

## Canine oral microbiome transplantation study - Report v4

Prepared by Yuki Chan (Dec 28 2018; updated Jan 11 2019; Feb 14 2019)

### Summary about Samples

Count of samples	T1	T2	T3	T4	Total
<b>Control</b>	<b>9</b>	<b>9</b>	<b>9</b>	<b>9</b>	<b>36</b>
DogB	1	1	1	1	4
DogD	1	1	1	1	4
DogG	1	1	1	1	4
DogI	1	1	1	1	4
DogK <sup>^</sup>	1	1	1	1	4
DogM <sup>^</sup>	1	1	1	1	4
DogO	1	1	1	1	4
DogR	1	1	1	1	4
DogS	1	1	1	1	4
<b>Donor</b>	<b>5</b>	<b>9</b>	<b>0</b>	<b>0</b>	<b>14</b>
DogE	5	9	0	0	14
<b>Recipient</b>	<b>9</b>	<b>9</b>	<b>9</b>	<b>9</b>	<b>36</b>
DogA	1	1	1	1	4
DogC	1	1	1	1	4
DogF	1	1	1	1	4
DogH	1	1	1	1	4
DogJ <sup>^</sup>	1	1	1	1	4
DogL <sup>^</sup>	1	1	1	1	4
DogN	1	1	1	1	4
DogP	1	1	1	1	4
DogQ	1	1	1	1	4
<b>Grand Total</b>	<b>23</b>	<b>27</b>	<b>18</b>	<b>18</b>	<b>86</b>

Donor samples From DogE	
T1 (baseline)	related to
MT-E-I-1	DogA,C
MT-E-II-1	DogF,H
MT-E-III-1	DogJ,L
MT-E-IV-1	DogN
MT-E-V-1	DogP, Q
MT-E-V-2 <sup>#</sup>	DogQ
Processed plaque From DogE	
T2 (transplantation)	transferred to
Transplant-A (MT-A)	DogA
Transplant-C (MT-C)	DogC
Transplant-F (MT-F)	DogF
Transplant-H (MT-H)	DogH
Transplant-J (MT-J)	DogJ
Transplant-L (MT-L)	DogL
Transplant-N (MT-N)	DogN
Transplant-P (MT-P)	DogP
Transplant-Q (MT-Q)	DogQ

<sup>^</sup>Dog L had been reported to had contact to dog J, K or M after transplantation

<sup>#</sup> This is a baseline sample that was taken at a wrong day, not included in the analysis

### Timepoints

**T1: Week 0 = baseline**

**T2: Week 2 = the day of transplantation**

**T3: Week 4 = 2 weeks after transplantation**

**T4: Week 14 = 12 weeks after transplantation**

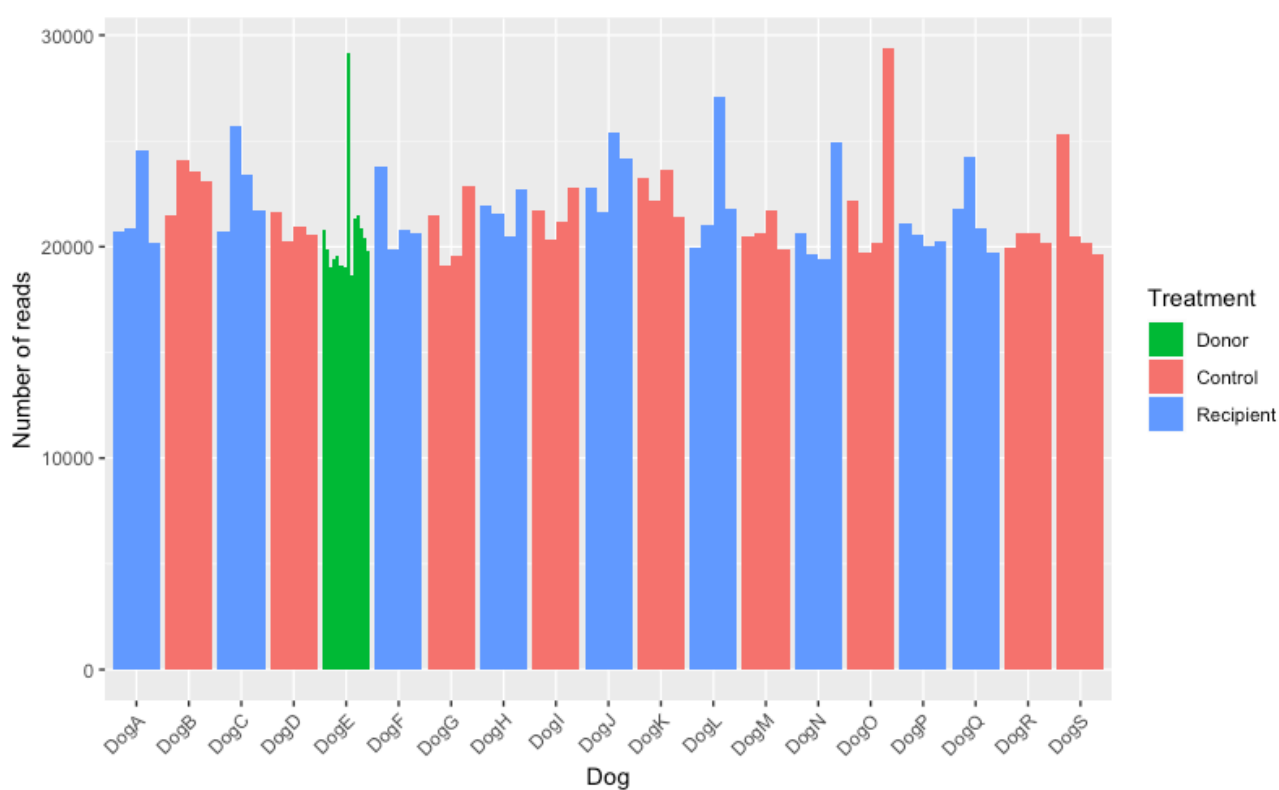
### Sequence data summary

The number of sequencing reads and the data statistics passed through the DADA2 pipeline for each sample was summarized in TableS1.

**How many sequencing reads are there?** (1 read  $\approx$  1 bacterial cell in theory, in reality there are multiple ribosomal RNA operons range from 1-16 copies<sup>1</sup>)

number of reads	All samples (n=86)	Donor (n=14)	Recipient (n=36)	Control (n=36)
min	18614	18614	19408	19090
max	29353	29165	27059	29353
mean	21525	20605	21846	21561

All samples carry equivalent numbers of sequences.

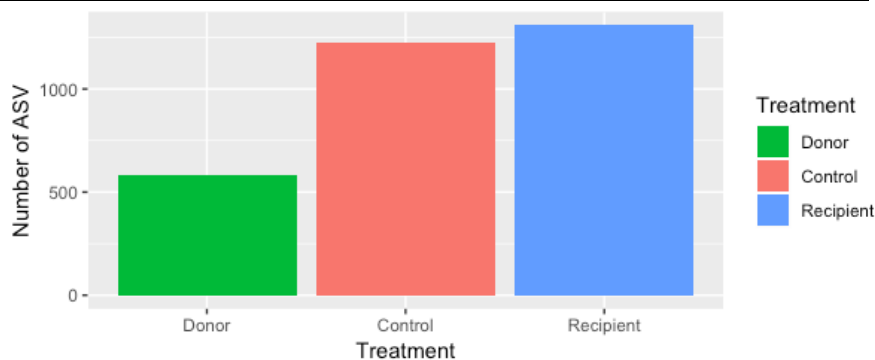


<sup>1</sup> Espejo RT, Plaza N. Multiple Ribosomal RNA Operons in Bacteria; Their Concerted Evolution and Potential Consequences on the Rate of Evolution of Their 16S rRNA. *Front Microbiol.* 2018 Jun 8;9:1232. doi: 10.3389/fmicb.2018.01232. PMID: 29937760; PMCID: PMC6002687.

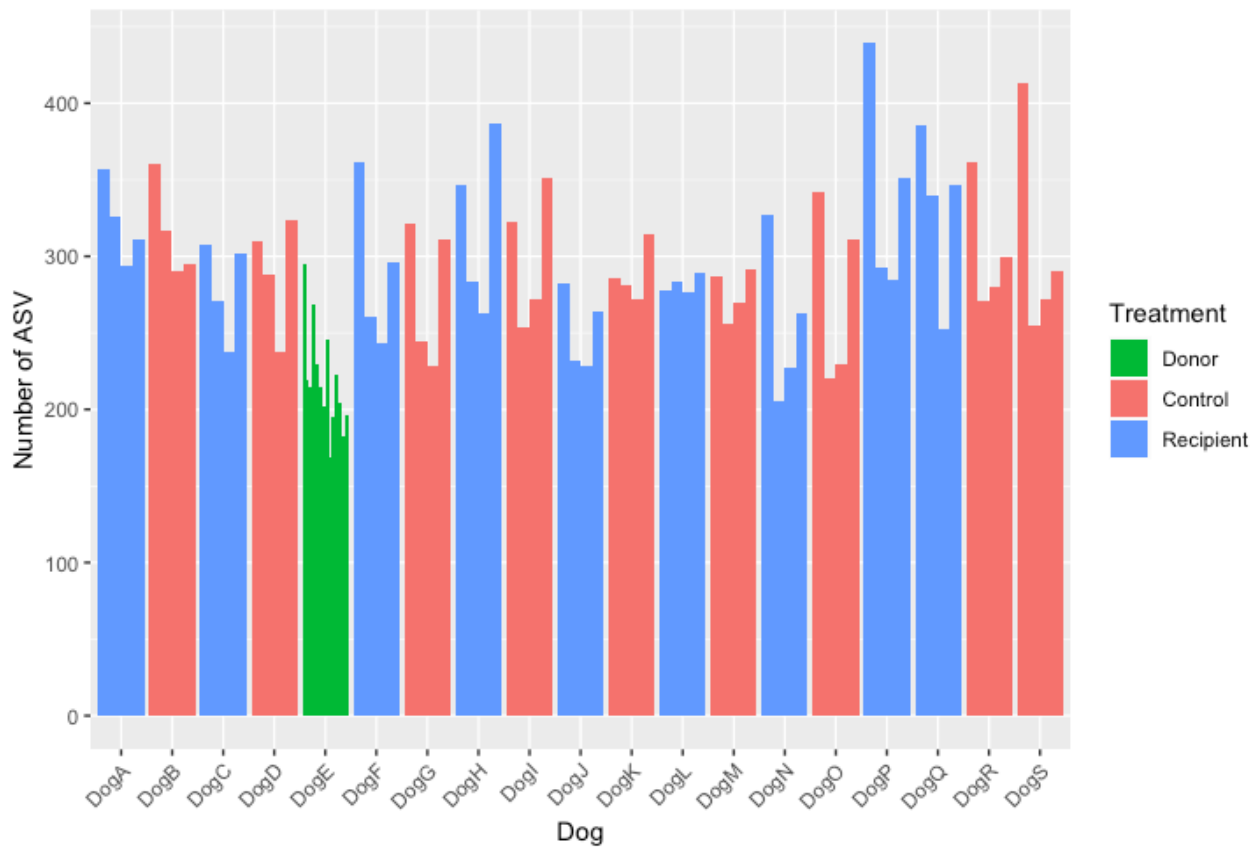
In these 86 samples there were **1753 ASVs (Amplicon Sequence Variants)**, conceptually equivalent to operational taxonomic units (OTUs) defined at 100% sequence identity; which can be interpreted as individual bacterial species/strains.

**How many species/strains of bacteria are there in the samples?** (number of ASVs)

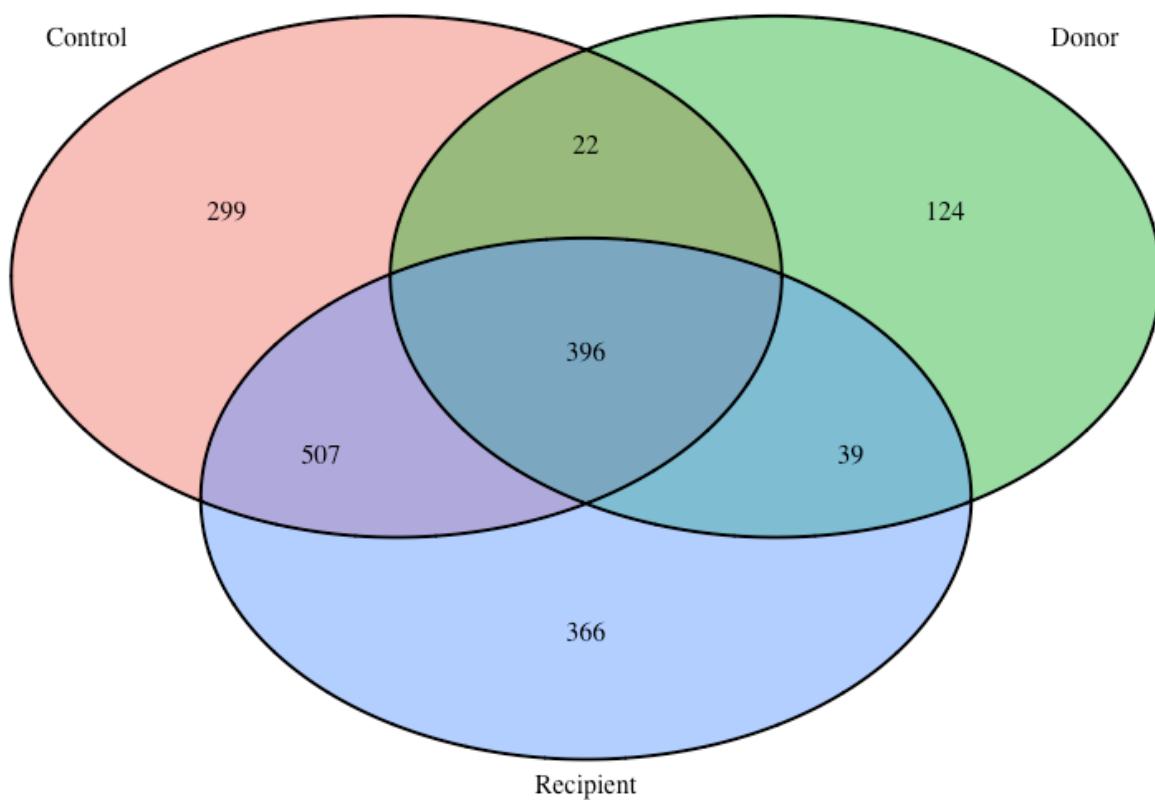
	All samples (n=86)	Donor (n=14)	Recipient (n=36)	Control (n=36)
min	169	169	205	220
max	439	295	439	413
mean	282	218.5	297.1	292.4
Total	1753	581	1308	1224



The samples from the donor dog had the lowest number of bacterial strains (ASVs).

