

20180419 Report on the Canine transplant 16S MiSeq data analysis (preliminary data)

(1) Primer sequences were trimmed from the clean reads using cutadapt

Ref: Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17(1), pp-10.

(2) Sequences were processed using the DADA2 R package

Ref: Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods*, 13(7), 581.

Ref: Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME journal*, 11(12), 2639.

“DADA2 is a software package that models and corrects Illumina-sequenced amplicon errors. DADA2 infers sample sequences exactly, without coarse-graining into OTUs, and resolves differences of as little as one nucleotide....DADA2 implements a new quality-aware model of Illumina amplicon errors. Sample composition is inferred by dividing amplicon reads into partitions consistent with the error model. The full workflow involves: filtering, error learning and modeling, dereplication, merging paired-end reads, and chimera identification”

Data stats after DADA2 pipeline:

SampleID	RawReads	clean	filtered	denoised	merged	nonchim	%output
MT-A-1	50265	35144	34247	34247	24877	20685	41.2%
MT-A-2	50968	35500	34708	34708	25757	20851	40.9%
MT-A-3	54617	35000	34061	34061	27890	24538	44.9%
MT-A-4	49865	34805	33956	33956	25122	20178	40.5%
MT-B-1	49589	34992	34160	34160	25570	21456	43.3%
MT-B-2	53257	35478	34450	34450	28033	24079	45.2%
MT-B-3	55239	35060	34231	34231	27703	23592	42.7%
MT-B-4	49250	34646	33925	33925	27474	23103	46.9%
MT-C-1	53289	34817	34269	34269	26331	20717	38.9%
MT-C-2	60002	34883	34239	34239	29144	25740	42.9%
MT-C-3	57681	34992	34470	34470	28566	23427	40.6%
MT-C-4	51699	34849	34314	34314	26374	21686	41.9%
MT-D-1	51900	35000	34393	34393	26761	21670	41.8%
MT-D-2	51878	35009	34387	34387	26129	20253	39.0%
MT-D-3	54702	34617	33961	33961	26869	20929	38.3%
MT-D-4	47834	34408	33849	33849	25489	20566	43.0%
MT-E-I-1	51331	34772	33987	33987	25426	20800	40.5%
MT-E-II-1	54934	34848	34011	34011	27058	19876	36.2%
MT-E-III-1	51587	34949	34087	34087	26424	19000	36.8%
MT-E-IV-1	51141	34941	33998	33998	25444	19443	38.0%
MT-E-V-1	56223	35034	34177	34177	27120	19545	34.8%
MT-E-V-2	59157	34975	34252	34252	29908	22116	37.4%
MT-F-1	53046	34633	34083	34083	27705	23819	44.9%
MT-F-2	52271	34933	34305	34305	26367	19841	38.0%

