Figure S1, Related to Figure 2. Comparison of viral taxonomic assignment using blastx and VirusSeeker.

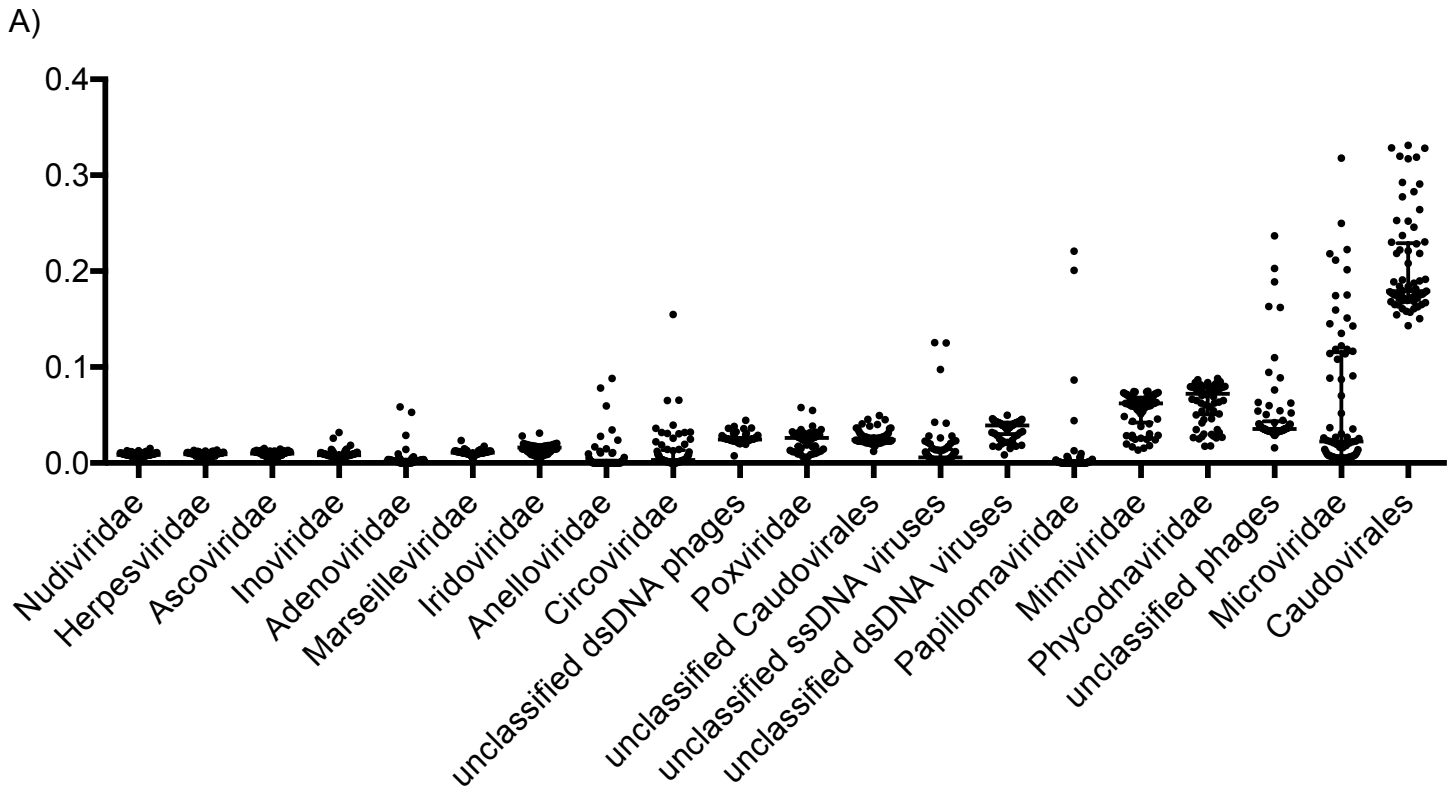


Figure S2, Related to Figure 3. Quantification and comparison of eukaryotic virus sequences detected in fecal samples.

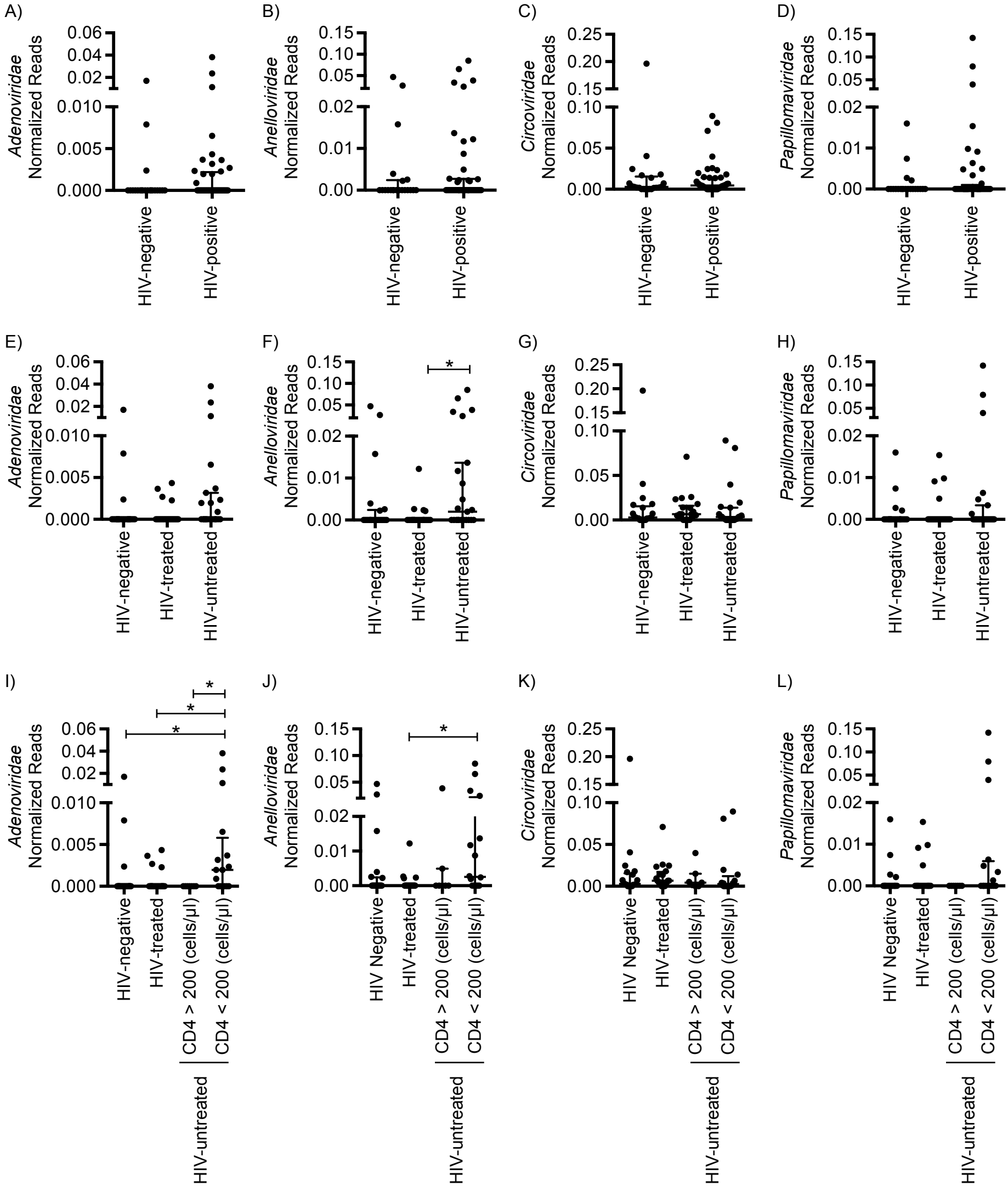


Figure S3, Related to Figure 4. Bacterial class-level taxa barplot by CD4 T cell number.