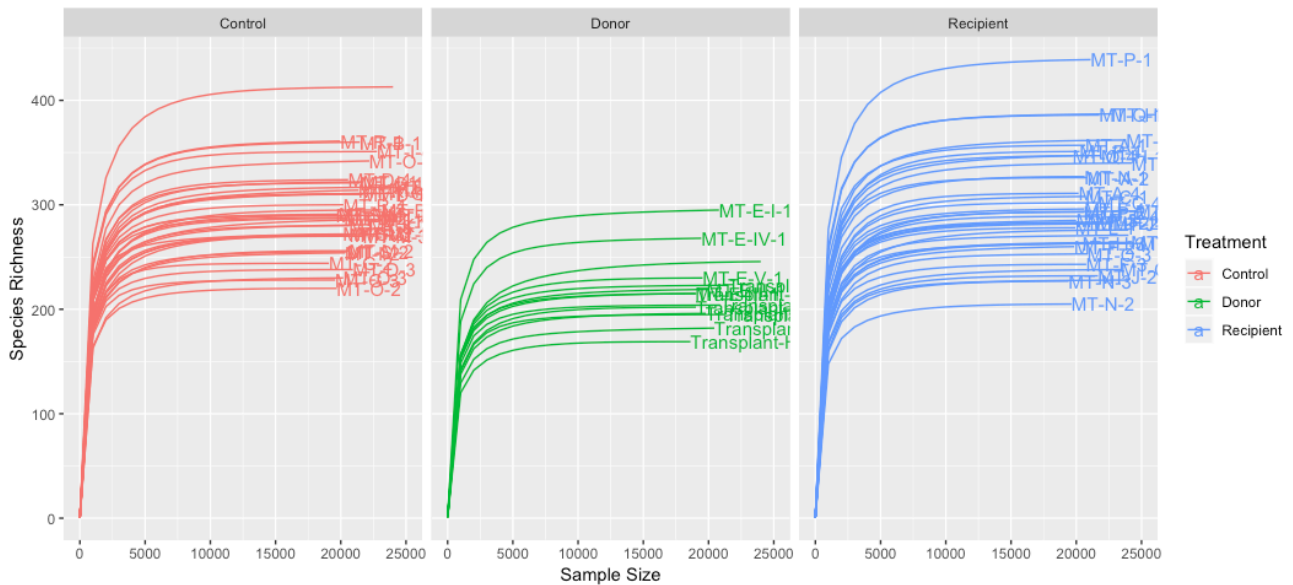


**How many ASVs are found in common among the samples from the donor, control and recipient dogs?**



**Is the sampling depth enough? i.e. did we recover most of the bacteria from the samples?**

**> Yes**

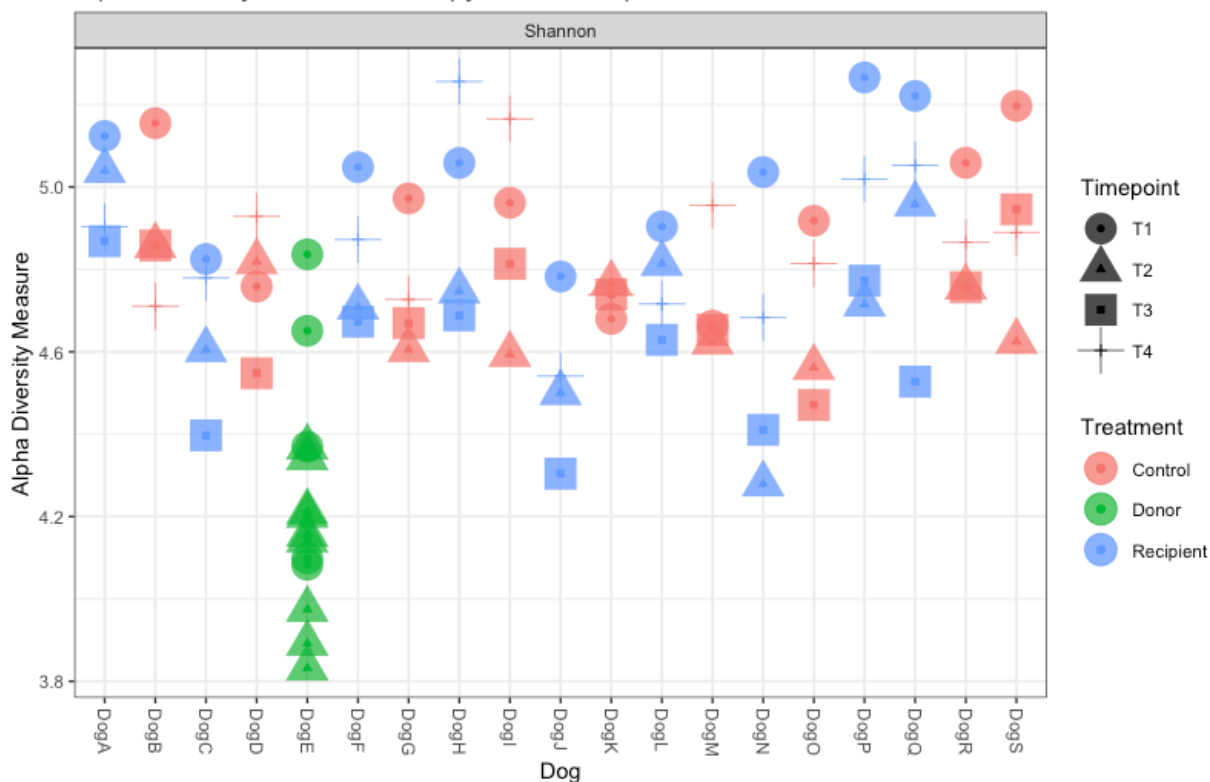


Rarefaction curves = plot of the number of ASVs (representing the species richness) as a function of the number of sequencing reads in the sample. These curves for all the samples reached plateau indicating high coverage of the microbiota by sequencing.

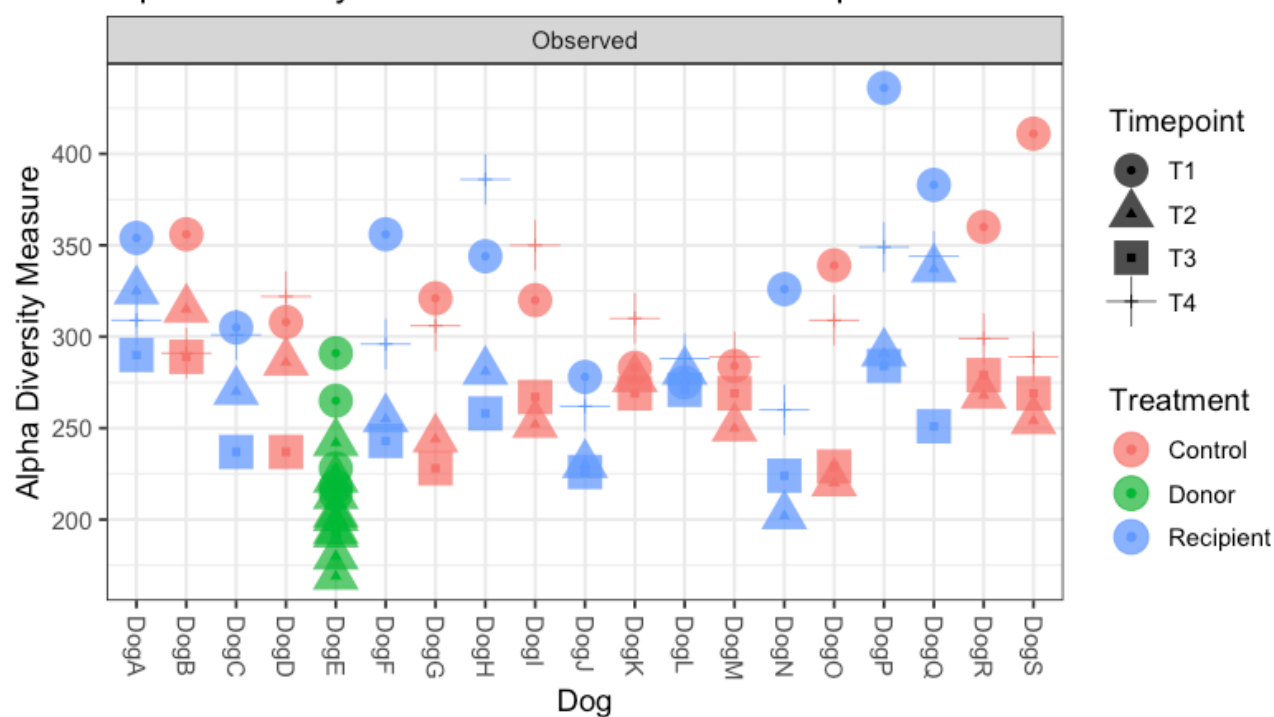
## Comparison of bacterial diversities in the microbiomes

Shannon Entropy (or Shannon diversity index) measures the species richness and evenness in the bacterial communities in the sample. Observed ASVs are the count of ASVs in the samples