MT-F-3	52542	35009	34373	34373	26276	20763	39.5%
MT-F-4	51180	35169	34549	34549	26088	20615	40.3%
MT-G-1	51628	35246	34628	34628	26390	21473	41.6%
MT-G-2	52810	34549	33977	33977	25882	19090	36.1%
MT-G-3	49457	34389	33817	33817	25868	19570	39.6%
MT-G-4	55495	34808	34175	34175	27205	22848	41.2%
MT-H-1	52976	35244	34612	34612	26429	21923	41.4%
MT-H-2	52791	35304	34571	34571	26487	21527	40.8%
MT-H-3	52368	34638	34047	34047	26388	20480	39.1%
MT-H-4	51536	35280	34631	34631	25986	22714	44.1%
MT-I-1	52996	35289	34608	34608	26597	21711	41.0%
MT-I-2	55780	34808	34115	34115	26696	20330	36.4%
MT-I-3	49221	34732	34293	34293	26238	21201	43.1%
MT-I-4	51675	35017	34359	34359	26118	22772	44.1%
MT-J-1	54336	35135	34479	34479	27661	22780	41.9%
MT-J-2	53817	35060	34351	34351	27544	21665	40.3%
MT-J-3	56563	34899	34138	34138	29498	25406	44.9%
MT-J-4	55658	35560	34805	34805	28791	24203	43.5%
MT-K-1	54890	35518	34645	34645	27321	23256	42.4%
MT-K-2	54327	35051	34362	34362	26543	22164	40.8%
MT-K-3	51341	34807	34169	34169	27723	23614	46.0%
MT-K-4	53320	35026	34277	34277	26154	21370	40.1%
MT-L-1	50316	34672	33911	33911	25401	19924	39.6%
MT-L-2	51657	35102	34263	34263	26928	21030	40.7%
MT-L-3	58197	35043	34052	34052	29715	27059	46.5%
MT-L-4	53352	35552	34848	34848	27182	21792	40.8%
MT-M-1	52059	35035	34049	34049	25566	20515	39.4%
MT-M-2	53180	34589	33788	33788	26763	20618	38.8%
MT-M-3	50304	34984	34350	34350	27257	21687	43.1%
MT-M-4	51884	34580	33777	33777	25292	19853	38.3%
MT-N-1	52130	35457	34721	34721	25555	20643	39.6%
MT-N-2	55280	35432	34608	34608	27423	19635	35.5%
MT-N-3	52272	34831	33986	33986	26424	19408	37.1%
MT-N-4	56101	35630	34656	34656	28763	24902	44.4%
MT-O-1	53803	35605	34687	34687	26806	22185	41.2%
MT-O-2	55477	35085	34204	34204	26594	19698	35.5%
MT-O-3	51406	34705	34086	34086	26953	20183	39.3%
MT-O-4	57804	30635	29707	29707	29442	29353	50.8%
MT-P-1	49664	34687	33925	33925	24797	21097	42.5%
MT-P-2	53456	35255	34454	34454	27137	20575	38.5%
MT-P-3	51636	35196	34306	34306	26042	19996	38.7%
MT-P-4	50803	35690	34867	34867	26050	20257	39.9%
MT-Q-1	49872	35359	34365	34365	25526	21759	43.6%
MT-Q-2	56068	35296	34422	34422	27882	24249	43.2%
MT-Q-3	49932	34747	34019	34019	27181	20859	41.8%

MT-Q-4	49970	34400	33805	33805	24521	19718	39.5%
MT-R-1	49775	34750	34182	34182	24948	19961	40.1%
MT-R-2	50983	34825	34290	34290	26408	20666	40.5%
MT-R-3	51757	35202	34644	34644	26760	20644	39.9%
MT-R-4	50754	35236	34708	34708	26132	20203	39.8%
MT-S-1	53333	34874	34275	34275	27725	25310	47.5%
MT-S-2	49249	34300	33739	33739	26851	20475	41.6%
MT-S-3	47863	34909	34372	34372	25541	20172	42.1%
MT-S-4	50925	34430	33799	33799	24804	19641	38.6%
Transplant-A	54312	34958	34077	34077	26218	19070	35.1%
Transplant-C	54635	34958	34087	34087	26529	19030	34.8%
Transplant-F	62548	34815	33942	33942	30944	29165	46.6%
Transplant-H	53950	34755	33992	33992	27057	18614	34.5%
Transplant-J	56797	34907	34068	34068	28091	21314	37.5%
Transplant-L	56037	35545	34629	34629	27779	21479	38.3%
Transplant-N	57660	35093	34325	34325	27844	20902	36.3%
Transplant-P	57327	34916	33943	33943	28430	20442	35.7%
Transplant-Q	56452	35457	34637	34637	28008	19784	35.0%

Stats on the number of reads in all the samples:

min	18614
max	29353
mean	21532
SD	2098

These 87 samples resulted in 1759 ASVs (Amplicon Sequence Variants), conceptually equivalent to OTUs defined at 100% identity.