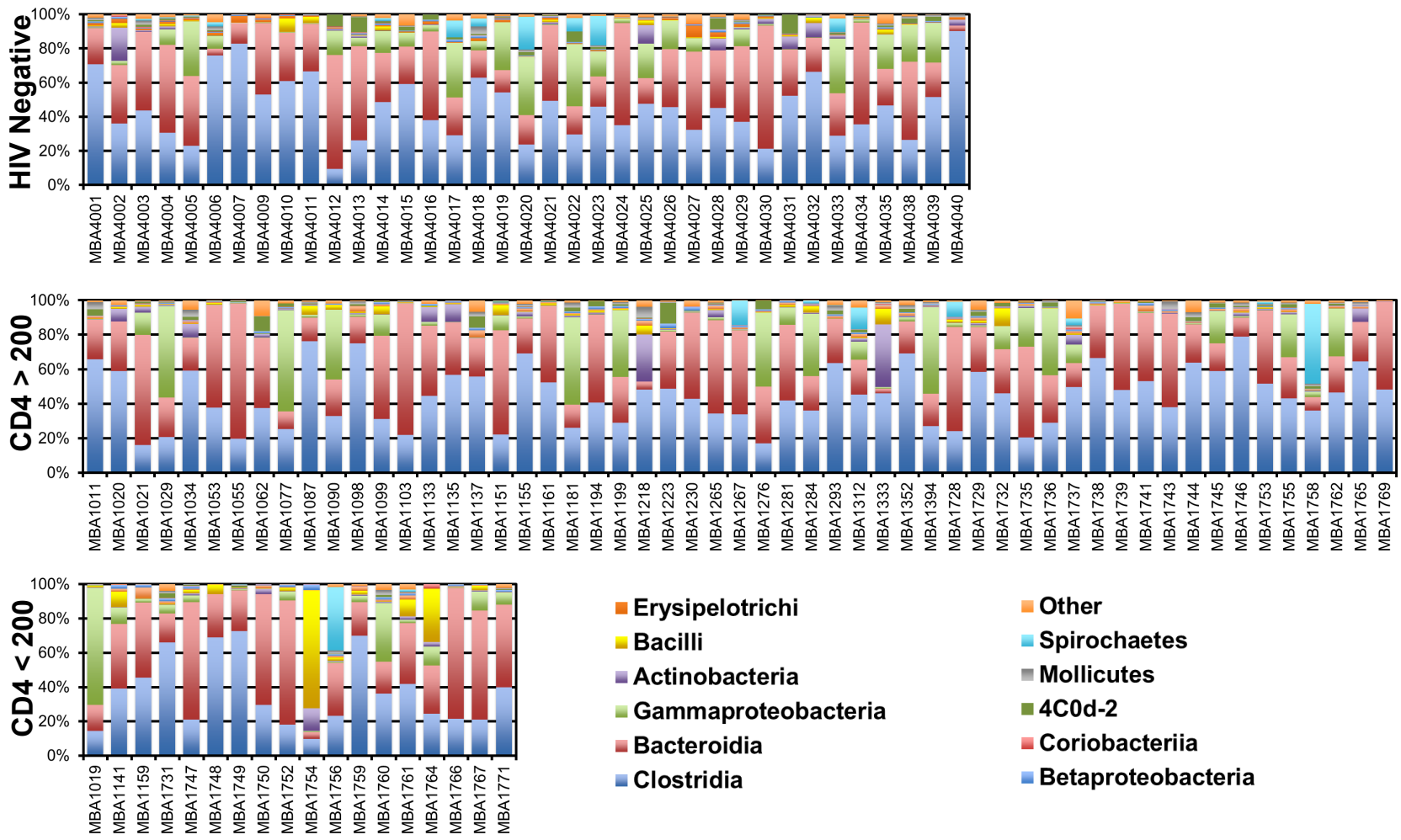


Figure S4, Related to Figure 4. Environmental and clinical associations with bacterial beta diversity.