Alpha Diversity: Shannon Entropy across samples

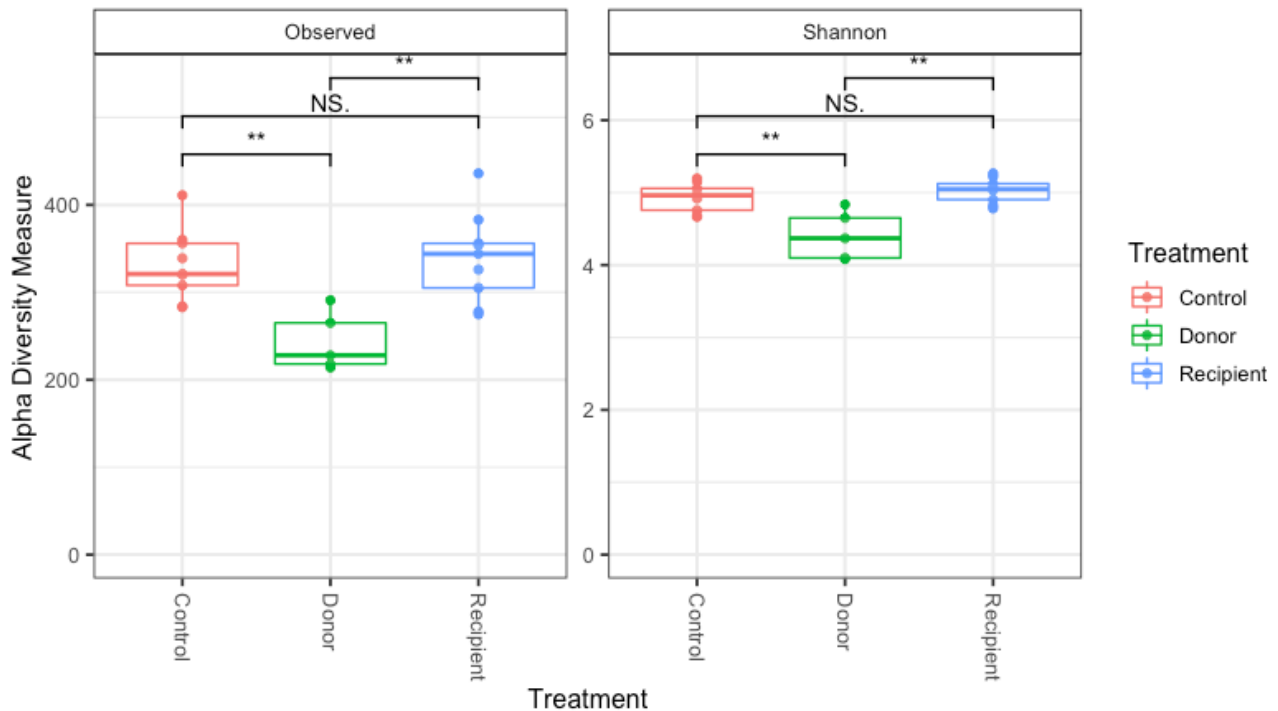


Alpha Diversity: Observed ASVs across samples

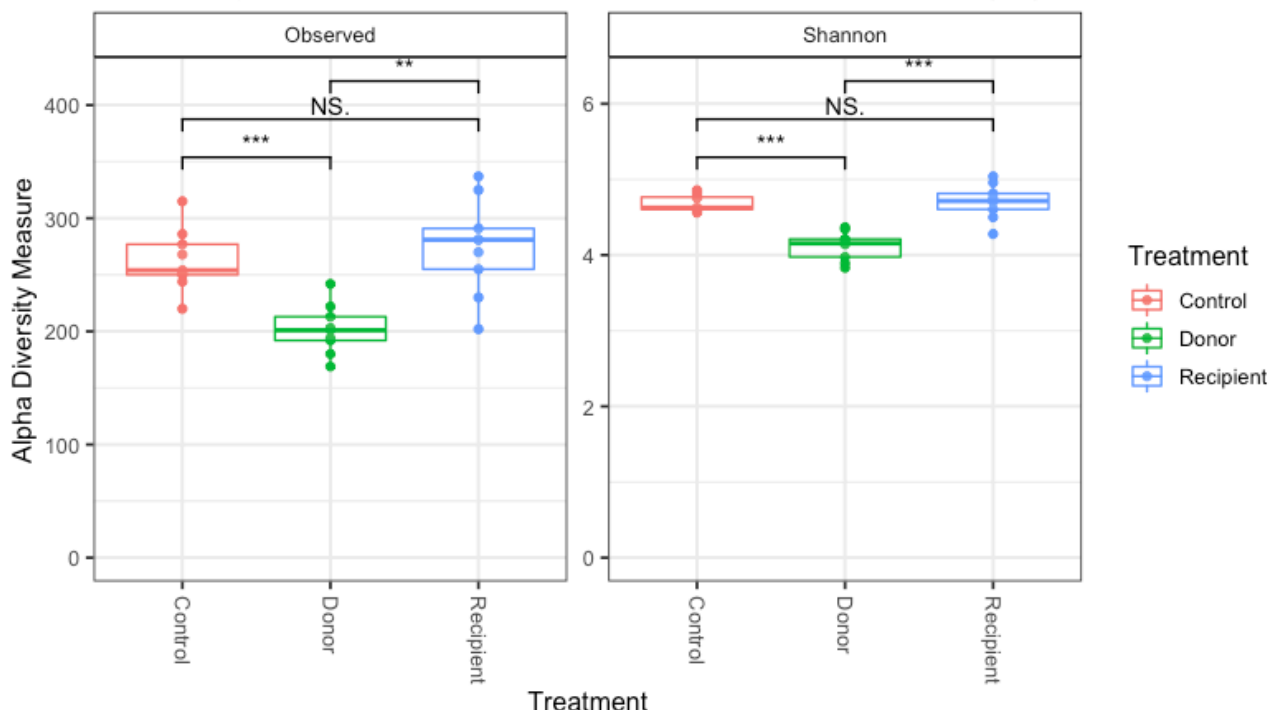


The diversity and species richness (number of different species) is significantly lower in the donor samples; whereas the Control and Recipient samples are similar at both the baseline (T1) and before transplant (T2). \*\* indicated statistical significant difference inferred with Wilcoxon-signed rank test with  $p \leq 0.01$ ; NS means no significant difference. This is consistent to what is known: healthy periodontal plaque microbiome has lower diversities than diseased ones. Note that Donor samples at T2 (in the figure at the bottom of this page) refer to the processed plaque samples from donor that were transferred to the recipient dogs.

Comparing alpha-diversities of microbiome at baseline (T1)

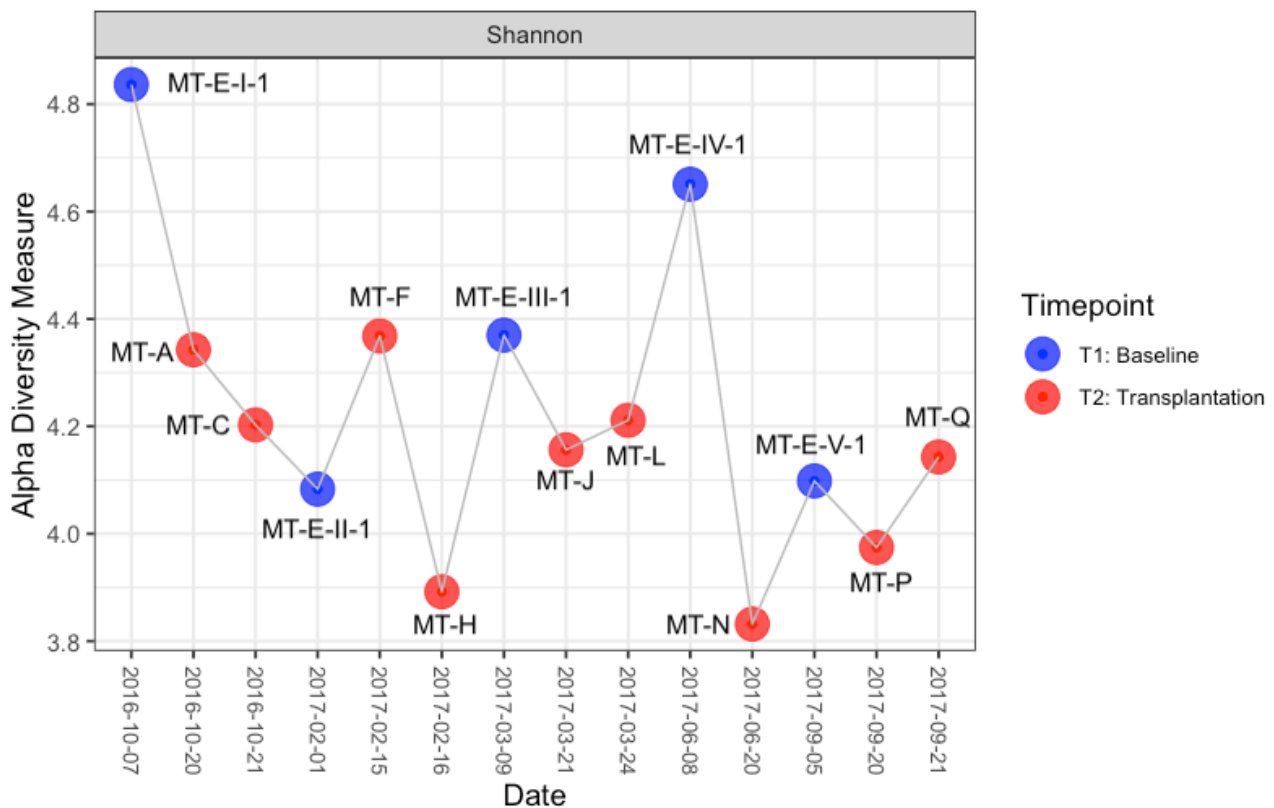


Comparing alpha-diversities of microbiome at before transplant (T2)

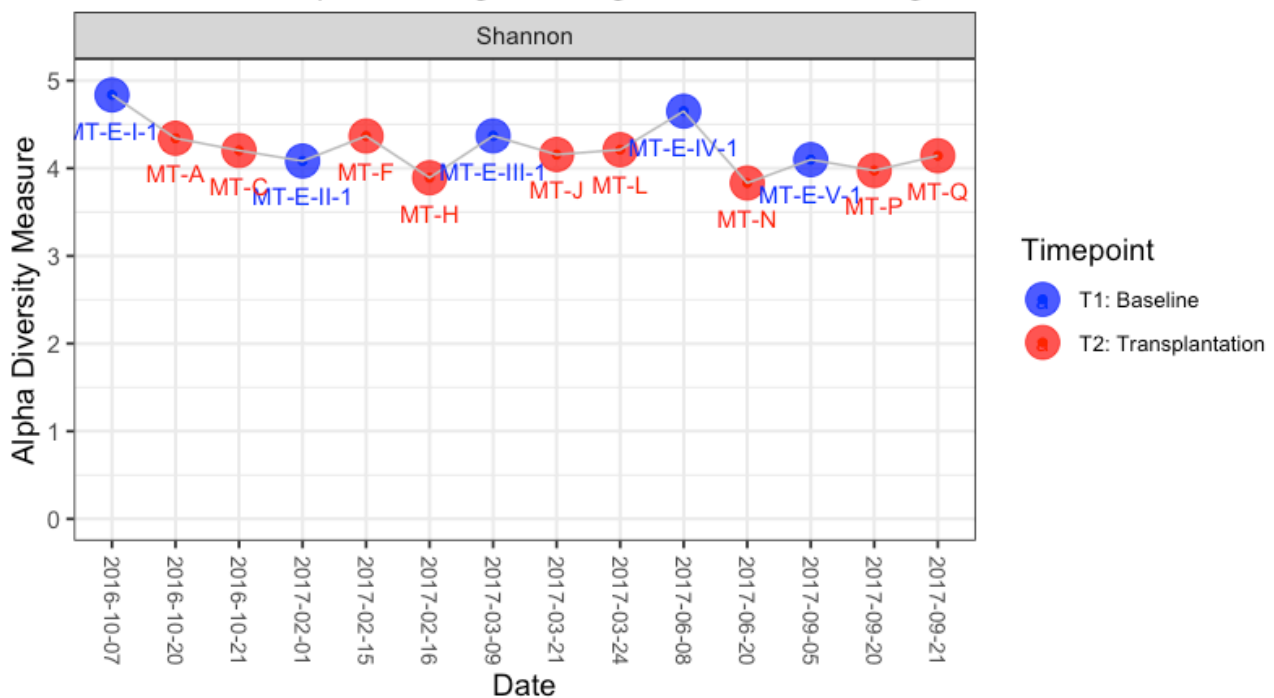


There are intrinsic changes in the diversities of microbiome in the donor dog. The increase in MT-F relative to MT-R-II-1 is awkward. However the overall range of variation should be acceptable.

Shannon Entropies changes along time in Donor dog E



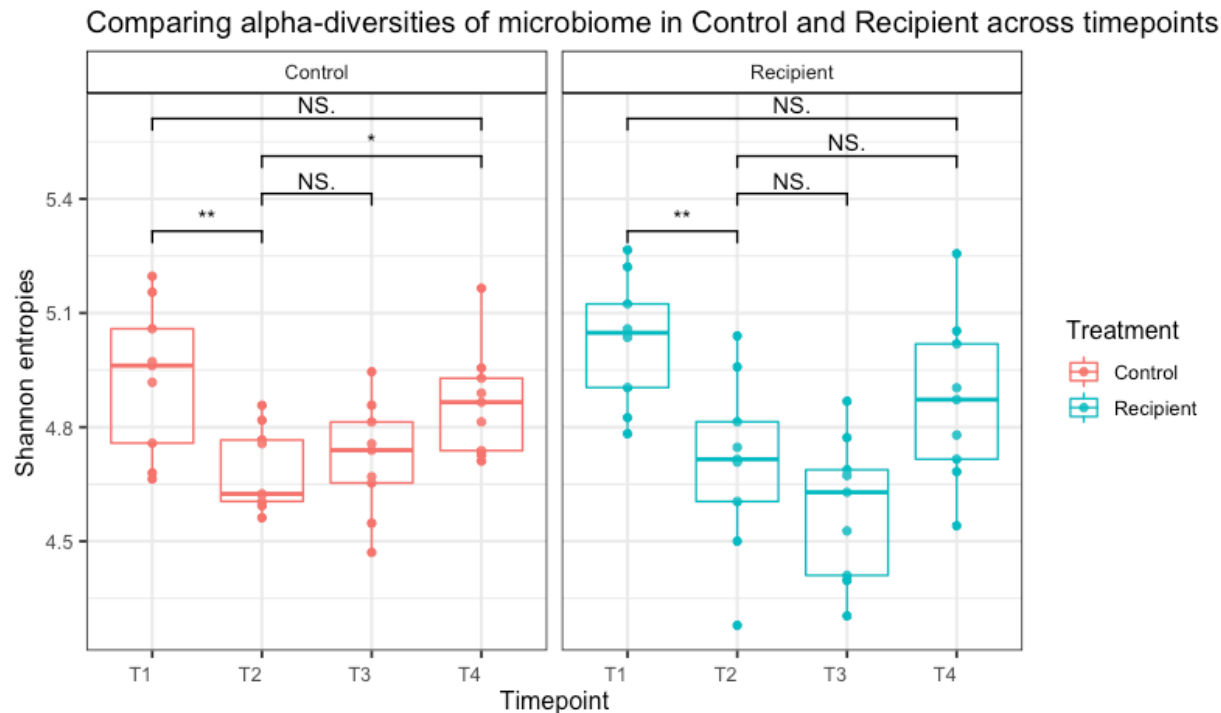
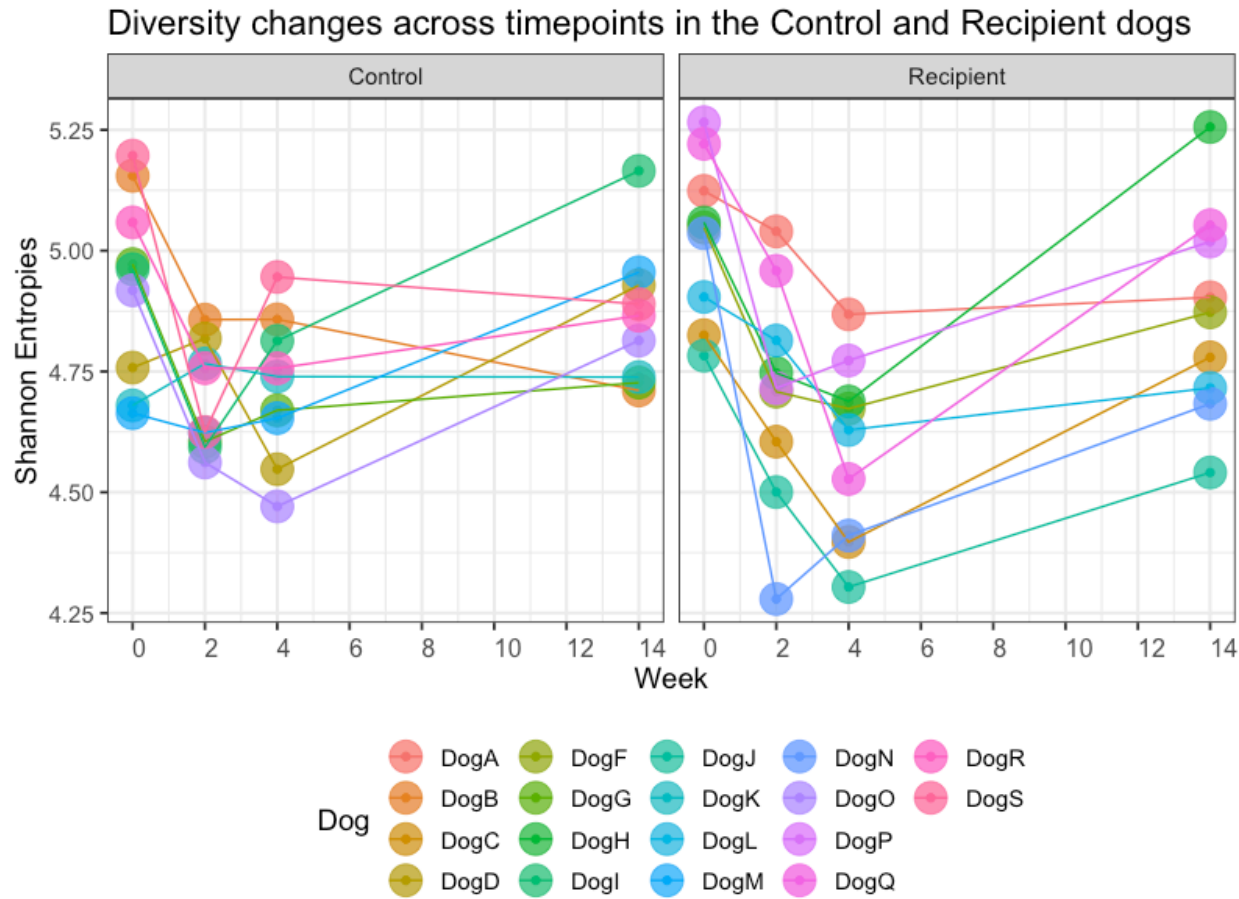
Shannon Entropies changes along time in Donor dog E