Stats on the number of ASVs in all the samples:

min	139
max	439
mean	279
SD	55

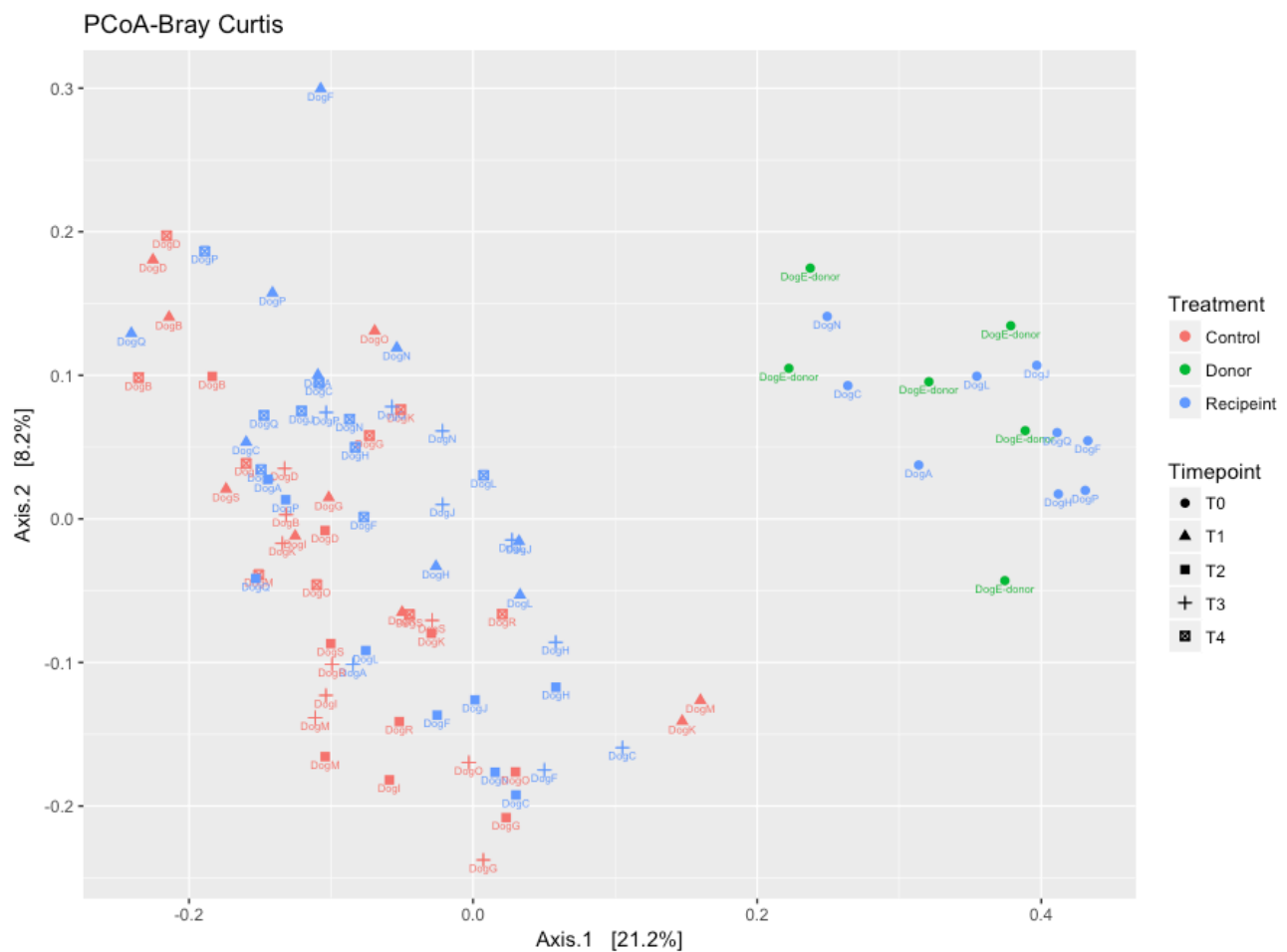


Figure 1 Principal Coordinates Analysis (PCoA) based on the Bray–Curtis dissimilarity measure of the ASV data; which takes into account the composition of taxa/ASVs in the samples.



Figure 2 PCoA based on the unweighted UniFrac measure of the ASV data; which takes into account the phylogenetic distance but not the abundances of taxa/ASVs.

