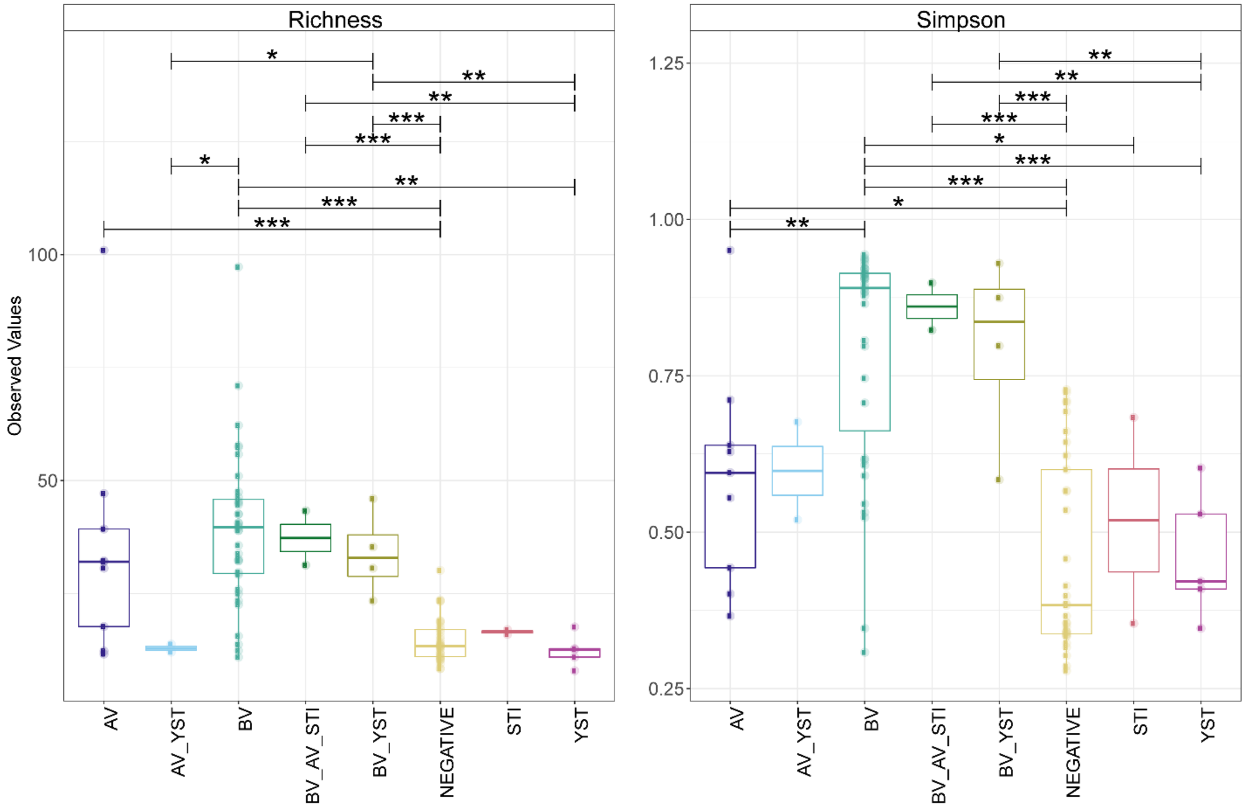
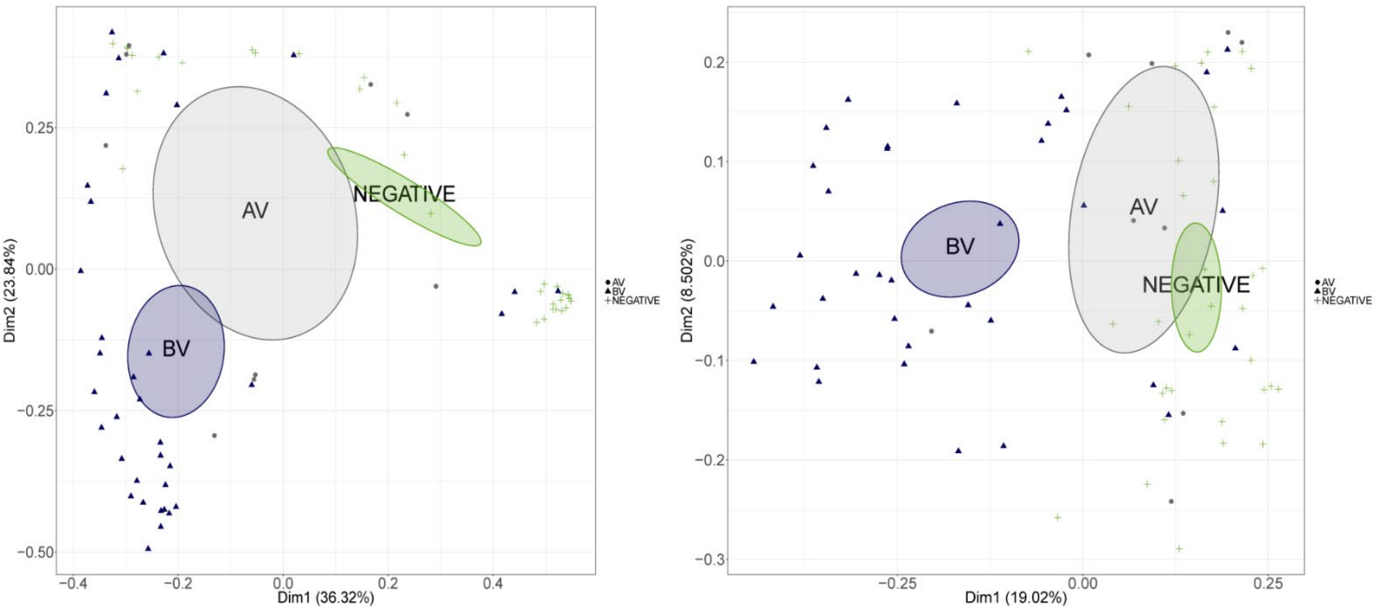
Supplementary Table 1. Relative abundance of each Genus within each biome group. Percentages are calculated by biome group, not individual sample. Only genera present in at least 2% of the total dataset are included. This is a tabular form of the data shown in the stacked bar graph in Figure 1.

