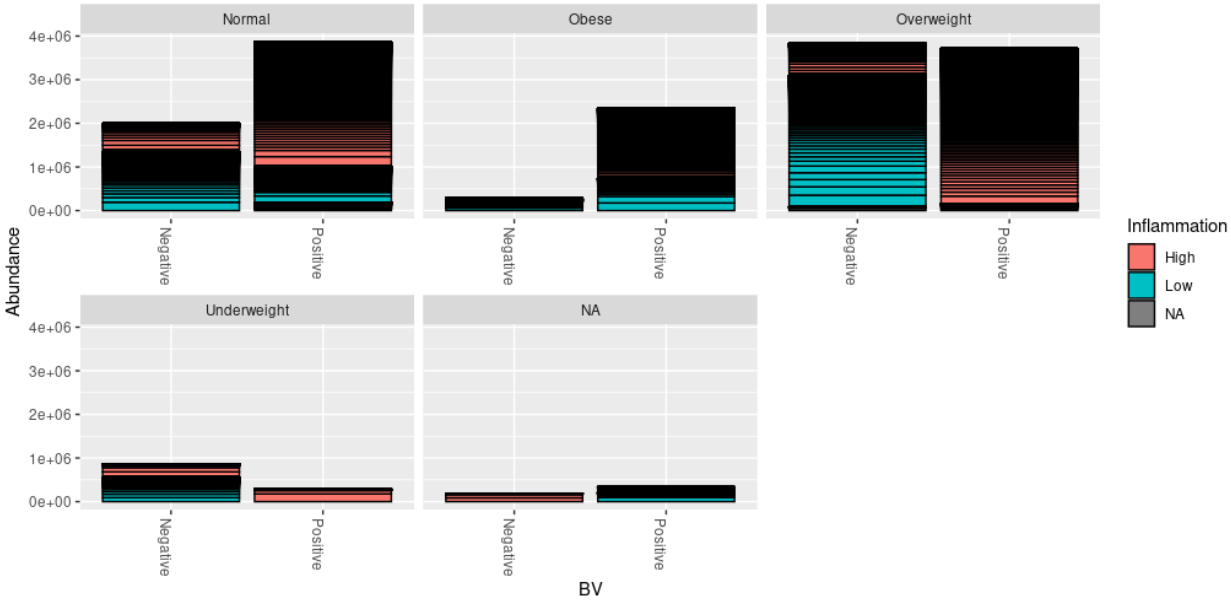

Figure 8: bv_bmi

Figure 8 Richness Barplot by abundance by BV using inflammation as fill and status i.e. categories of BMI as facet wrap

Figure 9 Richness Barplot by abundance by BV using BMI as fill and inflammation as facet wrap

Figure 10. Richness Boxplot

The plots' median values and interquartile ranges show a significant difference between Positive and negative BV samples (p-value <0.05). Negative BV samples showed greater CHAO/ACE values; this leads to an expected higher species richness of the microbiota. The Simpson index and the Shannon index showed a higher diversity of the microbiota in BV positive samples.

Beta diversity

Principal Coordinates Analysis (PCoA) was plotted to offer multidimensional scaling that operates on dissimilarities or distances. We used the created phyloseq object to generate the PCoA plot since it is very convenient for displaying beta diversity among samples.

Figure 11. PCoA plot for beta diversity visualization

There was clustering observed. This is due to a positive correlation between high inflammation and positive BV.

Rarefaction analysis

Rarefaction analysis revealed that the majority of rarefaction curves flattened. However, there are about six troublesome samples with very low sequencing depth.