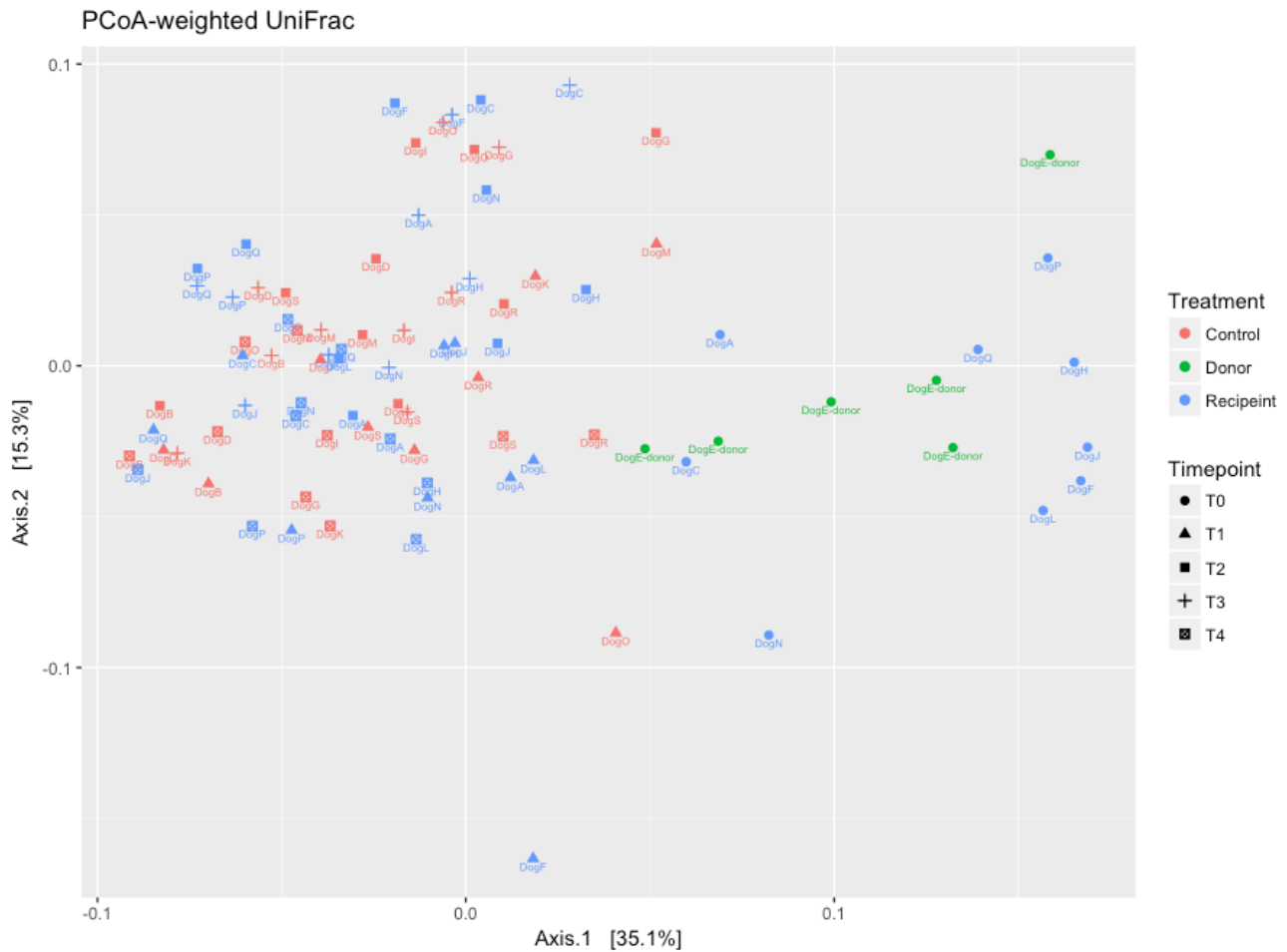


Figure 3 PcoA based on the weighted UniFrac measure of the ASV data; which takes into account the phylogenetic distance AND the abundances of taxa/ASVs.

Each dot in the ordination represents an individual sample. All ordination analyses based on three different metrics indicated higher similarities among the microbiota of the samples Transplant-A, C, F, H, J, L, N, P, Q and the donor dog E (MT-E-I-1, MT-E-II-1, MT-E-III-1, MT-E-IV-1, MT-E-V-1, MT-E-V-2), distinctive from the microbiota from transplant recipient dogs and control dogs A-S.