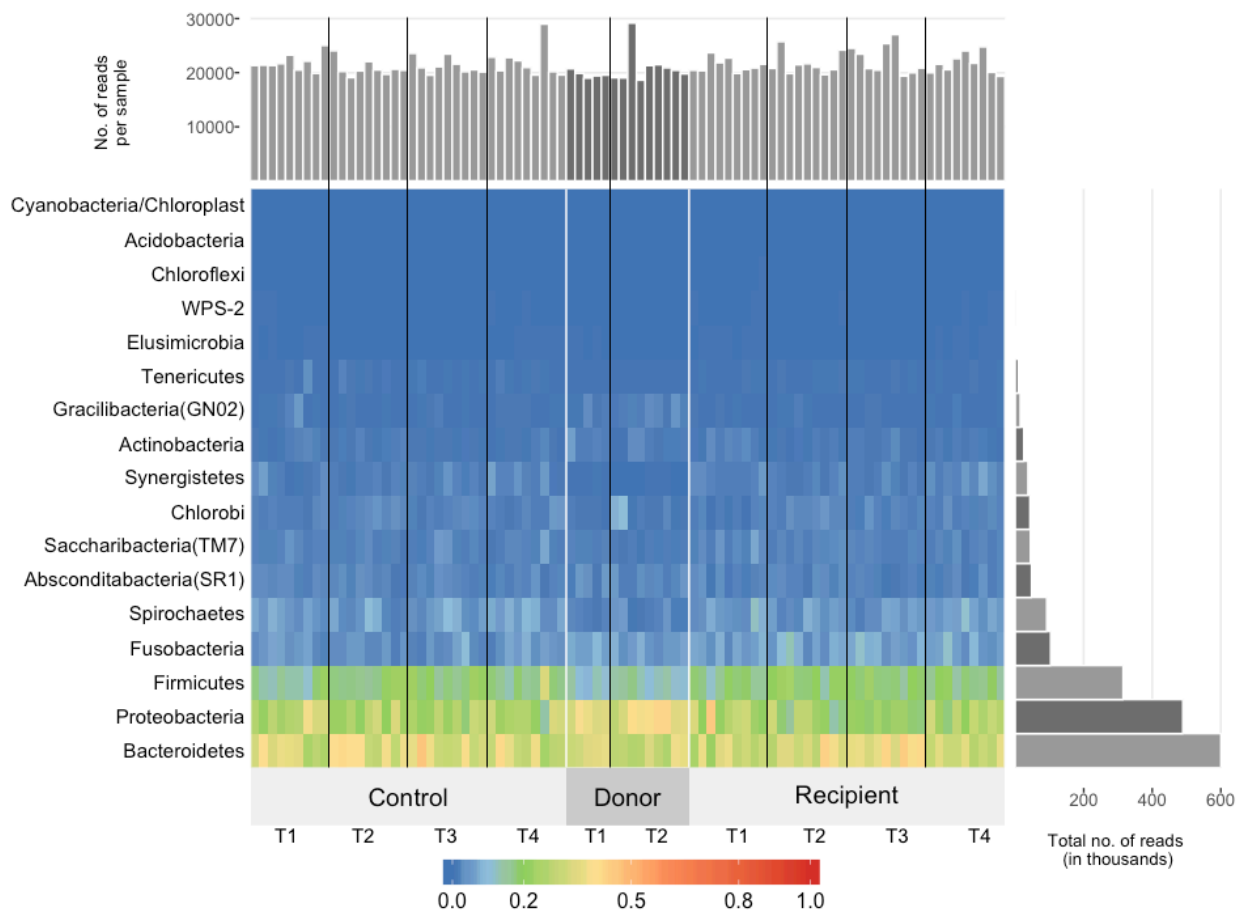
\*Updated scale of y-axis to start from 0



The effect of the first cleaning at baseline was obvious in reducing the diversity of microbiomes (week0vs2 = T1vsT2) in both control and recipient dogs ( $p < 0.01$ ). The effect of transplantation was less clear (week2vs4 = T2vsT3). It seemed to further lower diversities in the recipient (NS in Wilcoxon test). The diversity restored at 14 weeks (T4vsT1) in both Control and Recipient dogs.

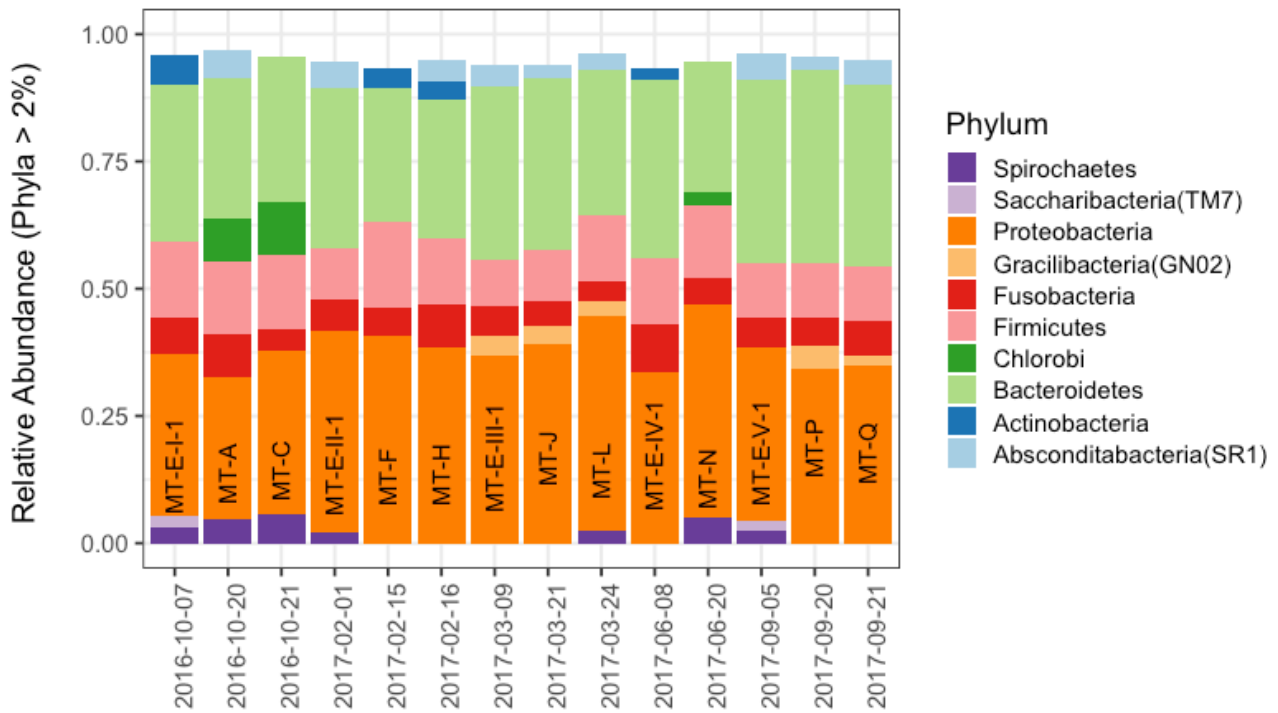


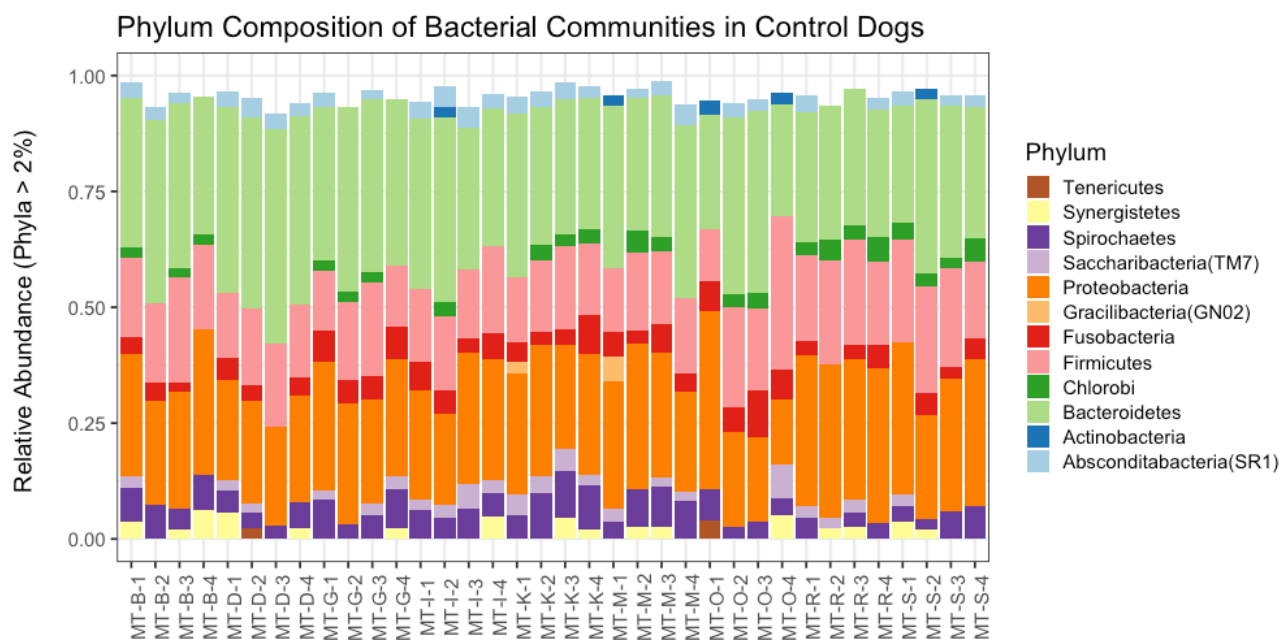
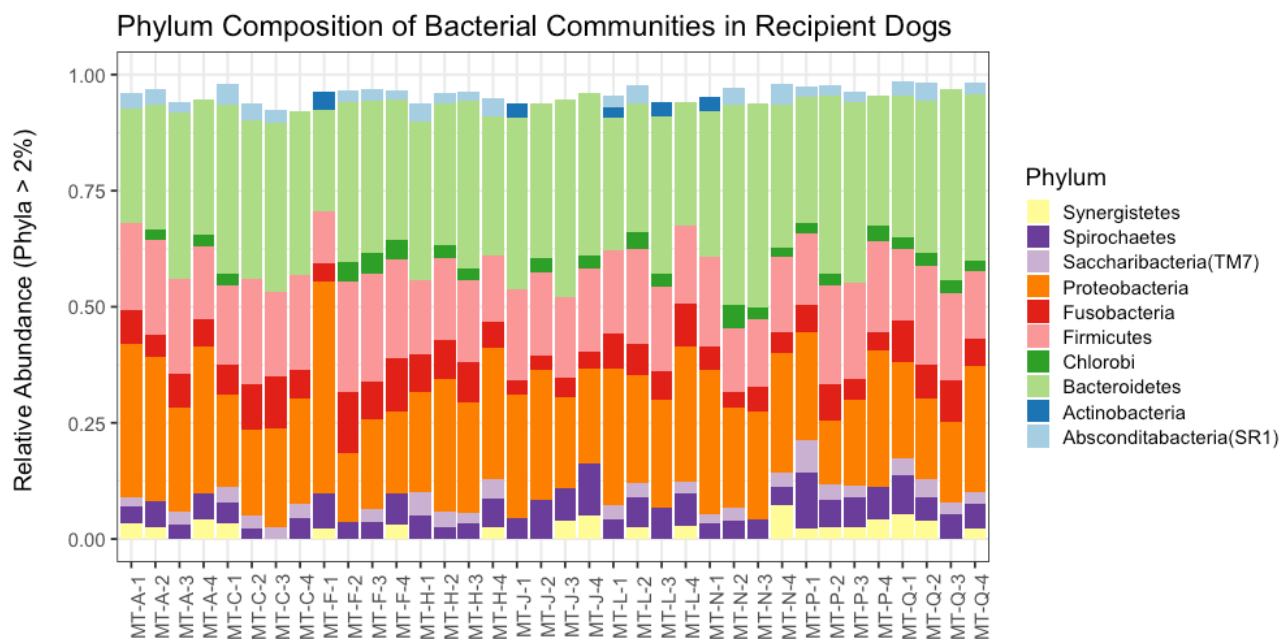
**What bacteria are in the samples?** This heatmap shows the phylum composition of all the samples.