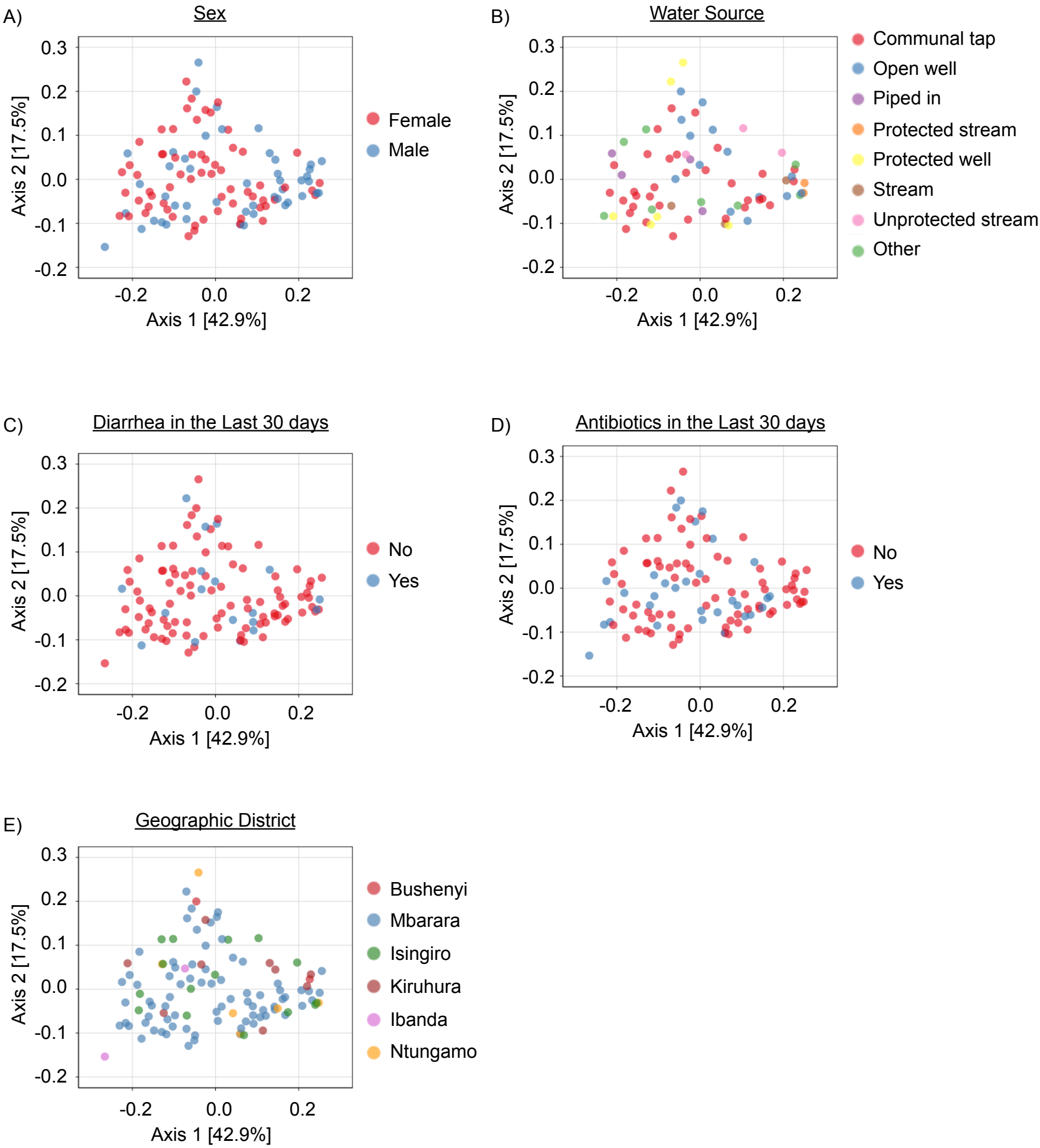
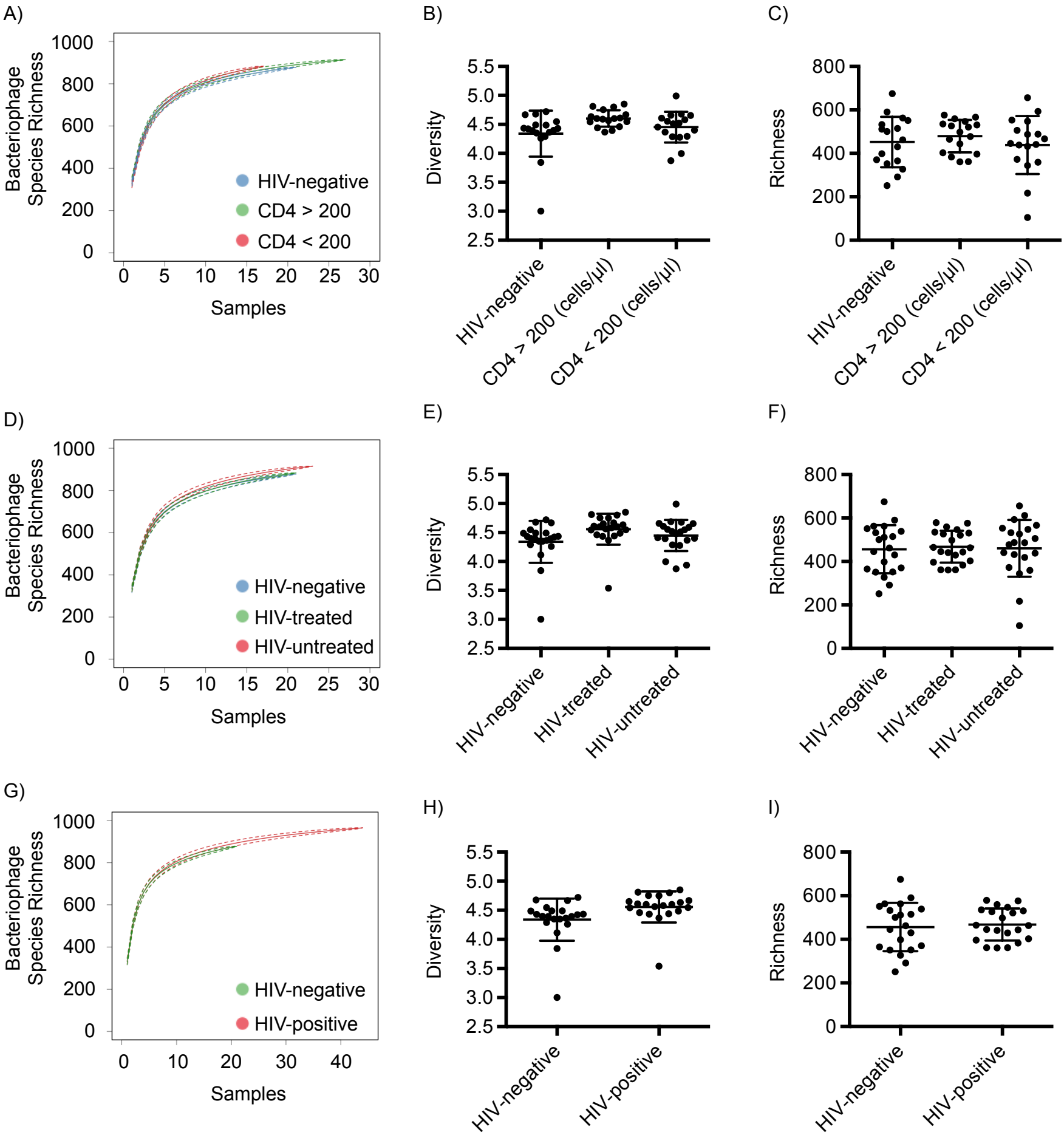


Figure S5, Related to Figure 3. Bacteriophage richness and diversity are unchanged in HIV.



SUPPLEMENTAL INFORMATION

SUPPLEMENTAL FIGURE LEGENDS

Figure S1, Related to Figure 2. Comparison of viral taxonomic assignment using blastx and VirusSeeker. (A) Dereplicated reads were compared using BLASTx to a virus protein database. The 20 most abundant virally-assigned taxa detected from all samples are shown, grouped at NCBI's taxonomic level of "family". (B) Comparison of the BLASTx (Bx) results in (A) to results obtained after running sequences through the VirusSeeker (Vs) pipeline. Only eukaryotic viruses and NCBI classifiable viral families are displayed. $p \leq 0.0001$ = ****. Median is indicated by the horizontal line. See also Table S1.

Figure S2, Related to Figure 3. Quantification and comparison of eukaryotic virus sequences detected in fecal samples. Abundance of (A, E, I) Adenoviridae, (B, F, J) Anelloviridae, (C, G, K) Circoviridae and (D, H, L) Papillomaviridae virus sequences were normalized to total quality-controlled sequences and analyzed by HIV infection status (top row) and ART therapy (middle row). ART-naïve, HIV-untreated patients were further subdivided by CD4 T cell number >200 (n=7) or <200 (n=16) in a subgroup analysis (bottom row) as compared to samples from HIV-negative or HIV-treated subjects. As only one sample in the HIV-treated group had a CD4 T cell number < 200, this group was not further subdivided. Sequences were normalized by dividing by the number of dereplicated (<95% identical), high-quality sequences. Statistical analysis was performed on untransformed data, and data was graphed after square root transformation. $p \leq 0.05$ = *. Bar indicates median.

Figure S3, Related to Figure 4: Bacterial class-level taxa barplot by CD4 T cell number. Relative abundance of bacterial taxa (y-axis) assigned to subjects (x-axis) as determined by 16S V4 sequencing was plotted and grouped by HIV Negative subjects (upper panel) and HIV-infected subjects with CD4 T cell number > 200 (middle panel) or <200 (lower panel). Color key is located at bottom right.

Figure S4, Related to Figure 4: Environmental and clinical associations with bacterial beta diversity. Principle Coordinate Analysis (PCoA) plots of the weighted UniFrac distances colored by (A) gender, (B) patient's home water source, (C) reported diarrhea or (D) antibiotics usage in the 30 days preceding sample collection, and (E) patient's home geographic district in Uganda.

Figure S5, Related to Figure 3: Bacteriophage Richness and Diversity are unchanged in HIV. Bacteriophage species accumulation plots rarefied by samples number over 1000 permutations were graphed by CD4 T cell number (A), HIV status and ART treatment group (D), and HIV infection status (G) with dotted lines representing 95% CI. Shannon diversity (y-axis) of bacteriophage species was determined and grouped by CD4 T cell count (B), HIV status and ART therapy (E), and HIV status (H). Species richness was determined and grouped by CD4 T cell count (C), HIV status and ART therapy (F), and HIV status (I). Bars indicate median ± interquartile range (IQR).

SUPPLEMENTAL TABLES:

Table S1, related to Table 1 and Figures 1: Viral sequence read statistics.