Biome group	**Genus**	**Relative abundance by group (%)**
AV	*Lactobacillus*	34.29
AV	*Streptococcus*	14.29
AV	*Gardnerella*	8.57
AV	*Veillonella*	5.71
AV	*Actinomyces*	2.86
AV	*Anaerococcus*	2.86
AV	*Atopobium*	2.86
AV	*Dialister*	2.86
AV	*Escherichia*	2.86
AV	*Finegoldia*	2.86
AV	*Jonquetella*	2.86
AV	*Mobiluncus*	2.86
AV	*Peptoniphilus*	2.86
AV	*Peptostreptococcus*	2.86
AV	*Porphyromonas*	2.86
AV	*Prevotella*	2.86
AV	*Raoultella*	2.86
BV	*Lactobacillus*	14.89
BV	*Gardnerella*	14.04
BV	*Atopobium*	12.34
BV	*Prevotella*	11.49
BV	*Megasphaera*	10.64
BV	*Sneathia*	8.51
BV	*BVAB3*	5.96
BV	*Dialister*	5.53
BV	*Parvimonas*	2.98
BV	*BVAB1*	2.55
BV	*Eggerthella*	2.55
BV	*TM7 vaginal*	1.70
BV	*Aerococcus*	1.28
BV	*Anaerococcus*	0.85