Figure 4 Number of observed ASVs in all samples

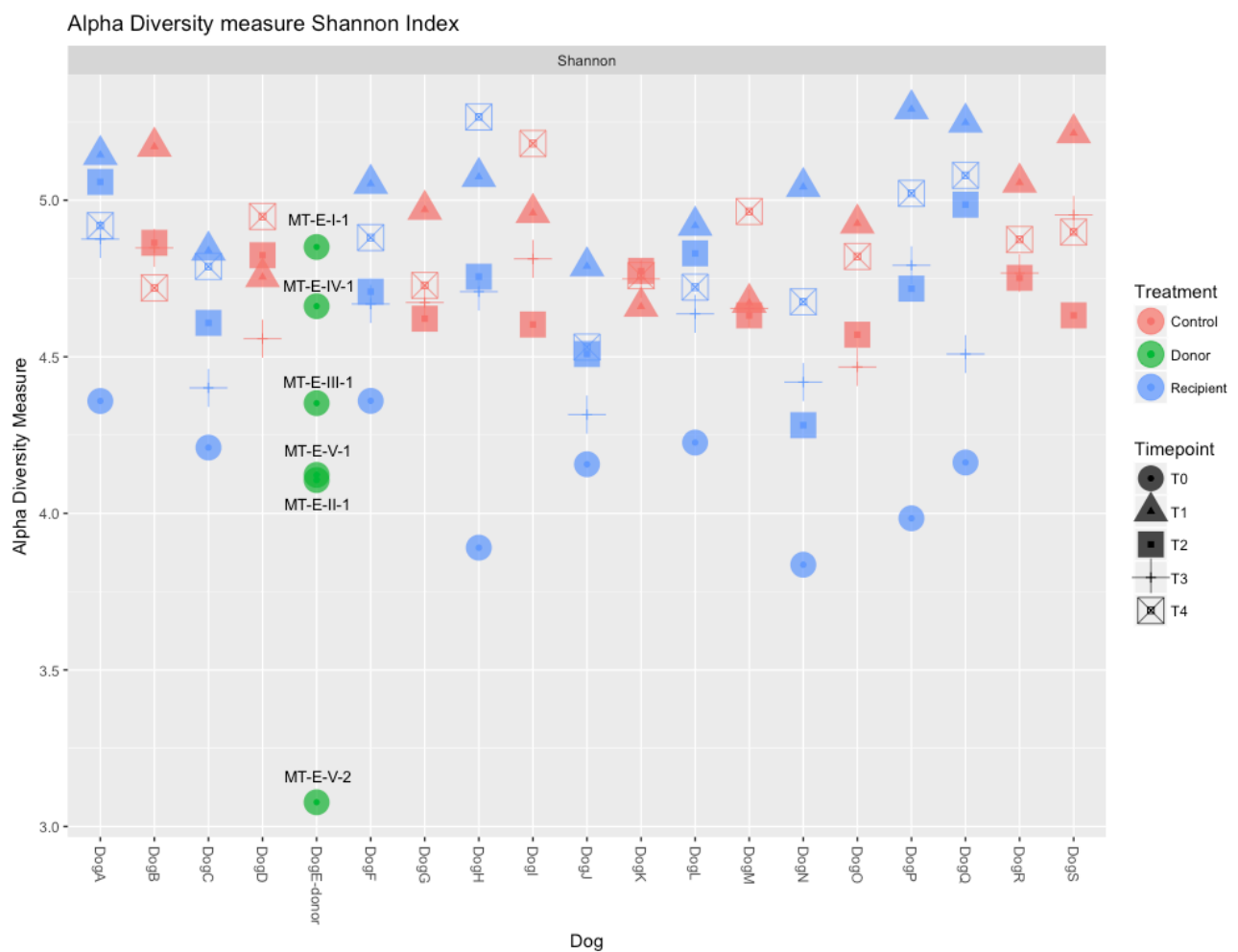


Figure 5 Shannon entropy (alpha diversity measure) in all samples (by individual Dogs)