Ragon ID	HIV phenotype	Index	Number of Paired reads	Number of reads after Stitching	number of QCed reads	% of Stitched reads	Number of deduplicated reads	% of Stitched reads
MBA1011	HIV-treated	TGACCA	887278	1358660	1159458	85.34	130432	9.6
MBA1019	HIV-treated	TAGCTT	986347	1291422	1229675	95.22	385904	29.88
MBA1021	HIV-treated	GGCTAC	867024	1165106	983803	84.44	178015	15.28
MBA1029	HIV-treated	GATCAG	916813	1327735	1179217	88.81	164662	12.4
MBA1034	HIV-treated	TAGCTT	833915	1267056	1213408	95.77	230176	18.17
MBA1053	HIV-treated	CTTGTA	1092591	1389298	1277642	91.96	138327	9.96
MBA1055	HIV-treated	CAGATC	839016	1237461	1138972	92.04	528988	42.75
MBA1062	HIV-treated	ATTCCT	917080	1241436	1094998	88.2	201317	16.22
MBA1090	HIV-treated	ATCACG	1401128	1841681	1754951	95.29	492692	26.75
MBA1098	HIV-treated	CGATGT	1262741	1688080	1542193	91.36	745878	44.18
MBA1099	HIV-treated	TTAGGC	1085055	1432604	1312876	91.64	520828	36.36

MBA1103	HIV-treated	TGACCA	1222037	1701853	1295897	76.15	250226	14.7
MBA1133	HIV-treated	GGCTAC	1334658	1849467	1709706	92.44	728140	39.37
MBA1151	HIV-treated	GGCTAC	1817015	2431224	2259607	92.94	865595	35.6
MBA1155	HIV-treated	AGTCAA	2119670	2941809	2426746	82.49	578439	19.66
MBA1194	HIV-treated	CAGATC	158740	245201	232044	94.63	145416	59.3
MBA1218	HIV-treated	ACTTGA	950837	1270738	1136246	89.42	159421	12.55
MBA1267	HIV-treated	ATGTCA	1195091	1555410	1469115	94.45	300758	19.34
MBA1276	HIV-treated	CCGTCC	818417	1166397	1116503	95.72	352567	30.23
MBA1281	HIV-treated	GTCCGC	1074371	1423265	1348973	94.78	653313	45.9
MBA1312	HIV-treated	GTGAAA	1070296	1396835	1291352	92.45	192329	13.77
MBA1728	HIV-Untreated	GATCAG	1614597	2161450	2029126	93.88	772803	35.75
MBA1730	HIV-Untreated	TAGCTT	1374725	1820599	1712034	94.04	831928	45.7
MBA1731	HIV-Untreated	ATCACG	764253	997164	561687	56.33	39141	3.93
MBA1736	HIV-Untreated	CGATGT	653711	1123519	1085243	96.59	515153	45.85
MBA1738	HIV-Untreated	TTAGGC	815721	1061379	1028063	96.86	548074	51.64
MBA1742	HIV-Untreated	GTCCGC	2291496	3042169	2945985	96.84	1263523	41.53
MBA1747	HIV-Untreated	GTGAAA	1849711	2557903	2407033	94.1	1330295	52.01
MBA1749	HIV-Untreated	GCCAAT	53590	77144	67556	87.57	51043	66.17
MBA1750	HIV-Untreated	AGTTCC	890057	1562412	1512076	96.78	563318	36.05
MBA1751	HIV-Untreated	GGCTAC	1165771	1549874	1469432	94.81	976821	63.03
MBA1752	HIV-Untreated	GCCAAT	1036084	1346547	1050543	78.02	399374	29.66
MBA1753	HIV-Untreated	ACTTGA	949969	1401006	1307710	93.34	616029	43.97
MBA1754	HIV-Untreated	CAGATC	1174171	1595482	1503076	94.21	529697	33.2
MBA1755	HIV-Untreated	CGATGT	851350	1162091	1111754	95.67	566070	48.71
MBA1756	HIV-Untreated	CTTGTA	1297664	1852920	1475713	79.64	397954	21.48
MBA1758	HIV-Untreated	ACTTGA	989986	1358905	1206807	88.81	145747	10.73
MBA1759	HIV-Untreated	AGTCAA	1498550	2068138	1909440	92.33	738440	35.71
MBA1760	HIV-Untreated	CGTACG	970521	1255234	1176361	93.72	260012	20.71
MBA1761	HIV-Untreated	GAGTGG	1248156	1618701	1368571	84.55	183348	11.33
MBA1764	HIV-Untreated	ACTGAT	1384862	1808384	1731582	95.75	1069159	59.12
MBA1766	HIV-Untreated	GATCAG	864582	1152938	1033348	89.63	514394	44.62
MBA1767	HIV-Untreated	GGCTAC	785837	1170065	1106865	94.6	520997	44.53
MBA1771	HIV-Untreated	CTTGTA	857105	1313352	1234733	94.01	515975	39.29
MBA4001	HIV-negative	ACAGTG	1357848	1823927	1492501	81.83	198893	10.9
MBA4009	HIV-negative	GCCAAT	1465306	1953861	1693939	86.7	249601	12.77
MBA4010	HIV-negative	CAGATC	1453263	1879875	1625514	86.47	146245	7.78
MBA4011	HIV-negative	ACTTGA	1406838	1882138	1718968	91.33	161323	8.57
MBA4012	HIV-negative	AGTTCC	2296454	3183074	2920462	91.75	503355	15.81
MBA4013	HIV-negative	ATGTCA	2150661	2809970	2556579	90.98	887768	31.59
MBA4014	HIV-negative	CCGTCC	2103863	2769135	2653147	95.81	1559644	56.32
MBA4020	HIV-negative	TAGCTT	1089269	1415685	1380910	97.54	628801	44.42
MBA4021	HIV-negative	GGCTAC	874229	1178317	1132507	96.11	597540	50.71
MBA4024	HIV-negative	ACAGTG	929842	1275933	1143725	89.64	110449	8.66
MBA4025	HIV-negative	GCCAAT	888133	1354606	1258248	92.89	124114	9.16
MBA4026	HIV-negative	ATCACG	1072551	1471600	1280533	87.02	117249	7.97
MBA4027	HIV-negative	CGATGT	1022900	1356081	1276535	94.13	628308	46.33
MBA4028	HIV-negative	TTAGGC	785929	1334787	947717	71	257697	19.31
MBA4029	HIV-negative	AGTCAA	1044097	1393242	1094090	78.53	64319	4.62
MBA4030	HIV-negative	GTGGCC	950277	1225198	1151439	93.98	538827	43.98
MBA4031	HIV-negative	ATCACG	1217994	1851477	1781073	96.2	692072	37.38
MBA4037	HIV-negative	GTTTCG	1015898	1336728	1261423	94.37	711598	53.23
MBA4038	HIV-negative	TGACCA	806068	1105304	858404	77.66	152637	13.81
MBA4039	HIV-negative	ACAGTG	953760	1261375	1159538	91.93	596567	47.29
MBA4040	HIV-negative	TTAGGC	897440	1342816	1253838	93.37	483619	36.02

Table S2, Related to Figure 2B: Papillomaviruses identified in the fecal sample collected from MBA1759, an HIV-infected, untreated subject.

#	Contig Length (bp)	Fold Cove r-age	Complete Genome Length (bp)	Genus	Most Closely Related HPV strain	Accesion Number	Reference Genome Length (bp)	Sequence identity to Reference Genome (nt)
1	9325	654.5	7362	Gammapapillomavirus	Type 103	NC_008188.1	7263	7259/7263 (99%)
2	9055	9.7	7857	Alphapapillomavirus	Type 18	KC470213.1	7857	7850/7857 (99%)
3	8414	732.4	7790	Alphapapillomavirus	Type 56	EF177178.1	7790	7779/7790 (99%)
4	8142	468.7	8015	Alphapapillomavirus	Type 61	HPU31793	7989	7893/8015 (98%)
5	8025	12.3	7890	Alphapapillomavirus	Type 59	KC470262.1	7899	7890/7898 (99%)
6	8008	31.2	7882	Alphapapillomavirus	Type 30	X74474.1	7852	7781/7890 (99%)
7	7985	50.6	7858	Alphapapillomavirus	Type 45	EF202156.1	7858	7854/7858 (99%)
8	7966	5.1	7839	Alphapapillomavirus	Type 33	HQ537702.1	7838	7834/7839 (99%)
9	7942	233.8	7815	Alphapapillomavirus	Type 51	M62877.1	7808	7728/7816 (99%)
10	7844	116	7717	Alphapapillomavirus	Type 54	AF436129.1	7717	7711/7717 (99%)
11	7833	66.5	7706	Alphapapillomavirus	Type 73	X94165.1	7700	7619/7706 (99%)
12	7805	148.7	7678	Alphapapillomavirus	Type 69	AB027020.1	7700	7616/7705 (99%)

HPV: human papillomavirus

Table S3, Related to Figure 2D: Number of ORF-1 verified anellovirus contigs per sample.