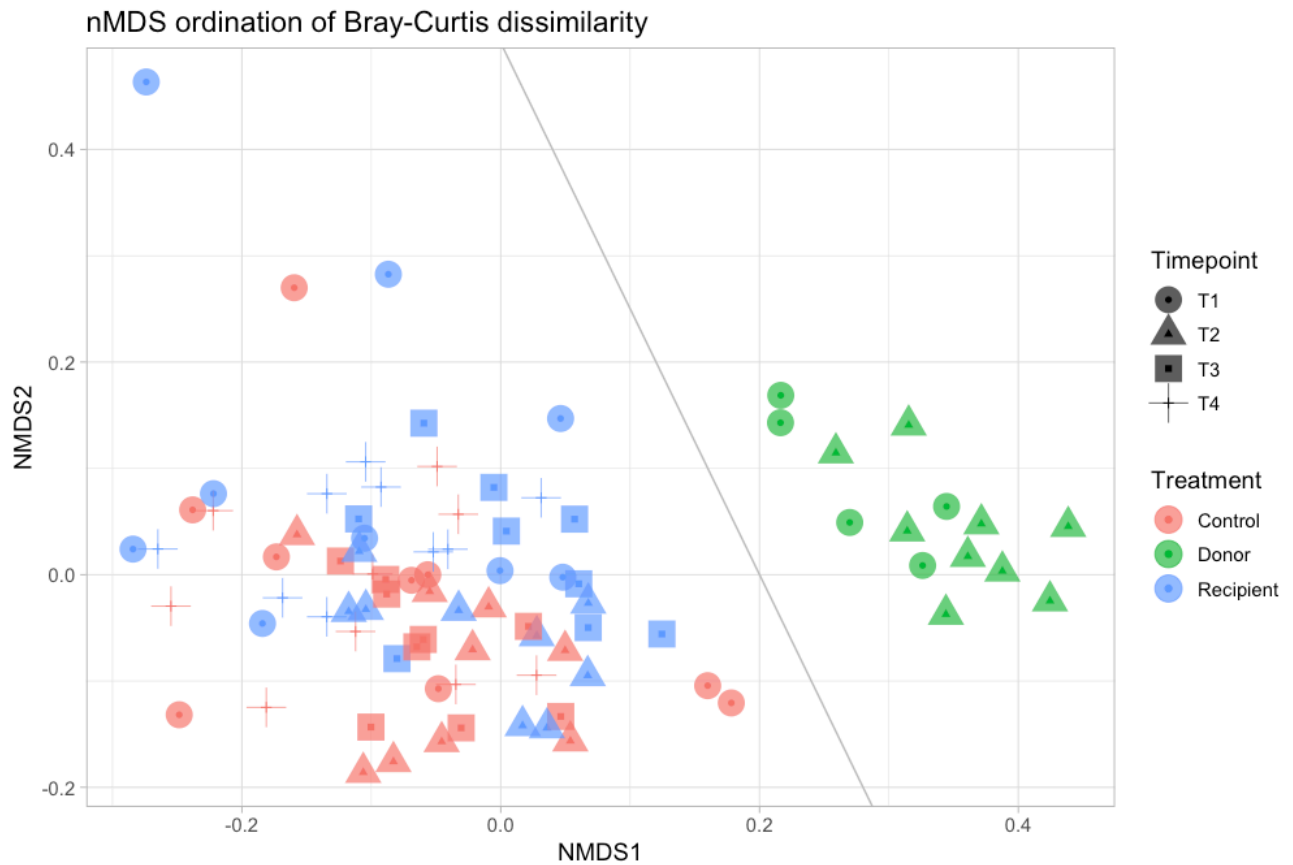
The three bacterial phyla **Bacteroidetes**, **Proteobacteria** and **Firmicutes** are dominant and prevalent in all samples. **Synergistetes** is absent in the donor samples. There are less Spirochaetes in the donor samples

## Phylum Composition of Bacterial Communities in the Donor Dog

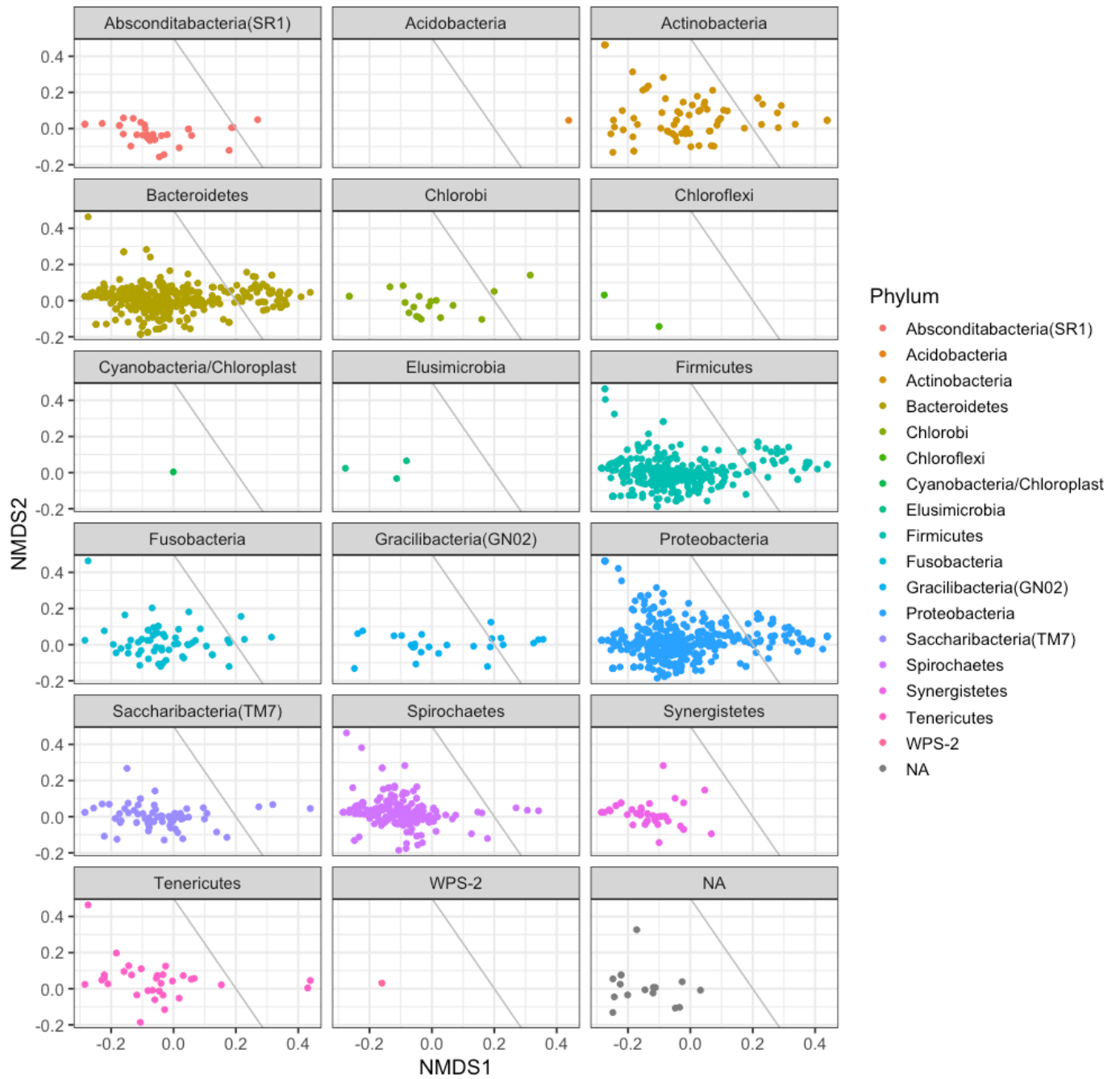




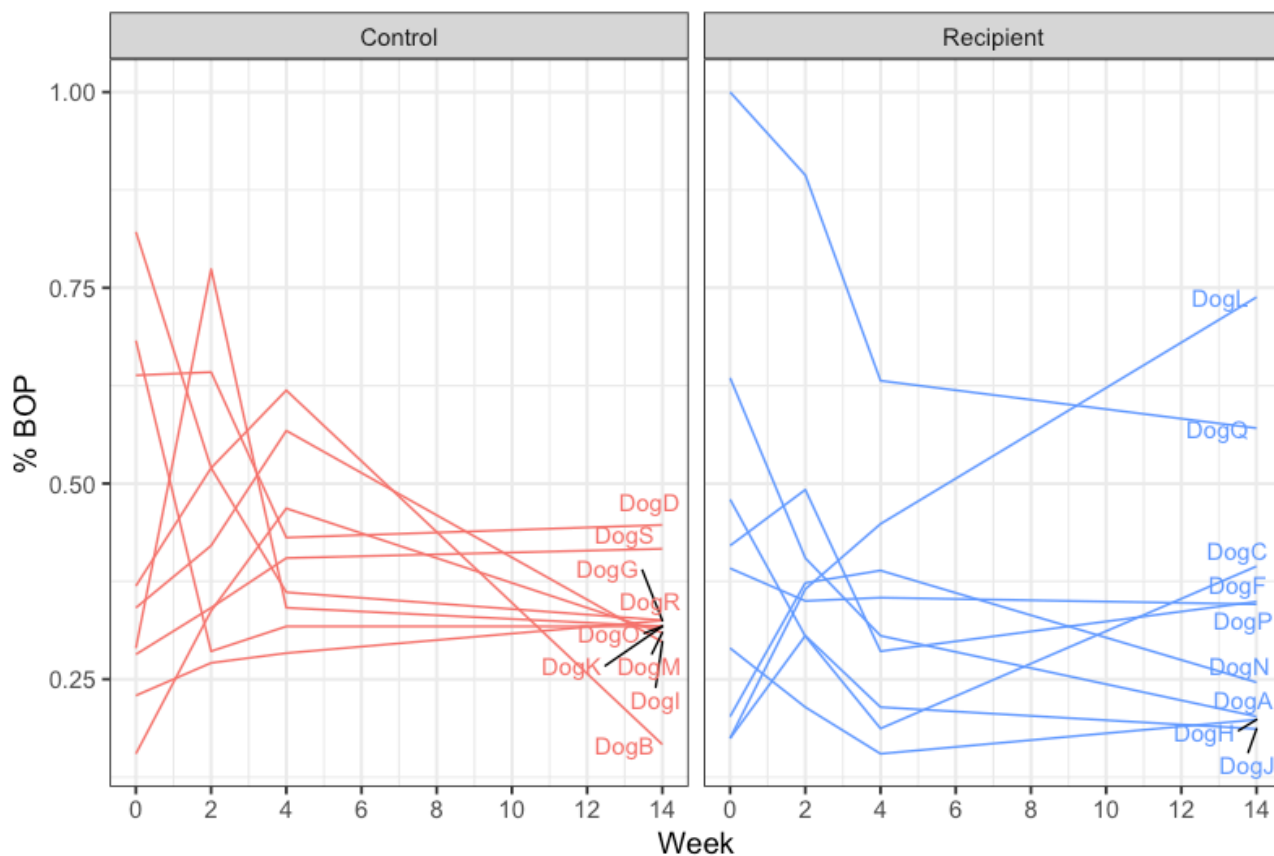
Donor samples are distinctive from the control and recipient samples. There is no obvious trend of microbiome shift in the recipient dog samples (blue square and cross) after transplantation in the unconstrained ordination analyses.