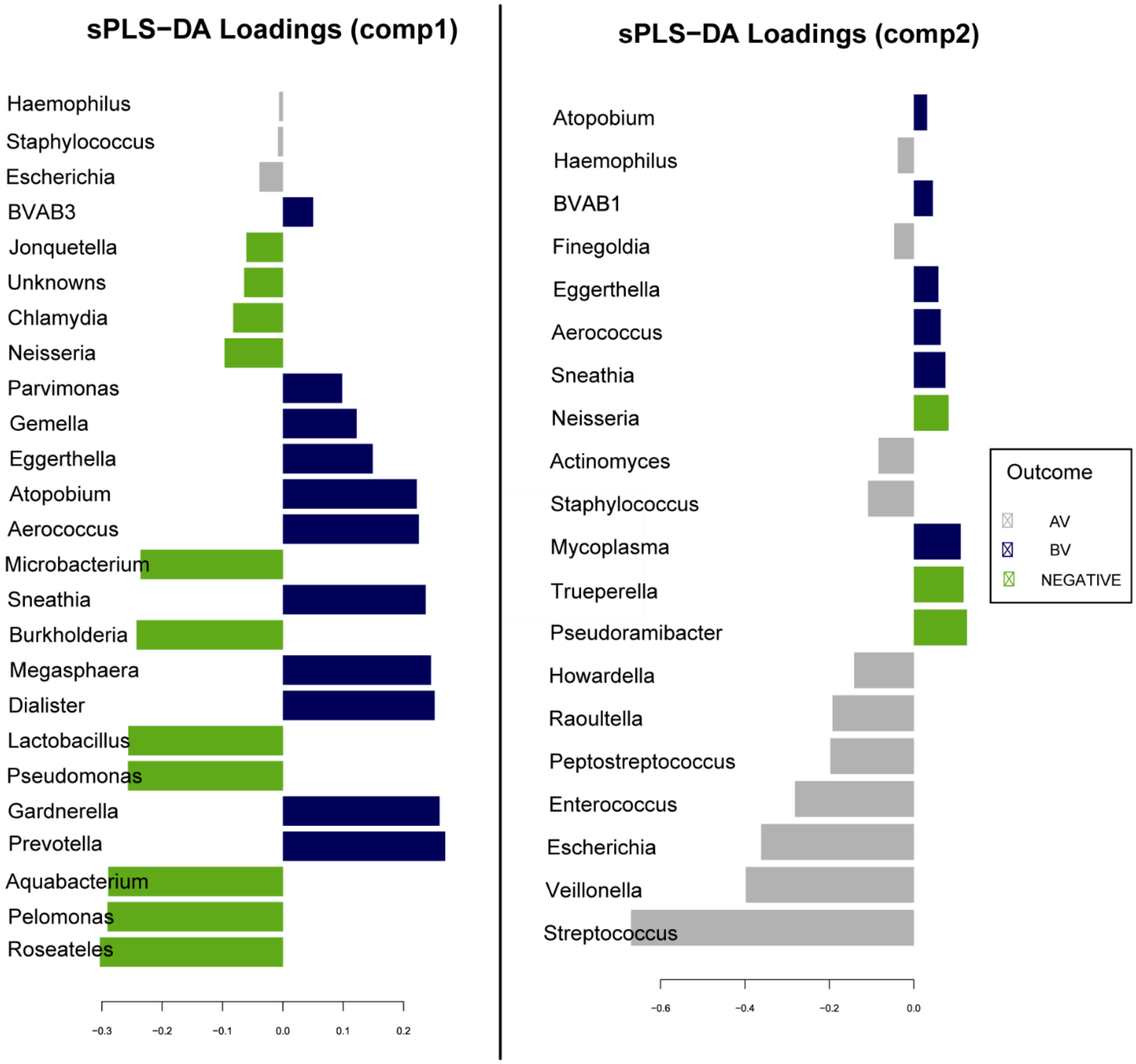
BV	*Peptostreptococcus*	0.85
BV	*Finegoldia*	0.43
BV	*Gemella*	0.43
BV	*Mobiluncus*	0.43
BV	*Mycoplasma*	0.43
BV	*Peptoniphilus*	0.43
BV	*Porphyromonas*	0.43
BV	*Pseudoramibacter*	0.43
BV	*Streptococcus*	0.43
BV	*Ureaplasma*	0.43
Negative	*Lactobacillus*	92.86
Negative	*Gardnerella*	4.76
Negative	*Aerococcus*	2.38



Supplementary Figure 1. The α-diversity by richness and evenness (Simpson diversity index) was calculated for each sample. Samples are categorized based on positivity by qPCR panels. These resultant groups of combined panels were compared to each other by ANOVA (* p<0.1, **p<0.05, ***p<0.001). Due to the small number of samples in some groups and lack of significant statistical differences; samples with only STI, YST or mixed panel positivity were grouped with their predominant biome group being: AV, BV and Negative.



Supplementary Figure 2. Principal Co-ordinate Analysis (PCoA) plotted at the ASV level for each biome group using Bray-Curtis and Unweighted UniFrac distances. The observed variation between groups by abundance and phylogenetic diversity with distinct ellipses (representing 95% confidence interval of the standard error) present for both distance measures. Further analysis using PERMANOVA suggests the biome groups were significant sources of variation by both Bray-Curtis (unadjusted R2=0.1978, $p<0.001$) and unweighted UniFrac (unadjusted R2=0.1216, $p<0.001$). The AV and Negative groups were not significantly different from each other but were from the BV group for both distance matrices examined ($p < 0.01$).



Supplementary Figure 3. The most discriminative features found by the sPLS-DA model for components 1 and 2. Coefficients from the loading vector are displayed as relative proportions between the biome groups, therefore the greater magnitude of the bar represents higher discrimination of that genera for that biome group.