Ragon ID	**CD4 group (cells/ml)**	**ORF-1 verified contigs (n)**
MBA4024	HIV Negative	4
MBA4026	HIV Negative	2
MBA1758	Greater than 200	4
MBA1731	Less than 200	5
MBA1747	Less than 200	6
MBA1751	Less than 200	1
MBA1752	Less than 200	7
MBA1761	Less than 200	19
MBA1759	Less than 200	3

Table S4, Related to Figure 3: Association between presence of specific viral sequences and patients with low CD4 T cells (< 200).

	Overall	**CD4 < 200 vs. HIV negative**			**CD4 < 200 vs. CD4 > 200**		
Virus	**Fisher's exact *p* value**	**Fisher's exact *p* value**	**FDR *p* value**	**Odds Ratio**	**Fisher's exact *p* value**	**FDR *p* value**	**Odds Ratio**
Adenoviridae	0.0031**	0.0062**	0.0129*	8.022	0.0064**	0.0129*	7.746
Anelloviridae	0.0409*	0.0990	0.0990	3.445	0.0238*	0.0317*	4.796
Circoviridae	0.8746	NA	NA	NA	NA	NA	NA
Papillomaviridae	0.1253	NA	NA	NA	NA	NA	NA

$p \leq 0.05$ = *, $p \leq 0.01$ = **

Table S5, Related to Figure 5: Differentially abundant bacterial taxa in subjects with low CD4 T cell counts and bacterial OTUs significantly associated with sCD14.
Tab 1: Differentially abundant bacterial taxa in subjects with CD4 <200 compared to HIV-negative subjects.
Tab 2: Differentially abundant bacterial taxa in subjects with CD4 <200 compared to subjects with CD4 > 200.
Tab 3: Bacterial OTUs significantly associated with sCD14.

Table S6, Related to Figure 5: OTUs significantly associated with CD4 T cell number using a multivariate model including CD4 T cell group, age, month stool collected, sequencing run, BMI, and other antibiotic use

Phyla	Class	Order	Family	Genus species	OTU ID	Raw *p* value	FDR *p* value
Firmicutes	Clostridia	Clostridiales	Clostridiaceae	NA	575768	9.52E-05	0.021052
Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Oscillospira	187267	0.000124	0.025939
Bacteroidetes	Bacteroidia	Bacteroidales	NA	NA	343635	0.000202	0.035365
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	NA	782953	1.99E-05	0.006693
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	NA	1111294	2.18E-05	0.007057
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	NA	292289	2.27E-05	0.007059
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	NA	1108656	2.43E-05	0.007285
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	NA	4354477	8.57E-05	0.020022
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Cronobacter dublinensis	667570	0.00016	0.030496
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	NA	581782	0.000235	0.039434
Firmicutes	Clostridia	Clostridiales	NA	NA	4480176	0.000295	0.047616
Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus parainfluenzae	4347099	0.000308	0.047998
Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	NA	199182	0.00033	0.048476
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	888300	0.000331	0.048476

Table S7, Related to Figure 4: Correlations between bacteriophage and bacterial richness and diversity by CD4 T cell grouping

Sample Group	Richness		Shannon Diversity	
	Spearman's *r*	*p* value	Spearman's *r*	*p* value
HIV-negative	-0.1687	0.4770	0.1378	0.5624
CD4 < 200	-0.4210	0.7550	-0.1093	0.6981
CD4 > 200	0.0123	0.9524	0.1600	0.4348

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

sCD14 Measurements in Plasma

Plasma was collected in acid citrate dextrose tubes and aliquots underwent a single freeze-thaw cycle before analysis. sCD14 concentration was measured by ELISA (R&D Systems Human sCD14 Quantikine ELISA kit #DC140). Samples were thawed on ice, centrifuged for 10 minutes at 1,000 x *g* and 4°C, and the supernatant was diluted 200-fold in Calibrator Diluent RD5P (1X, R&D Systems) per the manufacturer's instructions (10 µL sample + 1990 µL Calibrator Diluent RD5P). Samples were tested in duplicate with standards on every plate. sCD14

concentration was determined from optical density (O.D.) measurements by subtracting the average zero standard O.D. and performing a log-log transformation of the standards in order to fit a linear regression.

Virus-Like Particle Enrichment and Sequencing

Virus-Like Particle Preparation

VLPs were enriched from pulverized human stool as previously described (Reyes et al., 2010; Thurber et al., 2009). Approximately 200mg of stool was suspended in 400ul saline-magnesium buffer (0.1M NaCl, 0.008M MgSO4·7H2O, 0.002% gelatin, 0.05M Tris pH7.5) by vortexing for 10 minutes. Stool suspensions were then cleared by centrifugation at 2000 x *g* to remove debris and cells. Clarified suspensions were passed through one 0.45µm filter followed by two 0.22µm filters to remove residual host and bacterial cells. Samples were treated with lysozyme (1µg/ml at 37°C for 30 minutes) followed by chloroform (0.2x volume at RT for 10 minutes) to degrade remaining bacterial and host cell membranes. A DNase cocktail (10U Tubro DNaseI (Ambion), 1U Baseline zero DNase (Epicentre)) was used to remove contaminating host and bacterial DNA, followed by heat inactivation of DNases at 65°C for 10 minutes. VLPs were lysed (3.8% SDS plus 38µg/ml Proteinase K at 56°C for 20 minutes), treated with CTAB (2.5% CTAB plus 0.5M NaCl at 65°C for 10 minutes), and nucleic acid was extracted with phenol:chloroform pH 8.0 (Invitrogen). The aqueous fraction was washed once with an equal volume of chloroform and purified using DNeasy Blood and Tissue kit column (Qiagen). VLP DNA was amplified for 2 hours using Phi29 polymerase (GenomiPhi V2 kit, GE Healthcare) prior to sequencing. To reduce amplification bias, four independent reactions with 2µl of template were pooled for each sample (Reyes et al., 2010). Six samples failed amplification. Amplified VLP DNA (200ng) was fragmented by ultra-sonication (Covaris E210) before library construction (NEBNext Ultra DNA kit, New England Biolabs). Equimolar pools (ca. 12 samples/run) were sequenced on Illumina MiSeq platform (Washington University Center for Genome Sciences; 2 x 250bp run, loading at around 7pM, 1% PhiX spike-in) generating an average of over 1 million sequences per sample.

Assignment of VLP Reads Taxonomy

Detection of potentially ambiguous or false-positive viral sequences was done using VirusSeeker, a custom bioinformatics pipeline designed to detect sequences sharing nucleotide and protein level sequence similarity to known viruses (Zhao et al., submitted). Briefly, sequences are adapter-trimmed, quality controlled and dereplicated (removing sequences that are >95% identical). Potential unique viral reads were queried against the NCBI nt/nr databases, and only reads matching exclusively to viral sequences were kept for further analysis. All sequences aligning to viruses were further classified into viral families based on the NCBI taxonomic identity of the best hit.

VLP Sequence Analysis

Data were normalized by dividing individual taxon-assigned sequence counts by the total deduplicated quality controlled reads in a sample. Ecological parameters including richness and diversity were calculated using the diversityresult function of BiodiversityR package (Kindt, 2005). Heatmaps were generated using the vegan R package (Oksanen et al., 2013), and the heatmap.2 function of the gplots R packages (Warnes et al., 2015).