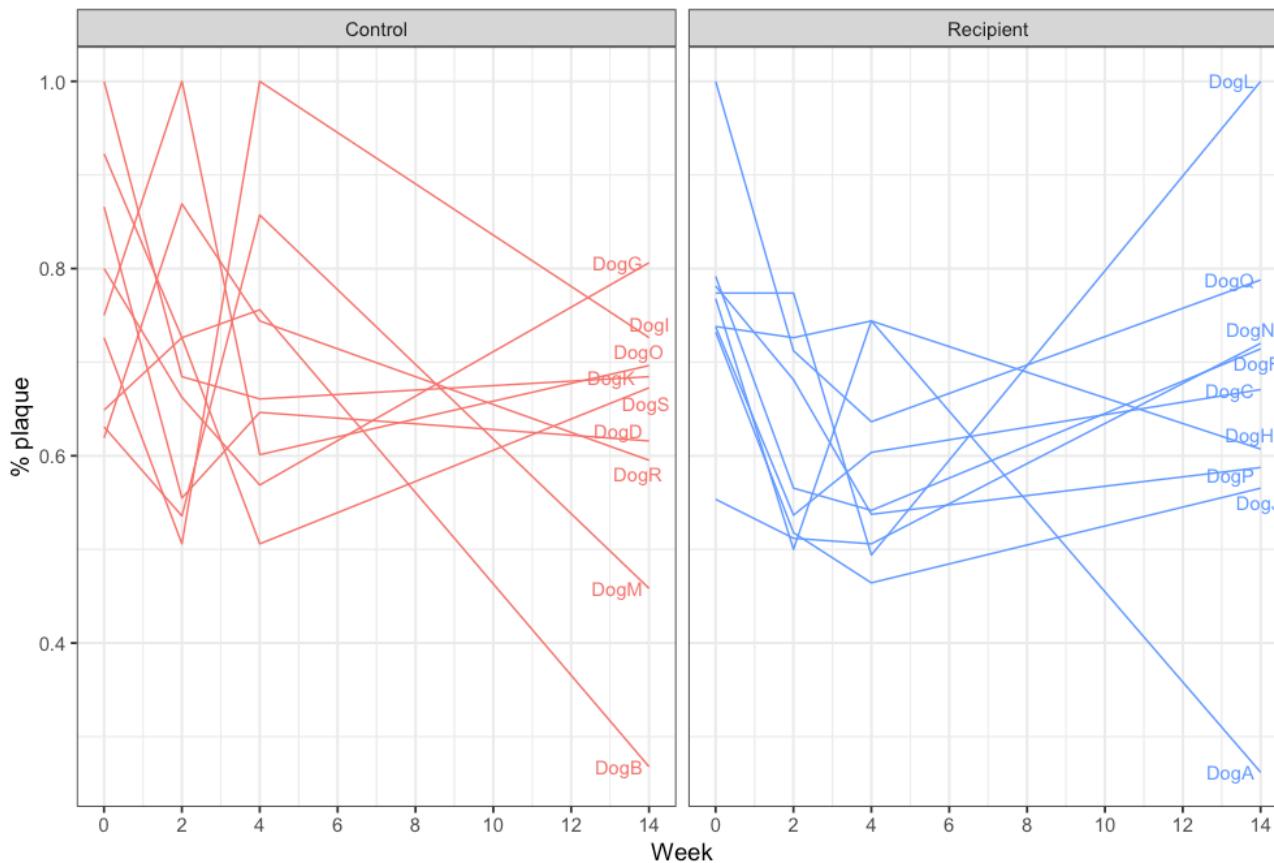
nMDS ordination showing the bacterial taxa



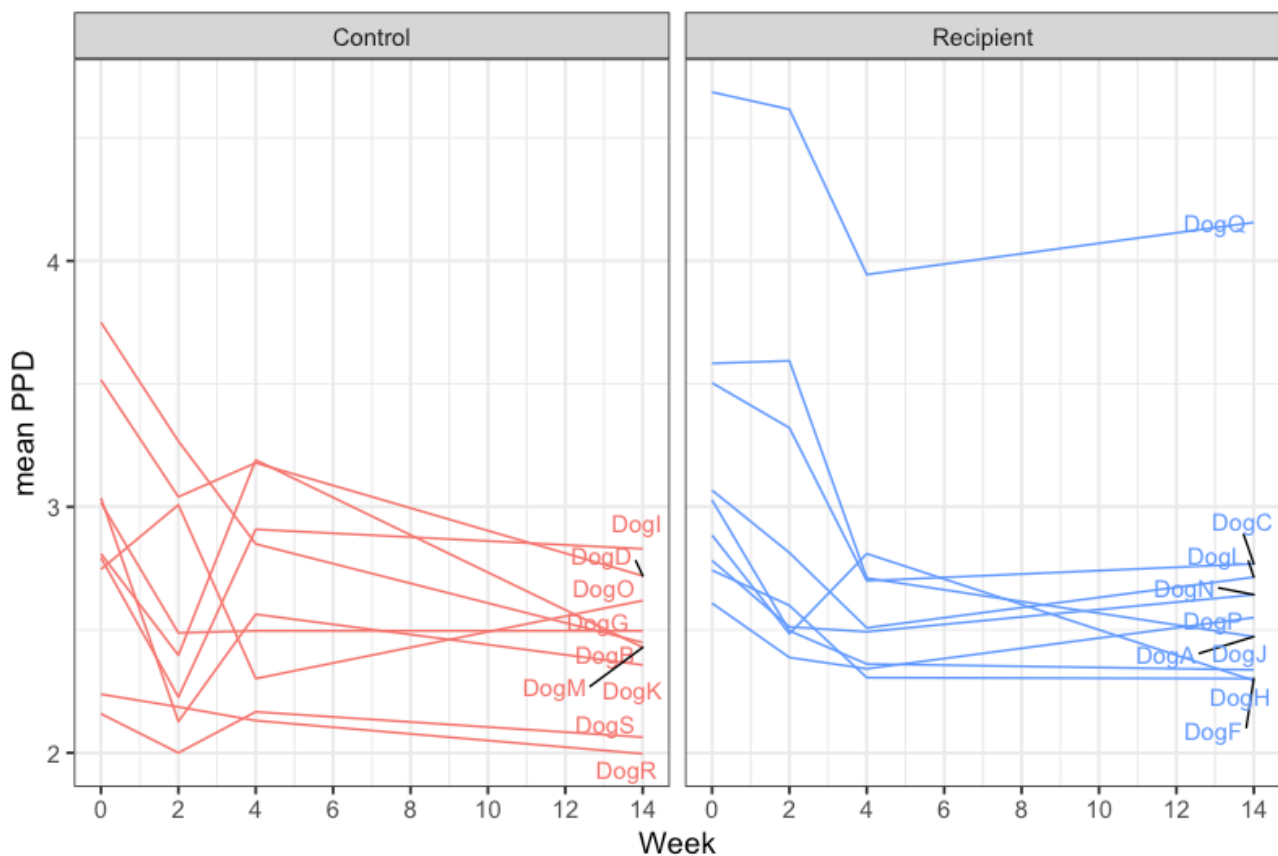
### Percentage BOP in the Control and Recipient dogs



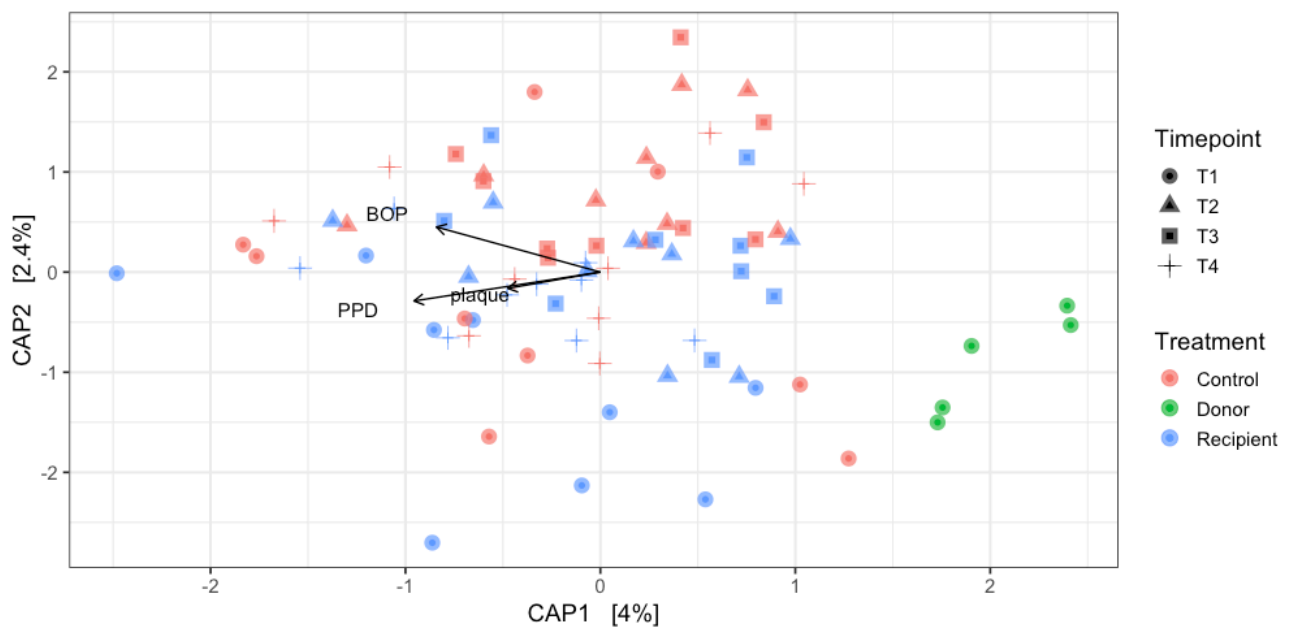
### Plaque percentage in the Control and Recipient dogs



## Mean PPD in the Control and Recipient dogs



With these clinical measurements data, next we sought to examine how these variables (BOP%, mean PPD and plaque %) are associated with the changes in the microbiome community composition; by constrained ordination analysis using Distance-based redundancy analysis.

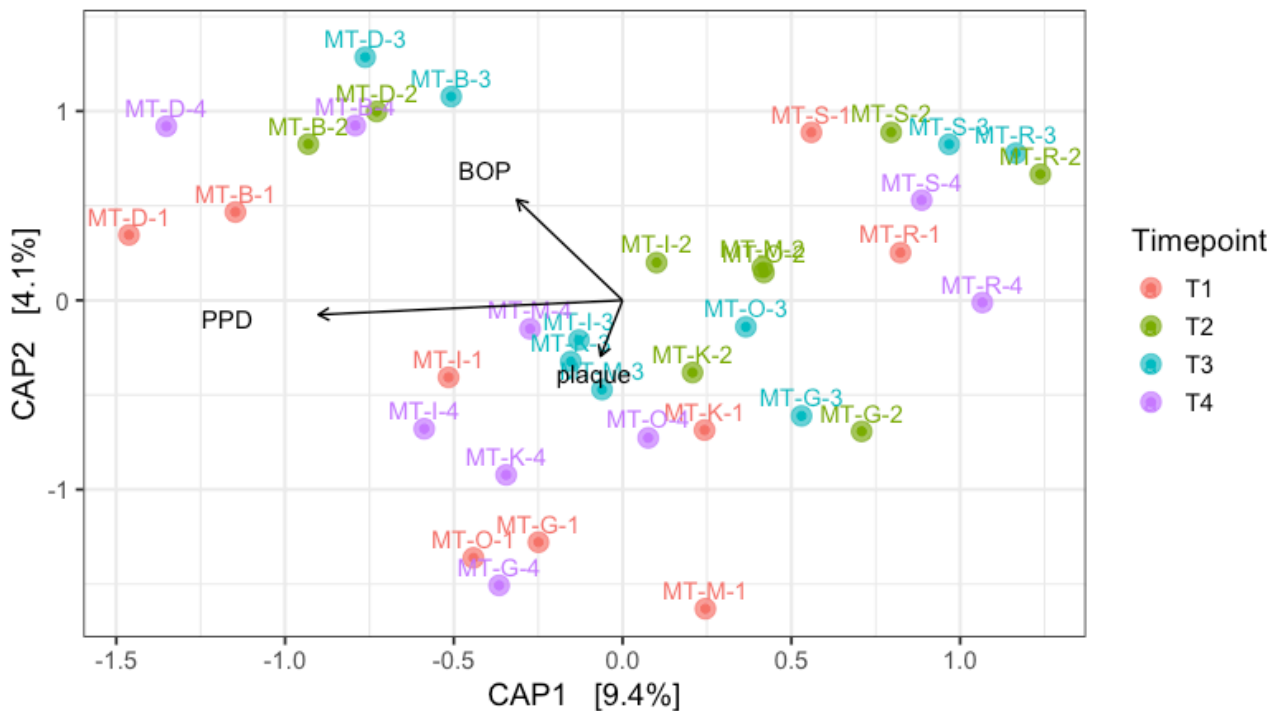


Donor samples cluster towards the right whereas the BOP & PPD measurements increase towards the left (gradient indicated by black arrows). The major differences in the microbiome

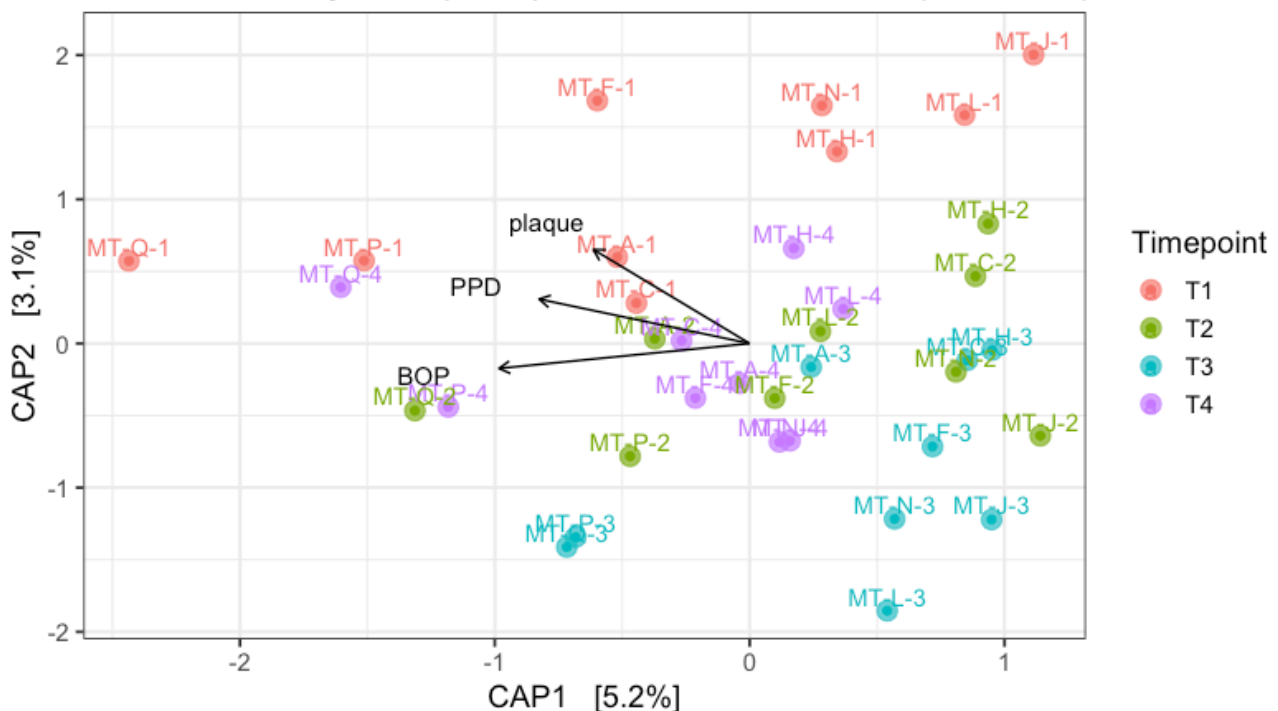


between the donor vs control and recipient are associated to lower % BOP ; % plaque and mean PPD measurements.