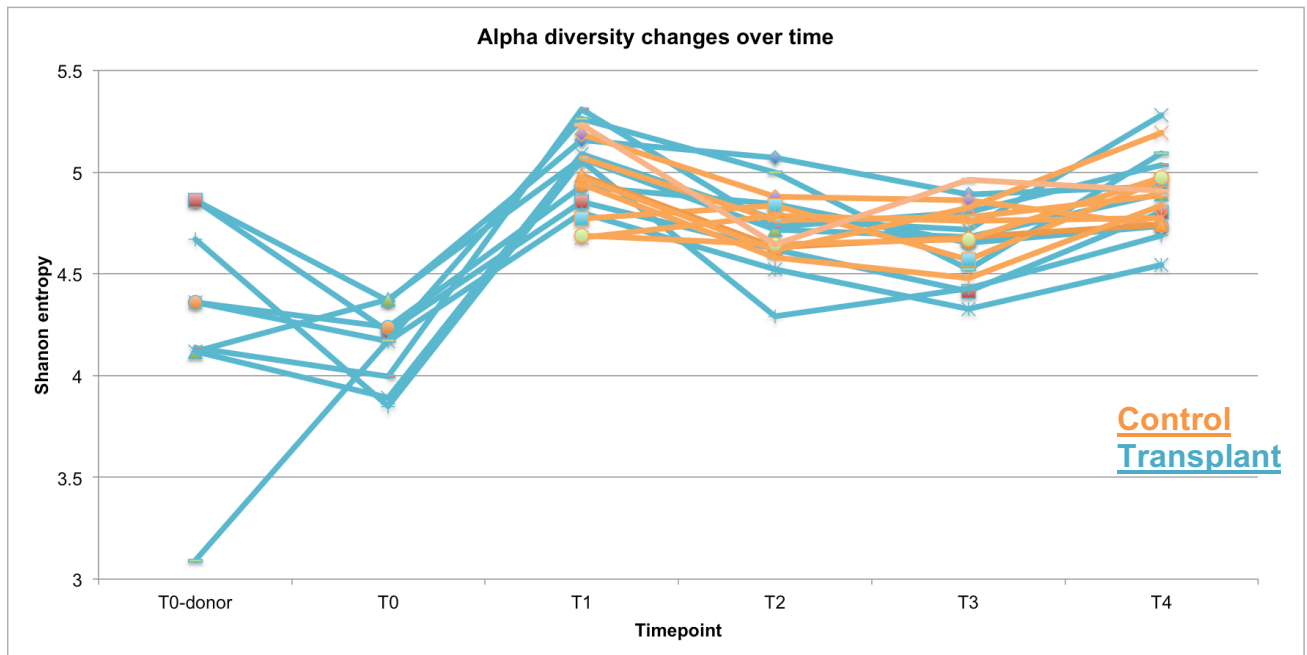


Figure 6 Shannon entropy in all samples (by Time-points)

The microbiota from donor dog E (MT-E-I-1, MT-E-II-1, MT-E-III-1, MT-E-IV-1, MT-E-V-1) except MT-E-V-2 have higher diversity than the transplant recipient dogs (Transplant-A, C, F, H, J, L, N, P, Q) initially. All recipient dogs have increased diversity in their microbiota after transplant, to equivalent levels compared to the controls at all time-points (T1-T4).