Virus contig analysis

Contigs were *de novo* assembled from paired-end reads for each sample using the SPAdes assembler (v 2.5.1) with kmer lengths of 77, 99, and 127 (Bankevich et al., 2012) and a minimum contig length of 500 bp. Sequences were deduplicated using CD-HIT (Fu et al., 2012; Li and Godzik, 2006) at 95% nucleotide identity and 95% length overlap cutoff. To determine the presence of *Adenoviridae*, *Anelloviridae*, *Circoviridae* and *Papillomaviridae* within the assembled sequence data, a blast database was created using reference genome sets by searching the NCBI nucleotide database for the selected family and narrowing search results by adding the filters “virus”, “genomic DNA/RNA”, and “RefSeq database”. tBlastx with a bit score cutoff of 100 was used to determine contigs with significant sequence similarity to reference genomes. Contigs were parsed in MEGAN (v5.8) (Huson et al., 2011) using the lowest-common ancestor algorithm with the following settings: Min Support: 1, Min Score: 100, Max Expected: 1e-10, Top Percent: 10.0, Min Complexity: 0.44 with the Min-Complexity Filter turned off. Candidate contig sequences were queried against the NCBI nr/nt databases to identify viral sequences that share similarity with only viruses.

Anellovirus Phylogenetic Analysis

Anellovirus-assigned contigs sharing greater than 95% sequence identity were deduplicated using the BBTools package (http://sourceforge.net/projects/bbmap/) and aligned to a conserved region of ORF1 (amino acid position 69

– 272 of reference TTV1 ORF1, NC_002076) of 40 representative anellovirus genomes using MUSCLE (Edgar, 2004). Phylogenetic trees were constructed from an amino acid alignment with the following reference anelloviruses: (alphatorquevirus) TTV LTT6 (EU305674), TTV 1 (NC_002076), TTV 2 (NC_014480), TTV 3 (NC_014081), TTV 4 (NC_014069), TTV 6 (NC_014094), TTV 7 (NC_014080), TTV 8 (NC_014084), TTV 10 (NC_014076), TTV 12 (NC_014075), TTV 14 (NC_014077), TTV 15 (NC_014096), TTV 16 (NC_014091), TTV 19 (NC_014078), TTV 25 (NC_014083), TTV 26 (NC_014079), TTV 27 (NC_014074), TTV 28 (NC_014073); (betatorquevirus) TTmV 6 (NC_014095), TTmV 7 (NC_014082), TTmV 9 (NC_002195); (gammatorquevirus) small anellovirus 1 (NC_007013), TTmV 1 (NC_009225), TTmV 2 (NC_014093). Maximum likelihood phylogenies were constructed with PhyML (version 20120412) (Guindon and Gascuel, 2003) using the LG substitution model. A discrete γ distribution of 4 rate categories was used to model heterogeneity among sites and support was assessed by 1000 nonparametric bootstraps. Phylogenies were visualized with FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Torque Teno Virus Real-Time PCR

TaqMan quantitative real-time PCR was performed to detect and quantify anellovirus in MDA amplified samples (Maggi et al., 2003; Walton et al., 2014) using primers specific for alphatorqueviruses. All reactions were performed in at least duplicate in a blinded manner. Each reaction was performed in 25ul total volume including 5ul of MDA product, AmpliTaq gold (Life Technologies), 0.9uM AMTS forward primer (5' GTGCCGIAGGTGAGTTTA 3'; I is inosine), 0.9uM AMTAS reverse primer (5' AGCCCGGCCAGTCC 3'), 250nM AMTPTU probe 5' 6-FAM/ZEN/3IBFQ (5' TCAAGGGGCAATTCGGGCT 3'; Integrated DNA Technologies). The primers and probe target a highly conserved TTV genome segment described previously (Maggi et al., 2003; Walton et al., 2014). Cycling conditions were as follows: 50C for 2 minutes, 95C for 10 minutes then 40 cycles of 95C for 15 sec and 60C for 1 minutes. Reactions were performed on a StepOne Plus real-time PCR system (Applied Biosystems) and analyzed with StepOne Software v2.3. A plasmid control containing the TTV target was synthesized and used to generate a standard curve. The linear range of the assay was from 2e9 to 2e3 copies/ml. When both PCR reactions were positive we counted this as a positive (n = 31). When both were negative we counted this as a negative (n = 34). When there were disparate results from the two reactions, we called these indeterminate (n = 9). For purposes of analysis we took the conservative approach of requiring both reactions to be positive to call a sample 'present' for anellovirus. For statistical purposes we therefore grouped negative and indeterminate together.

Bacterial 16S rRNA Analysis

Stool Pulverization

Aliquots of pulverized human stool (100-200mg) were processed as previously described (Reyes et al., 2013). Briefly, stool was chipped from RNAlater on liquid nitrogen, samples were pulverized, aliquotted (approximately 200mg stool each) into 2-3 separate 2mL collections tubes (Starstedt) and stored at -80°C until use. Aliquots were used for total nucleic acid (TNA) extraction and VLP preparation, ensuring that similar parts of the stool samples were used for both extraction methods.

Human Stool Total Nucleic Acid Extraction

Stool TNA was extracted from aliquots of pulverized human stool (~200mg) as previously described (Reyes et al., 2013) with modification. Briefly, 200µl of 1 mm diameter zirconia/silica beads (Biospec) were added to individual pulverized stool aliquots. 500µL of phenol:chloroform:isoamyl alcohol (Fisher Scientific, 25:24:1, pH 8.0), 500µL of 0.2µm-filtered 2x Buffer A (200mM NaCl, 200mM Tris, 20mM EDTA), and 210µL of 20% SDS were added to each sample. Samples were chilled on ice and homogenized using the highest setting on a BioSpec Mini-Beadbeater for 2 minutes at 4ºC. The homogenized samples were then centrifuged at 4ºC for 3 minutes at 7000 x *g*, and the aqueous phase was transferred to a clean tube. An equal volume of phenol:chloroform:isoamyl alcohol was added and mixed by vortexing. Samples were centrifuged at 16,000 x *g* for 5 minutes at room temperature and the aqueous phase transferred to a clean tube. Nucleic acid was precipitated with isopropanol and 3M sodium acetate (pH 5.5, Ambion) at -80°C for 20 minutes, then spun at maximum speed at 4°C for 30 minutes. The pellet was washed with 500µl 100% ethanol, centrifuged at 16,000 x *g* for 15 minutes at 4ºC, dried, and resuspended in 200ul of molecular grade Tris-EDTA buffer (Ambion). DNA was isolated from the total nucleic acid preparation using an AllPrep DNA/RNA Micro Kit (Qiagen) according to the manufacturer's instructions. Nine samples resulted in insufficient quantity of DNA for 16S studies.

16S rRNA Amplification and Sequencing

Primer selection and polymerase chain reaction was performed as described previously (Caporaso et al., 2011). Briefly, each sample was amplified in triplicate, pooled, and confirmed by gel electrophoresis. PCR reactions contained 2.5µL 10X High Fidelity PCR Buffer (Invitrogen), 18.8µL RNase/DNase-free water, 0.5µL 10 mM dNTPs, 1µL 50mM MgSO4, 0.5µL each of the forward and reverse Golay-barcoded primers specific for the V4 region (F515/R806, 10µM final concentration), 0.1µL Platinum High Fidelity Taq (Invitrogen) and 3µL extracted total nucleic acid. Reactions were held at 94°C for 2 minutes to denature the DNA, with amplification for 26 cycles of 94°C for 15s, 50°C for 30s, and 68°C for 30s; a final extension of 2 minutes at 68°C (to ensure complete amplification). Amplicons were pooled and purified using 0.6x Agencourt Ampure XP beads (Beckman-Coulter) according to the manufacturer's instructions. The final pooled samples were sequenced on the Illumina MiSeq platform (Washington University Center for Genome Sciences; 2x250 standard run) in two separate runs.