### Canonical analysis of principal coordinates for Control samples



### Canonical analysis of principal coordinates for Recipient samples

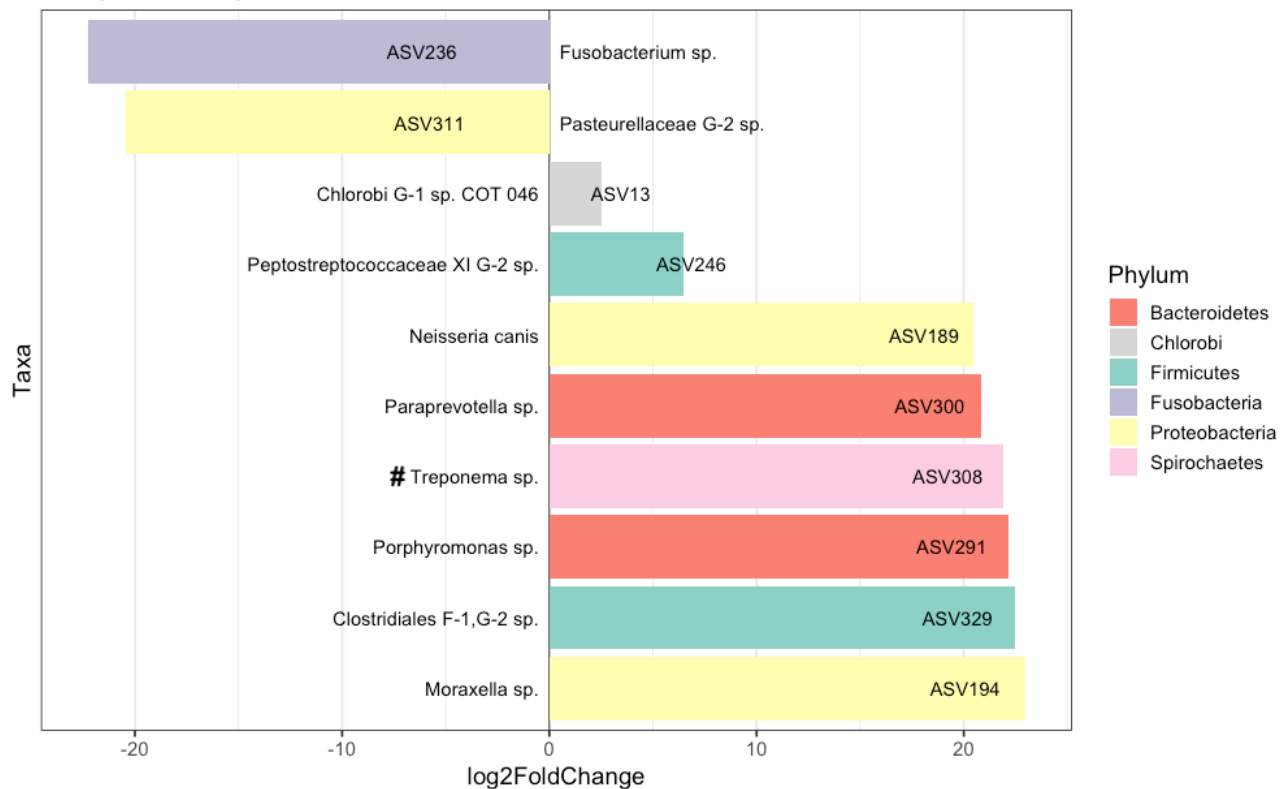


Excluding the donor samples and looking at the control and recipient samples separately: these two constrained ordination plots indicate the effects of transplantation.

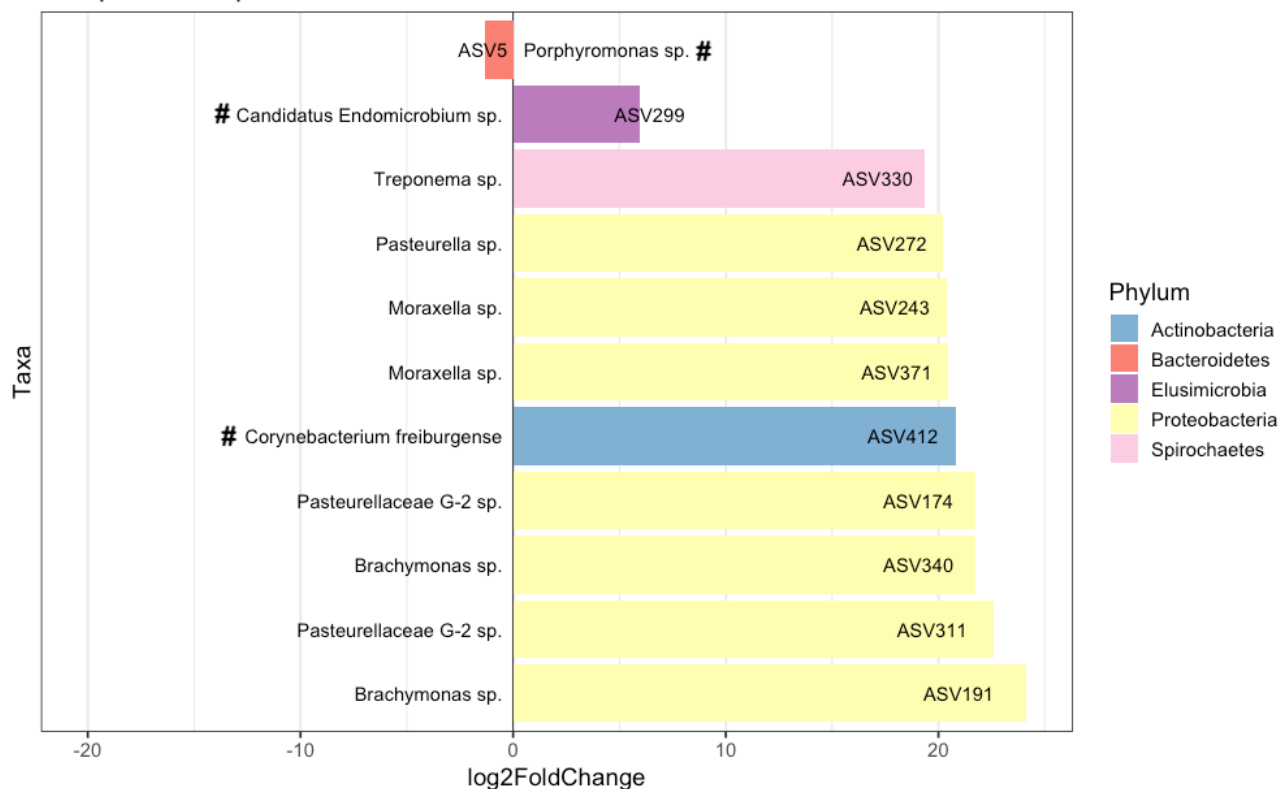
## Differential Abundances analysis

These ASVs are showing significant difference in their abundances in the recipient dogs at different timepoints

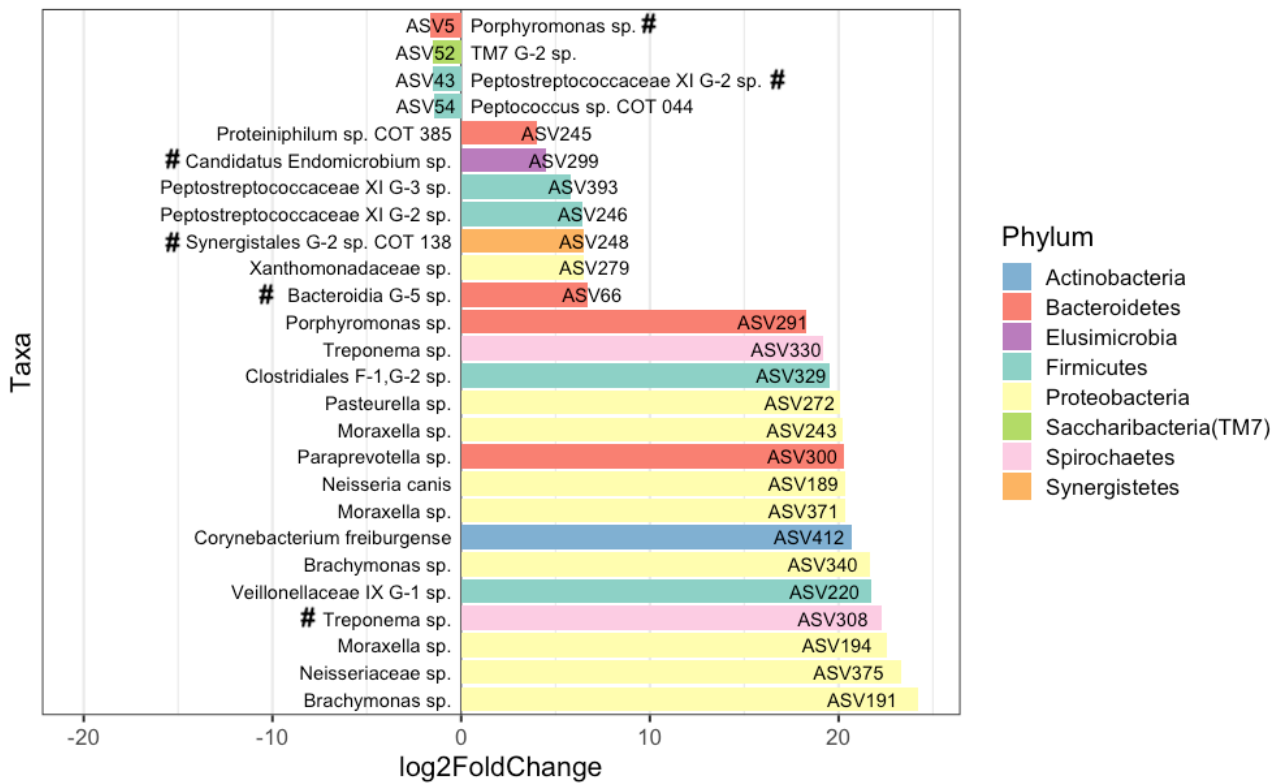
Recipient Timepoint 2 vs 3



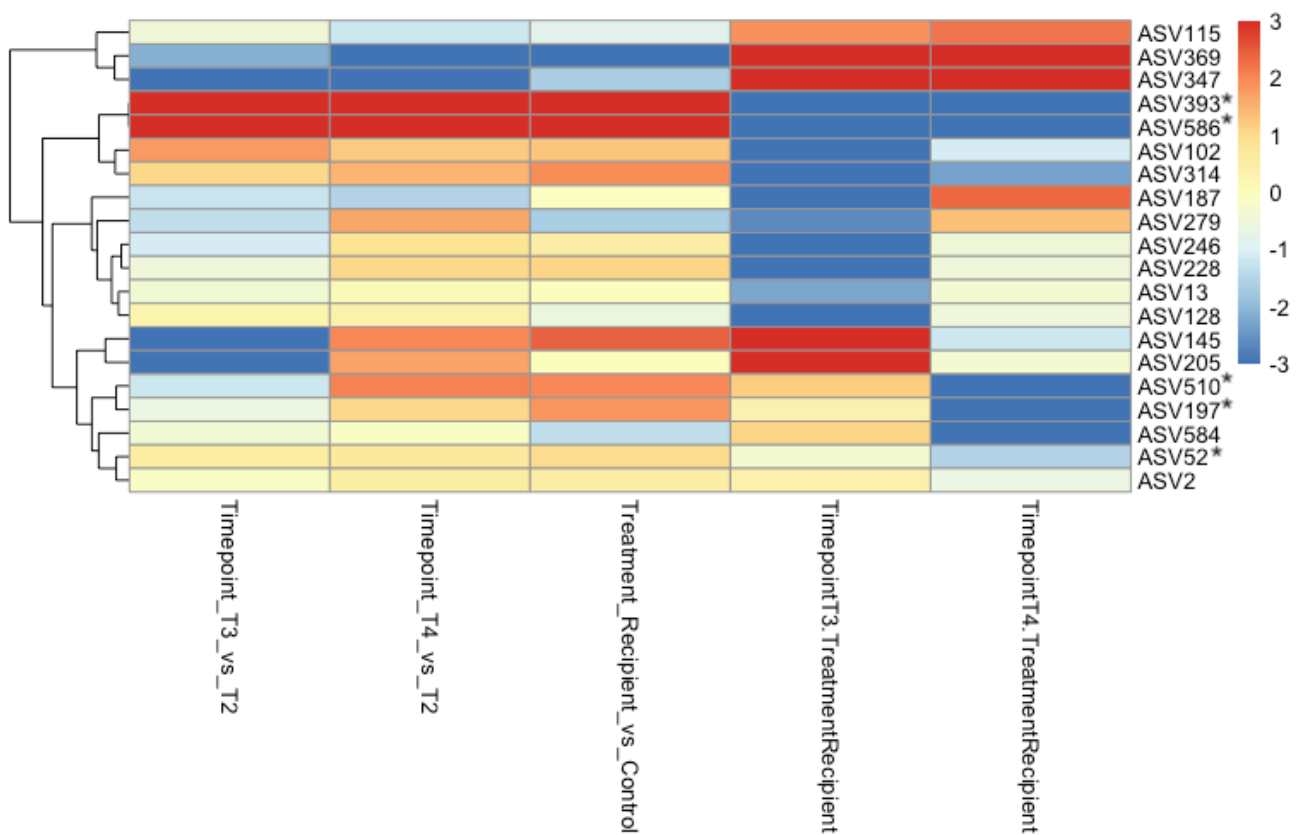
Recipient Timepoint 2 vs 4



## Recipient Timepoint 3 vs 4



## Time-series comparison of bacterial abundances



## Supplementary

### General description of the methods used for the analysis:

#### 1) Quality check of the sequences obtained from BGI (HK), using fastQC on the fastq files

#### 2) 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.

#### 3) 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” (Quote from Callahan et al. 2016)

#### 4) Statistical analyses using phyloseq and other R package

Ref: McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one*, 8(4), e61217.