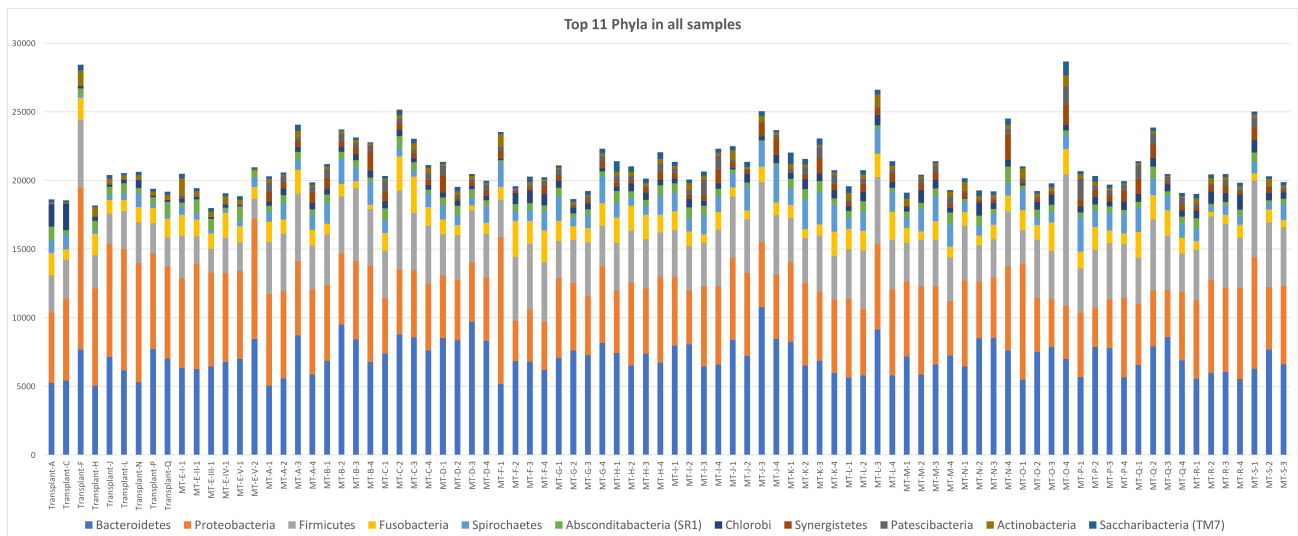


Figure 7 Bar-plots of the top 11 phyla detected in all samples (absolute read counts)

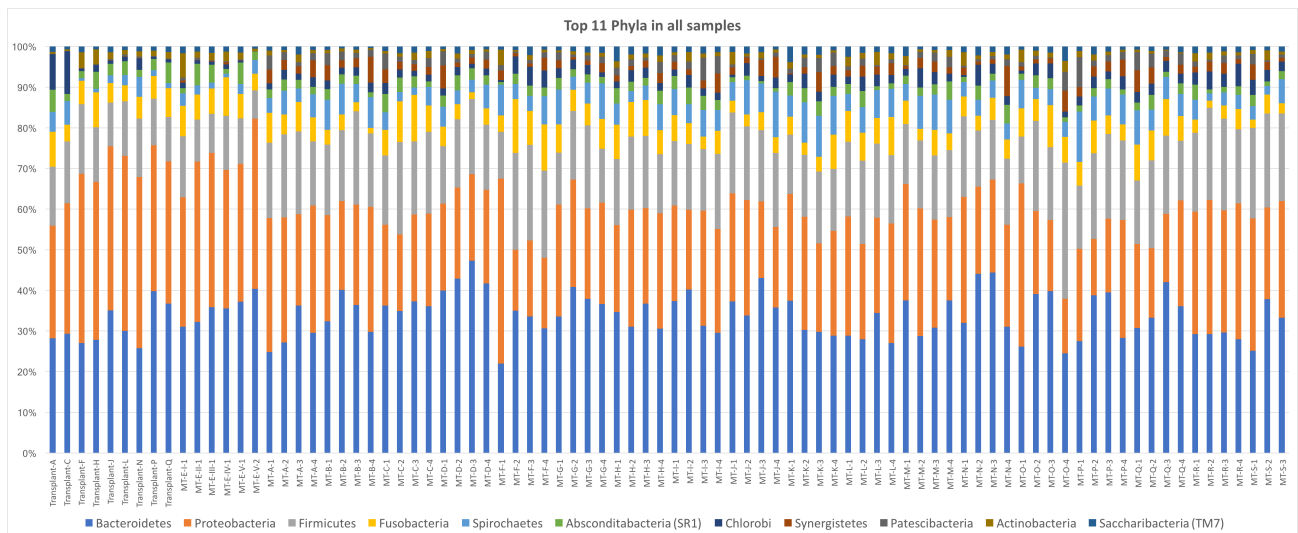


Figure 8 Bar-plots of the top 11 phyla detected in all samples (relative abundances)