16S rRNA Analysis

Analysis of R1 16S sequence data was performed using QIIME (Quantitative Insights Into Microbial Ecology, version 1.9.1) (Caporaso et al., 2010). Raw sequence fastq files were quality filtered and demultiplexed using default parameters with the following exceptions: PHRED quality score cut-off at 20, and reverse-complement mapping barcodes were used. Closed reference operational taxonomic units (OTUs) sharing 97% identity were clustered using the UCLUST algorithm (Edgar, 2010) and assigned taxonomy according to the Greengenes database (version 13.8) (McDonald et al., 2012). To standardize differences in the number of OTUs between sequencing runs, all samples were rarefied to 5000 OTUs (10 iterations without replacement; maximum of 5000 OTUs per sample; 10 rarefaction steps) and the relative number of sequences assigned to each OTU was calculated for each sample. Two samples did not achieve high enough OTUs for downstream analysis. Alpha diversity analysis was performed on rarefied data. Faith's phylogenetic diversity (Faith and Baker, 2006) and the Chao1 richness metric were calculated for all ten rarefied tables. Statistical analysis between groups was performed using the compare_alpha_diversity.py function of QIIME. Species accumulation rarefactions plots were determined using the specaccum function of the vegan R package (Oksanen et al., 2013). Beta-diversity was determined in Phyloseq (v1.10.0) (McMurdie and Holmes, 2012) using weighted UniFrac distances. Differential abundance of bacterial taxa between experimental groups was determined using the PhyloSeq DESeq2 extension using the Wald significance test and a parametric fit type (v.1.6.3) (Anders and Huber, 2010; McMurdie and Holmes, 2012).

Oligotyping

We performed oligotyping analyses (Eren et al., 2011) on differentially abundant 16S V4 sequencing reads assigned by QIIME to the *Ruminococcus* genera or *Enterobacteriaceae* family that were not previously resolved at the species level. Sequences shorter than the indicated length when trimmed to Phred score >30 were removed before analysis to prevent excessive variation due to sequencing error. Representative sequences for each oligotype were searched in the BLAST nr/nt database. The following table details the parameters for each performance of oligotyping. Representative sequences for each oligotype are indicated below.

Oligotyping Group	Minimum Read Length	Total Full-Length Reads	Reads Assigned to Oligotypes	Minimum Sequences per Oligotype (-A parameter)	Oligotype Base Locations of Interest	Taxa assigned
PhyloSeq, *Ruminococcus* sp. enriched in CD4 >200 vs. CD4 <200	230	191993	134162 (0.699)	2000	25 Bases: 0, 8, 9, 12, 57, 58, 68, 79, 94, 95, 98, 112, 158, 174, 177, 178, 181, 201, 212, 229, 232, 236, 237, 242, 249	*R. bromii* *R. callidus*
PhyloSeq, *Ruminococcus* sp. enriched in HIV-negative vs. CD4 <200	0	109848	80909 (0.737)	2000	19 Bases: 0, 9, 77, 98, 148, 158, 174, 178, 181, 201, 212, 225, 229, 232, 236, 237, 238, 242, 249	*R. bromii*
MaAsLin, *Enterobacteriaceae* enriched in CD4 <200	240	111936	100009 (0.893)	1000	15 Bases: 0, 28, 57, 113, 138, 183, 189, 223, 226, 228, 232, 238, 242, 243, 249	*Shigella* sp. or a closely-related *Escherichia* sp.

PhyloSeq, *Ruminococceae Ruminococcus* sp. enriched in CD4 >200 vs. CD4 <200

>Oligotype TGAATTAAAATTTGGAACTAACGTC
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGGCTTTGCAAGTCAG
ATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACTGTAGAGCTTGAGTGAAGTAGAGGCA
GGCGGAATTCCCCGTGTAGCGGTGAAATGCGTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGCC
TGCTGGGCTTTAACTGACGCTGAGGCACGAAAGCGTGGGTAGCAAAC

> Oligotype TGAGTGAAAGTTTCGAACTTACGGC
TACGTAGGGAGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGTAGGCGGGATGGCAAGTCA
GATGTGAAAACTATGGGCTCAACCCATAGACTGCATTTGAAACTGTTGTTCTTGAGTGAGGTAGAGGT
AAGCGGAATTCCTGGTGTAGCGGTGAAATGCGTAGAGATCAGGAGGAACATCGGTGGCGAAGGCGGC
TTACTGGGCCTTTACTGACGCTGAGGCTCGAAAGCGTGGGGAGCAAAC

> Oligotype TGAGCGAAAGTTTCGAACTTACGGC
TACGTAGGGAGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGTAGGCGGGACGGCAAGTCA
GATGTGAAAACTATGGGCTCAACCCATAGACTGCATTTGAAACTGTTGTTCTTGAGTGAGGTAGAGGT
AAGCGGAATTCCTGGTGTAGCGGTGAAATGCGTAGAGATCAGGAGGAACATCGGTGGCGAAGGCGGC
TTACTGGGCCTTTACTGACGCTGAGGCTCGAAAGCGTGGGGAGCAAAC

> Oligotype TGAAAAAAGATTTCGAACTAACGTC
TACATAGGGAGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGGCTAAGCAAGTCA
GATGTGAAATACACGGGCTCAACCCGTGAGCTGCATTTGAAACTGTTTAGCTTGAGTGAAGTAGAGGC
AGGCGGAATTCCCGGTGTAGCGGTGAAATGCGTAGAGATCGGGAGGAACACCAGTGGCGAAGGCGGC
CTGCTGGGCTTTAACTGACGCTGAGGCACGAAAGCGTGGGTAGCAAAC

> Oligotype TTGGAGAAAATTTTGAACTTACGGC
TACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGTGTAGGCGGGAAGGCAAGTCA
GAAGTGAAAATTATGGGCTTAACCCATAACCTGCTTTTGAAACTGTTTTTCTTGAGTGAGGCAGAGGC
AAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGC
TTGCTGGGCCTTTACTGACGCTGAGGCTCGAAAGCGTGGGGAGCAAAC

> Oligotype TGAGTGAAAGTCTCGAACTTACGGC
TACGTAGGGAGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGTAGGCGGGATGGCAAGTCA
GATGTGAAAACTATGGGCTCAACCCATAGACTGCATTTGAAACTGCTGTTCTTGAGTGAGGTAGAGGT
AAGCGGAATTCCTGGTGTAGCGGTGAAATGCGTAGAGATCAGGAGGAACATCGGTGGCGAAGGCGGC
TTACTGGGCCTTTACTGACGCTGAGGCTCGAAAGCGTGGGGAGCAAAC

> Oligotype TGAATTAAAATTTGGAACTAACGTA
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGGCTTTGCAAGTCAG
ATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACTGTAGAGCTTGAGTGAAGTAGAGGCA
GGCGGAATTCCCCGTGTAGCGGTGAAATGCGTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGCC
TGCTGGGCTTTAACTGACGCTGAGGCACGAAAGCGTGGGTAGCAAAA

> Oligotype TGAATTAAAATTTGGAACTAACGGC
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGGCTTTGCAAGTCAG
ATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACTGTAGAGCTTGAGTGAAGTAGAGGCA
GGCGGAATTCCCCGTGTAGCGGTGAAATGCGTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGCC
TGCTGGGCTTTAACTGACGCTGAGGCACGAAAGCGTGGGGAGCAAAC

PhyloSeq, *Ruminococceae Ruminococcus* sp. enriched in HIV-negative vs. CD4 <200

>Oligotype TACTCTGAACTAAACGTTC
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGGCTTTGCAAGTCAG
ATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACTGTAGAGCTTGAGTGAAGTAGAGGCA