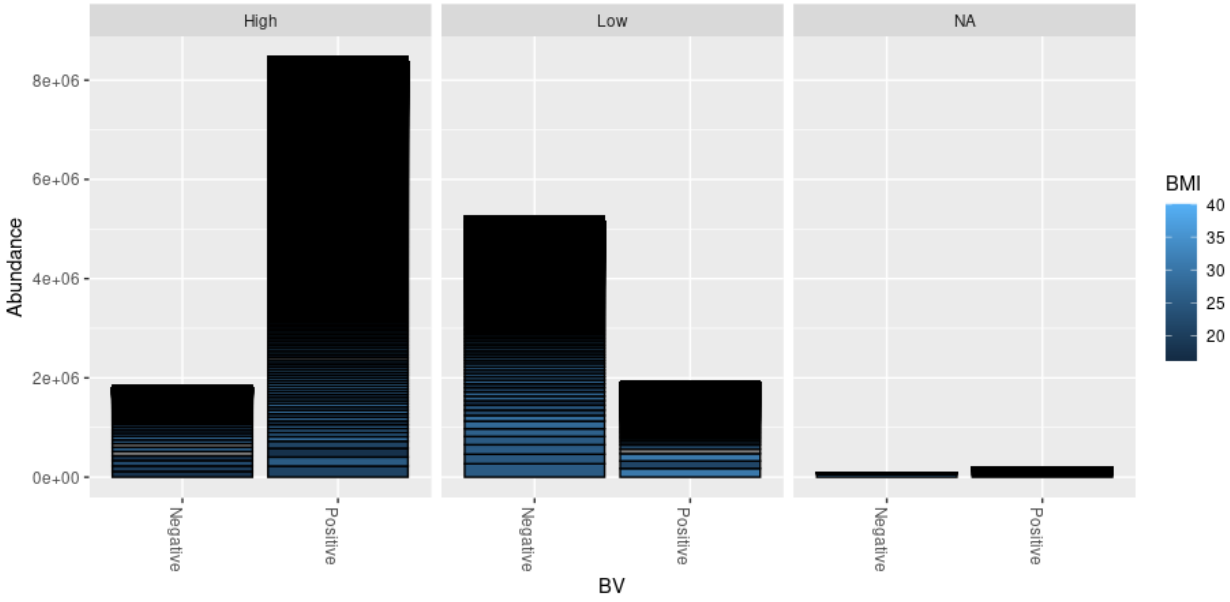

Figure 9: bmi_inflammation

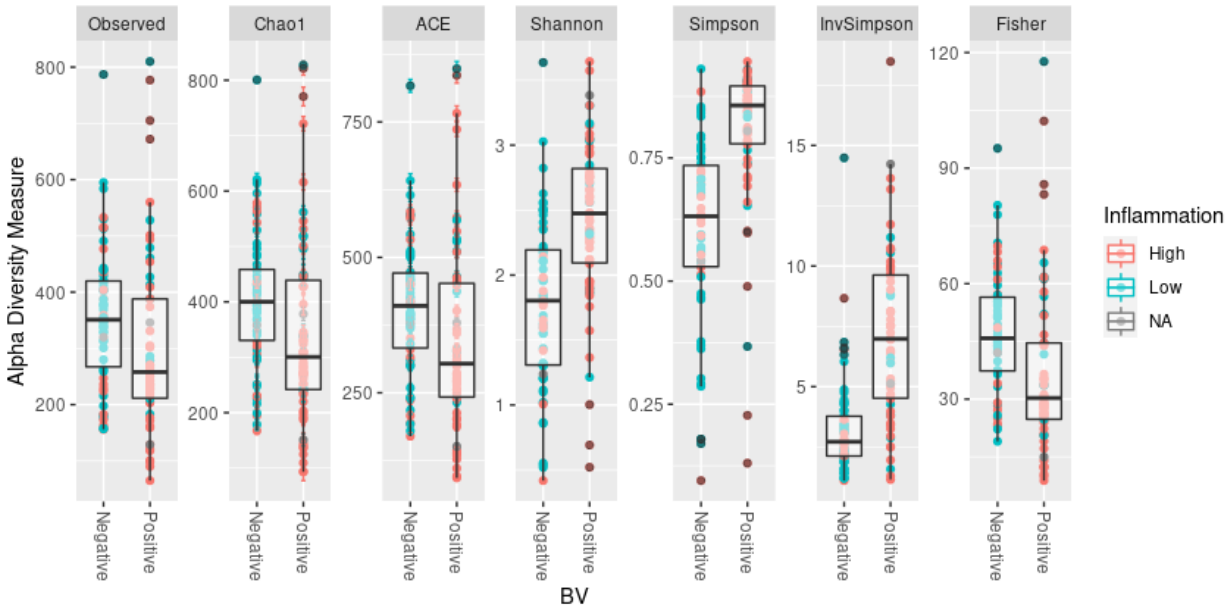

Figure 10: alpha-diversity

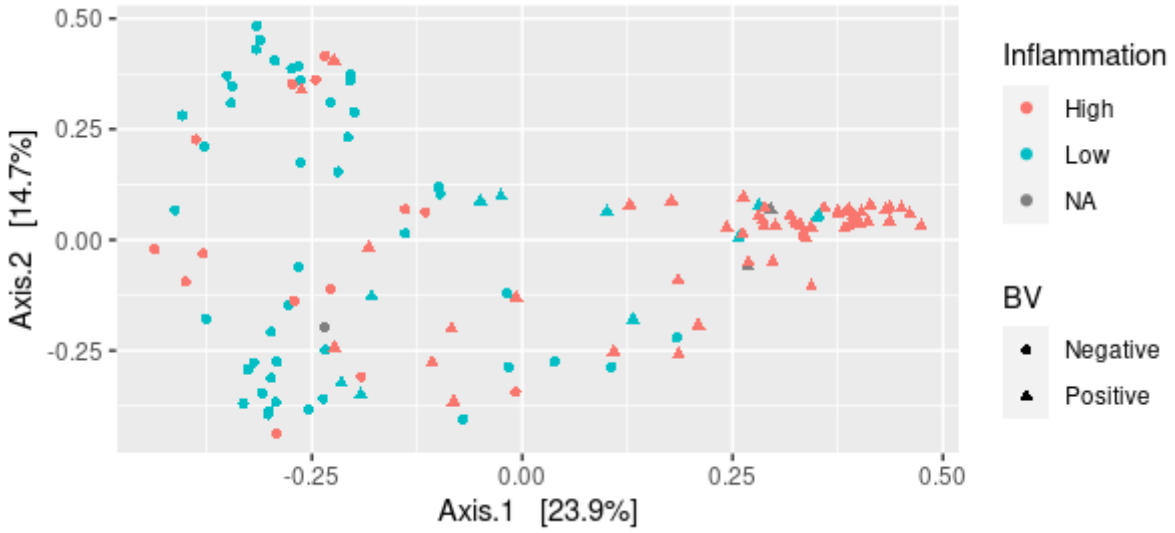

Figure 11: PCoa plot

Figure 12. Rarefaction curves

Taxanomic diversity

Figure 13. Phylum abundance in relation to BV status across the samples

The dominant phylum across the different samples regardless of the BV status is Fusobacteriota, while Bacteroidota is the second most dominant in BV positive samples.

Figure 14. Genus abundance in relation to BV status across the samples

There are different dominant genus according to the BV status, as shown above. Notably, lactobacillus is significantly reduced in the BV negative samples, in agreement with the literature.