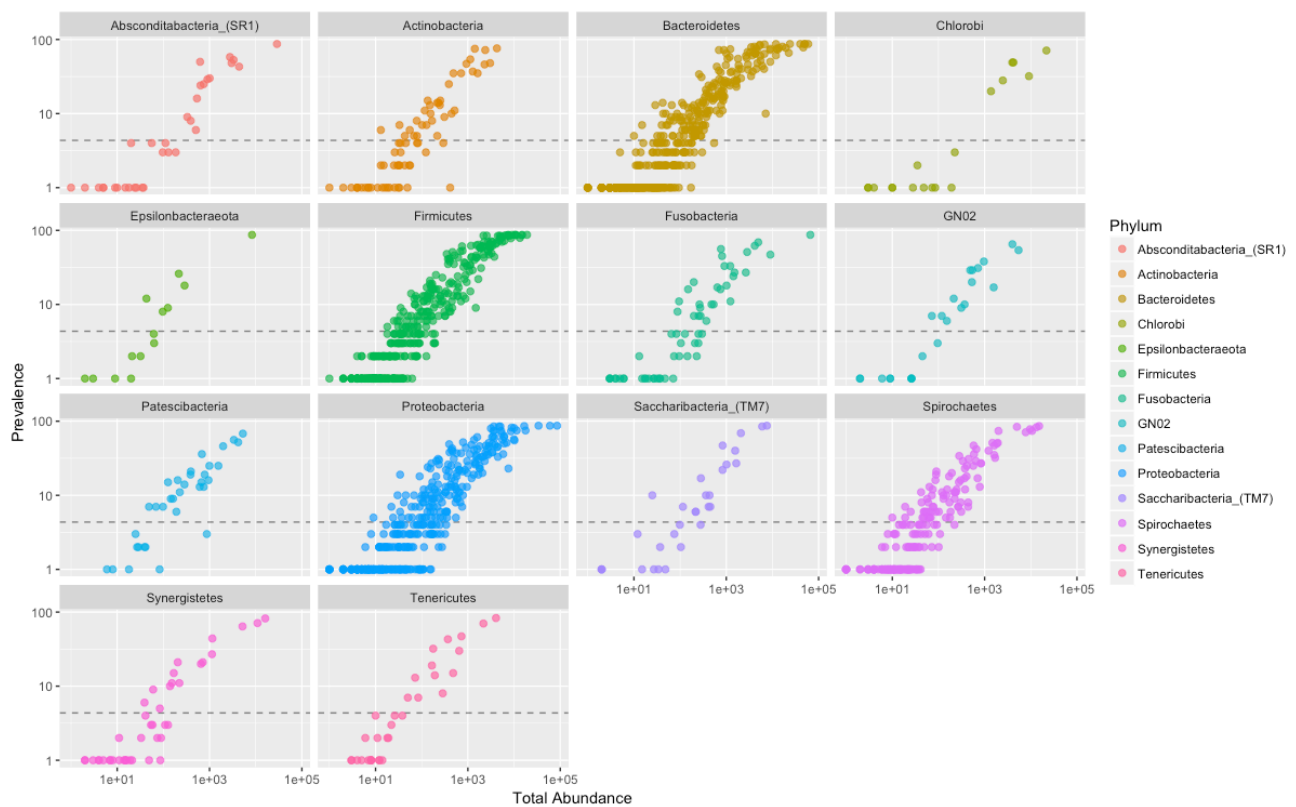


Figure 9 taxa prevalence vs total counts. Further filtering of the dataset could be performed based on the taxa prevalence (detection frequency) and the abundance. This plot indicates the filtering of taxa based on a prevalence threshold at 5% of the total samples (4.35). Each dots represent an ASV, the horizontal dotted line indicates the prevalence cutoff; this will filtered the data to 806 ASVs and removed the rare taxa from the following phyla (prevalence indicated in brackets): Acidobacteria (1), Chloroflexi (2), Cyanobacteria (1), Elusimicrobia (3), Planctomycetes (2) and WPS-2 (1).