Ref: Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J., & Holmes, S. P. (2016). Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *F1000Research*, 5.

Ref: Love, Michael I, Wolfgang Huber, and Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome biology* 15.12 (2014): 550.

#### 5) Taxonomy Assignment

“The DADA2 package provides a native implementation of the naive Bayesian classifier method (Wang et al. 2007) for taxonomy assignment. It takes as input a set of sequences to be classified, and a training set of reference sequences with known taxonomy, and outputs taxonomic assignments with at least minBoot (default at 50) bootstrap confidence. The dada2 package also implements a method to make species level assignments based on exact matching between ASVs and sequenced reference strains. Recent analysis (Edgar 2018) suggests that exact matching (or 100% identity) is the only appropriate way to assign species to 16S gene fragments.” (Quoted from DADA2 workflow)

We use a curated reference sequence database that comprised of the RDP trainset 16 (Cole 2014, derived from RDP Release 11.5 that consists of 3,356,809 aligned and annotated 16S rRNA sequences, updated 2016-09-30), eHOMD 16S rRNA RefSeq Version 15.1 (Chen et al. 2010, updated 2017-11-16) and the full set of canine oral microbiome 16S rRNA gene sequences (Dewhirst et al. 2012, 416 full-length sequences). Mothur (Schloss 2009) was used to process the reference sequences.

NB. The silva SSU reference database (v132 last updated 2017-12-13) is not used due to the incompatible taxonomy with the NCBI, HOMD, RDP databases, most notably due to the proposed superphylum, Patescibacteria that includes candidate phyla Gracilibacteria (GN02), Microgenomates, Parcubacteria, and Saccharibacteria (TM7)

**Ref:** Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, 73(16), 5261-5267.

**Ref:** Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J. and Sahl, J.W., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, 75(23), pp.7537-7541.

**Ref:** Edgar, R. C. (2018). Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics*, bty113, <https://doi.org/10.1093/bioinformatics/bty113>.

**Ref:** Cole, J. R., Q. Wang, J. A. Fish, B. Chai, D. M. McGarrell, Y. Sun, C. T. Brown, A. Porras-Alfaro, C. R. Kuske, and J. M. Tiedje. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis *Nucl. Acids Res.* 42(Database issue):D633-D642; doi: 10.1093/nar/gkt1244 [PMID: 24288368]

**Ref:** Chen, T., Yu, W-Han, Izard, J., Baranova, O.V., Lakshmanan, A., Dewhirst, F.E. (2010) The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database*, Vol. 2010, Article ID baq013, doi: 10.1093/database/baq013

**Ref:** Dewhirst, Floyd E., Erin A. Klein, Emily C. Thompson, Jessica M. Blanton, Tsute Chen, Lisa Milella, Catherine MF Buckley, Ian J. Davis, Marie-Lousie Bennett, and Zoe V. Marshall-Jones. "The canine oral microbiome." *PloS one* 7, no. 4 (2012): e36067.

## 6) Functional metagenomic prediction of canine oral microbiome

"PICRUSt (phylogenetic investigation of communities by reconstruction of unobserved states), a computational approach to predict the functional composition of a metagenome using marker gene data and a database of reference genomes. PICRUSt uses an extended ancestral-state reconstruction algorithm to predict which gene families are present and then combines gene families to estimate the composite metagenome. Using 16S information, PICRUSt recaptures key findings from the Human Microbiome Project and accurately predicts the abundance of gene families in host-associated and environmental communities, with quantifiable uncertainty."

A validated dada2-to-PICRUSt pipeline was used for the analysis:

[https://github.com/vmaffei/dada2\\_to\\_picrust](https://github.com/vmaffei/dada2_to_picrust)

Accuracy of this pipeline has been tested using paired 16S and shotgun metagenome datasets.

Among the 1759 ASVs (Amplicon Sequence Variants) in the canine oral microbiome dataset, 1374 (78.11%) were matched to the greengene database and functional metagenome information available.

**Ref:** Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Thurber RL, Knight R, Beiko RG. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology*. 2013 Sep;31(9):814.

**TableS1 Statistics of data after processing through the DADA2 pipeline:**

	<b>SampleID</b>	<b>RawReads</b>	<b>clean</b>	<b>filtered</b>	<b>denoised</b>	<b>merged</b>	<b>nonchimeric</b>	<b>%output</b>
1	MT-A-1	50265	35144	34247	34247	24877	<b>20685</b>	41.2%
2	MT-A-2	50968	35500	34708	34708	25757	<b>20851</b>	40.9%
3	MT-A-3	54617	35000	34061	34061	27890	<b>24538</b>	44.9%
4	MT-A-4	49865	34805	33956	33956	25122	<b>20178</b>	40.5%
5	MT-B-1	49589	34992	34160	34160	25570	<b>21456</b>	43.3%
6	MT-B-2	53257	35478	34450	34450	28033	<b>24079</b>	45.2%
7	MT-B-3	55239	35060	34231	34231	27703	<b>23592</b>	42.7%
8	MT-B-4	49250	34646	33925	33925	27474	<b>23103</b>	46.9%
9	MT-C-1	53289	34817	34269	34269	26331	<b>20717</b>	38.9%
10	MT-C-2	60002	34883	34239	34239	29144	<b>25740</b>	42.9%
11	MT-C-3	57681	34992	34470	34470	28566	<b>23427</b>	40.6%
12	MT-C-4	51699	34849	34314	34314	26374	<b>21686</b>	41.9%
13	MT-D-1	51900	35000	34393	34393	26761	<b>21670</b>	41.8%
14	MT-D-2	51878	35009	34387	34387	26129	<b>20253</b>	39.0%
15	MT-D-3	54702	34617	33961	33961	26869	<b>20929</b>	38.3%
16	MT-D-4	47834	34408	33849	33849	25489	<b>20566</b>	43.0%
17	MT-E-I-1	51331	34772	33987	33987	25426	<b>20800</b>	40.5%
18	MT-E-II-1	54934	34848	34011	34011	27058	<b>19876</b>	36.2%
19	MT-E-III-1	51587	34949	34087	34087	26424	<b>19000</b>	36.8%
20	MT-E-IV-1	51141	34941	33998	33998	25444	<b>19443</b>	38.0%
21	MT-E-V-1	56223	35034	34177	34177	27120	<b>19545</b>	34.8%
	<b>MT-E-V-2</b>	<b>59157</b>	<b>34975</b>	<b>34252</b>	<b>34252</b>	<b>29908</b>	<b>22116</b>	<b>37.4%</b>
22	MT-F-1	53046	34633	34083	34083	27705	<b>23819</b>	44.9%
23	MT-F-2	52271	34933	34305	34305	26367	<b>19841</b>	38.0%
24	MT-F-3	52542	35009	34373	34373	26276	<b>20763</b>	39.5%
25	MT-F-4	51180	35169	34549	34549	26088	<b>20615</b>	40.3%
26	MT-G-1	51628	35246	34628	34628	26390	<b>21473</b>	41.6%
27	MT-G-2	52810	34549	33977	33977	25882	<b>19090</b>	36.1%
28	MT-G-3	49457	34389	33817	33817	25868	<b>19570</b>	39.6%
29	MT-G-4	55495	34808	34175	34175	27205	<b>22848</b>	41.2%
30	MT-H-1	52976	35244	34612	34612	26429	<b>21923</b>	41.4%
31	MT-H-2	52791	35304	34571	34571	26487	<b>21527</b>	40.8%
32	MT-H-3	52368	34638	34047	34047	26388	<b>20480</b>	39.1%
33	MT-H-4	51536	35280	34631	34631	25986	<b>22714</b>	44.1%
34	MT-I-1	52996	35289	34608	34608	26597	<b>21711</b>	41.0%
35	MT-I-2	55780	34808	34115	34115	26696	<b>20330</b>	36.4%
36	MT-I-3	49221	34732	34293	34293	26238	<b>21201</b>	43.1%
37	MT-I-4	51675	35017	34359	34359	26118	<b>22772</b>	44.1%
38	MT-J-1	54336	35135	34479	34479	27661	<b>22780</b>	41.9%
39	MT-J-2	53817	35060	34351	34351	27544	<b>21665</b>	40.3%
40	MT-J-3	56563	34899	34138	34138	29498	<b>25406</b>	44.9%
41	MT-J-4	55658	35560	34805	34805	28791	<b>24203</b>	43.5%

42	MT-K-1	54890	35518	34645	34645	27321	<b>23256</b>	42.4%
43	MT-K-2	54327	35051	34362	34362	26543	<b>22164</b>	40.8%
44	MT-K-3	51341	34807	34169	34169	27723	<b>23614</b>	46.0%
45	MT-K-4	53320	35026	34277	34277	26154	<b>21370</b>	40.1%
46	MT-L-1	50316	34672	33911	33911	25401	<b>19924</b>	39.6%
47	MT-L-2	51657	35102	34263	34263	26928	<b>21030</b>	40.7%
48	MT-L-3	58197	35043	34052	34052	29715	<b>27059</b>	46.5%
49	MT-L-4	53352	35552	34848	34848	27182	<b>21792</b>	40.8%
50	MT-M-1	52059	35035	34049	34049	25566	<b>20515</b>	39.4%
51	MT-M-2	53180	34589	33788	33788	26763	<b>20618</b>	38.8%
52	MT-M-3	50304	34984	34350	34350	27257	<b>21687</b>	43.1%
53	MT-M-4	51884	34580	33777	33777	25292	<b>19853</b>	38.3%
54	MT-N-1	52130	35457	34721	34721	25555	<b>20643</b>	39.6%
55	MT-N-2	55280	35432	34608	34608	27423	<b>19635</b>	35.5%
56	MT-N-3	52272	34831	33986	33986	26424	<b>19408</b>	37.1%
57	MT-N-4	56101	35630	34656	34656	28763	<b>24902</b>	44.4%
58	MT-O-1	53803	35605	34687	34687	26806	<b>22185</b>	41.2%
59	MT-O-2	55477	35085	34204	34204	26594	<b>19698</b>	35.5%
60	MT-O-3	51406	34705	34086	34086	26953	<b>20183</b>	39.3%
61	MT-O-4	57804	30635	29707	29707	29442	<b>29353</b>	50.8%
62	MT-P-1	49664	34687	33925	33925	24797	<b>21097</b>	42.5%
63	MT-P-2	53456	35255	34454	34454	27137	<b>20575</b>	38.5%
64	MT-P-3	51636	35196	34306	34306	26042	<b>19996</b>	38.7%
65	MT-P-4	50803	35690	34867	34867	26050	<b>20257</b>	39.9%
66	MT-Q-1	49872	35359	34365	34365	25526	<b>21759</b>	43.6%
67	MT-Q-2	56068	35296	34422	34422	27882	<b>24249</b>	43.2%
68	MT-Q-3	49932	34747	34019	34019	27181	<b>20859</b>	41.8%
69	MT-Q-4	49970	34400	33805	33805	24521	<b>19718</b>	39.5%
70	MT-R-1	49775	34750	34182	34182	24948	<b>19961</b>	40.1%
71	MT-R-2	50983	34825	34290	34290	26408	<b>20666</b>	40.5%
72	MT-R-3	51757	35202	34644	34644	26760	<b>20644</b>	39.9%
73	MT-R-4	50754	35236	34708	34708	26132	<b>20203</b>	39.8%
74	MT-S-1	53333	34874	34275	34275	27725	<b>25310</b>	47.5%
75	MT-S-2	49249	34300	33739	33739	26851	<b>20475</b>	41.6%
76	MT-S-3	47863	34909	34372	34372	25541	<b>20172</b>	42.1%
77	MT-S-4	50925	34430	33799	33799	24804	<b>19641</b>	38.6%
78	Transplant-A	54312	34958	34077	34077	26218	<b>19070</b>	35.1%
79	Transplant-C	54635	34958	34087	34087	26529	<b>19030</b>	34.8%
80	Transplant-F	62548	34815	33942	33942	30944	<b>29165</b>	46.6%
81	Transplant-H	53950	34755	33992	33992	27057	<b>18614</b>	34.5%
82	Transplant-J	56797	34907	34068	34068	28091	<b>21314</b>	37.5%
83	Transplant-L	56037	35545	34629	34629	27779	<b>21479</b>	38.3%
84	Transplant-N	57660	35093	34325	34325	27844	<b>20902</b>	36.3%
85	Transplant-P	57327	34916	33943	33943	28430	<b>20442</b>	35.7%
86	Transplant-Q	56452	35457	34637	34637	28008	<b>19784</b>	35.0%

-----End Of Report-----