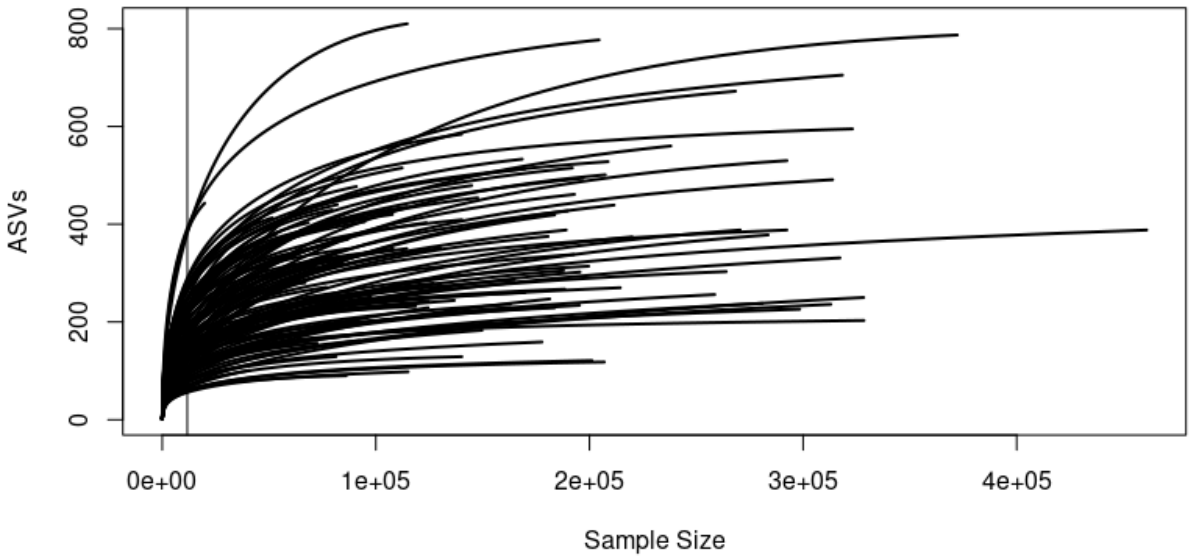

Figure 12: Rarefaction curve

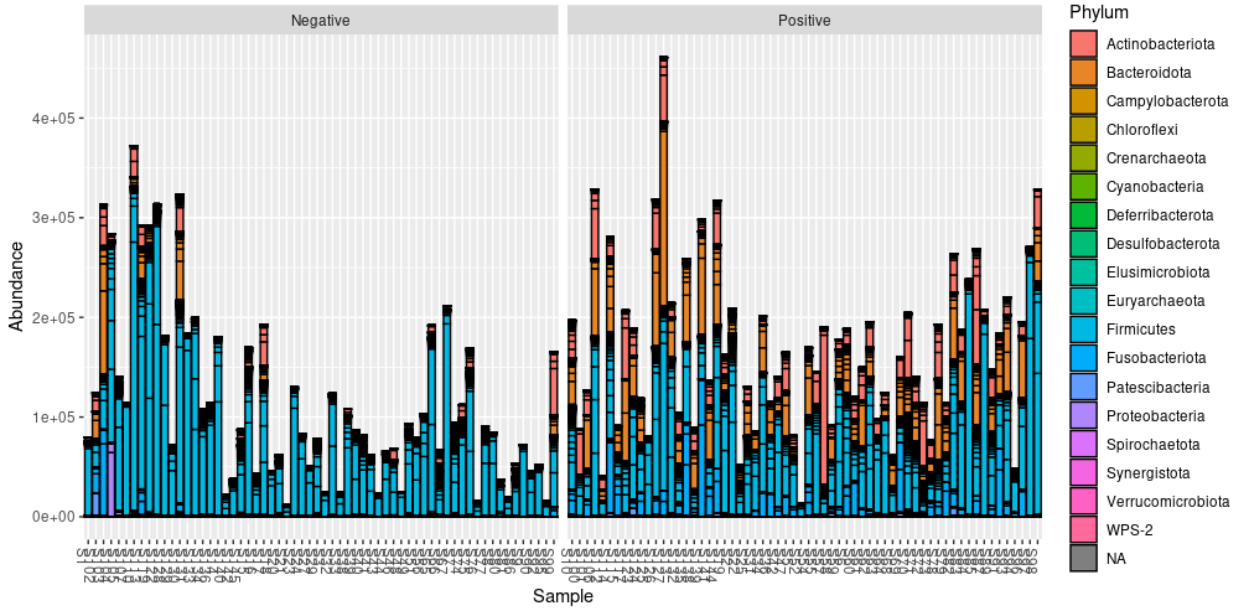

Figure 13: Phylum nice

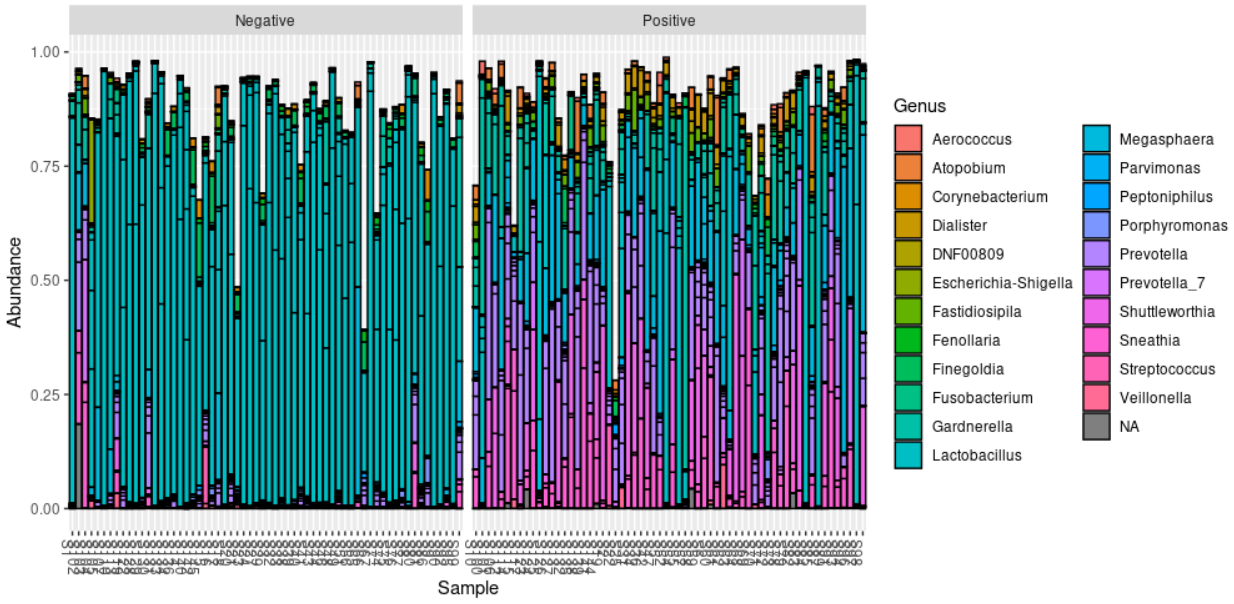

Figure 14: Genus