GGCGGAATTCCCCGTGTAGCGGTGAAATGCGTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGCC
TGCTGGGCTTTAACTGACGCTGAGGCACGAAAGCGTGGGTAGCAAAC

>Oligotype TAATGTCAACTAAACGTTC
TACATAGGGAGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGGCTAAGCAAGTCA
GATGTGAAATACACGGGCTCAACCCGTGAGCTGCATTTGAAACTGTTTAGCTTGAGTGAAGTAGAGGC
AGGCGGAATTCCCGGTGTAGCGGTGAAATGCGTAGAGATCGGGAGGAACACCAGTGGCGAAGGCGGC
CTGCTGGGCTTTAACTGACGCTGAGGCACGAAAGCGTGGGTAGCAAAC

>Oligotype TACTCTGAACTAAACGTTA
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGGCTTTGCAAGTCAG
ATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACTGTAGAGCTTGAGTGAAGTAGAGGCA
GGCGGAATTCCCCGTGTAGCGGTGAAATGCGTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGCC
TGCTGGGCTTTAACTGACGCTGAGGCACGAAAGCGTGGGTAGCAAAA

>Oligotype TACTCTGAACTAAACGTGC
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGGCTTTGCAAGTCAG
ATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACTGTAGAGCTTGAGTGAAGTAGAGGCA
GGCGGAATTCCCCGTGTAGCGGTGAAATGCGTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGCC
TGCTGGGCTTTAACTGACGCTGAGGCACGAAAGCGTGGGGAGCAAAC

>Oligotype TATGATTAATCAAACATGC
TACGTAGGGAGCGAGCGTTGTCCGGAATTATTGGGTGTAAAGGGTGCGTAGGCGGCTATGTAAGTCAG
GCGTGTAATTCAGAGGCTTAACCTCTTGACGGCGCTTGAAACTGTGTAGCTTGAGTGGAGTAGAGGCA
GATGGAATTTCCAGTGTAGCGGTGAAATGCGTAGATATTGGAAGGAACATCGGTGGCGAAGGCGATC
TGCTGGGCTCTAACTGACGCTGAGGCACGAAAGCATGGGGAGCAAAC

>Oligotype TACTCTGAATTAAACGTTC
TACGTAGGGAGCAAGCGTTGTCCGGATTTACTGGGTGTAAAGGGTGCGTAGGCGGCTTTGCAAGTCAG
ATGTGAAATCTATGGGCTCAACCCATAAACTGCATTTGAAACTGTAGAGCTTGAGTGAAGTAGAGGCA
GGCGGAATTCCCCGTGTAGCGGTGAAATGCGTAGAGATGGGGAGGAACACCAGTGGCGAAGGCGGTC
TGCTGGGCTTTAACTGACGCTGAGGCACGAAAGCGTGGGTAGCAAAC

>Oligotype TATTATTAATTAAACATGC
TACGTAGGGAGCGAGCGTTGTCCGGAATTATTGGGTGTAAAGGGTGCGTAGGCGGCTATGTAAGTCAG
GCGTGTAATTCAGAGGCTTAACCTCTTGACTGCGCTTGAAACTGTGTAGCTTGAGTGGAGTAGAGGCA
GATGGAATTTCCAGTGTAGCGGTGAAATGCGTAGATATTGGAAGGAACATCGGTGGCGAAGGCGATC
TGCTGGGCTTTAACTGACGCTGAGGCACGAAAGCATGGGGAGCAAAC

MaAsLin, *Enterobacteriaceae* enriched in CD4 <200

>Oligotype TTTCTTTTGTATGAC
TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAG
ATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGG
GGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCC
CCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAAC

>Oligotype TTTCTTTTGTATGAA
TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAG
ATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGG
GGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCC
CCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAAA

>Oligotype CTTCTTTTGTATGAC
CACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCA
GATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGG

GGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGC
CCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAAC

>Oligotype TTGTTTTTGTATGAC
TACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTGATTAAGTCA
GATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGTCAGCTTGAGTCTCGTAGAGGG
GGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGC
CCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAAC

Statistical Analysis

Descriptive measures were used to summarize the data. Continuous variables were summarized using median and IQR; categorical variables were summarized using frequency and percent (%). Spearman's rank correlations were used to examine bivariate associations between study variables. Fisher's exact and Chi-square tests were used to compare categorical variables between the study groups. Mann-Whitney test and Kruskal Wallis test (indicated by *p*-value in text) with Dunn's post hoc analyses (*p*-values in figures) were used for comparing continuous variables. PhyloSeq (v1.10.0) (McMurdie and Holmes, 2012) was used to calculate UniFrac distances between 16S samples and to perform principal coordinate analysis. MaAsLin (http://huttenhower.sph.harvard.edu/maaslin) was used for multivariate modeling by importing the relative bacterial abundance values and associated sample metadata. Minimum for feature relative abundance filtering was set to 1e-6, maximum false discovery rate at 0.05, minimum for feature prevalence filtering set to 0.1, and Benjamini-Hochberg FDR (BH-FDR) protocol was used for multiple comparison correction. Statistical significance of distance and dissimilarity metrics (beta-diversity) between groups was determined by PERMANOVA using the adonis function of QIIME. Differential abundance of bacterial taxa between experimental groups was determined using the PhyloSeq DESeq2 extension using the Wald significance test and a parametric fit type (v.1.6.3) (Anders and Huber, 2010; McMurdie and Holmes, 2012) with multiple comparison correction using BH-FDR. No correction for multiple comparisons was performed unless otherwise stated. Statistical analyses and graphing were performed in R (Team, 2013) and Prism version 6.05 for Windows (GraphPad Software, La Jolla California USA, www.graphpad.com). All p-values were two-sided and $p < 0.05$ was considered significant.

SUPPLEMENTAL REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., *et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology : a journal of computational molecular cell biology *19*, 455-477.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. Nature methods *7*, 335-336.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences of the United States of America *108*, 4516-4522.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460-2461.

Eren, A.M., Zozaya, M., Taylor, C.M., Dowd, S.E., Martin, D.H., and Ferris, M.J. (2011). Exploring the diversity of Gardnerella vaginalis in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. PloS one *6*, e26732.

Faith, D.P., and Baker, A.M. (2006). Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. Evolutionary Bioinformatics Online *2*, 121-128.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150-3152.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. SystBiol *52*, 696-704.

Huson, D.H., Mitra, S., Ruscheweyh, H.-J., Weber, N., and Schuster, S.C. (2011). Integrative analysis of environmental sequences using MEGAN4. Genome research *21*, 1552-1560.

Kindt, R.C., R. (2005). Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies. (Nairobi, World Agroforestry Centre (ICRAF)).
Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658-1659.
Maggi, F., Pifferi, M., Tempestini, E., Fornai, C., Lanini, L., Andreoli, E., Vatteroni, M., Presciuttini, S., Pietrobelli, A., Boner, A.*, et al.* (2003). TT virus loads and lymphocyte subpopulations in children with acute respiratory diseases. Journal of virology *77*, 9081-9083.
McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., and Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME journal *6*, 610-618.
McMurdie, P.J., and Holmes, S. (2012). Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data. Pac Symp Biocomput, 235-246.
Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H.*, et al.* (2013). vegan: Community Ecology Package. R package version 2.0-10.
Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., and Gordon, J.I. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature *466*, 334-338.
Reyes, A., Wu, M., McNulty, N.P., Rohwer, F.L., and Gordon, J.I. (2013). Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. Proceedings of the National Academy of Sciences of the United States of America *110*, 20236-20241.
Team, R.C. (2013). R: A language and environment for statistical computing (Vienna, Austria, R Foundation for Statistical Computing).
Thurber, R.V., Haynes, M., Breitbart, M., Wegley, L., and Rohwer, F. (2009). Laboratory procedures to generate viral metagenomes. Nature protocols *4*, 470-483.
Walton, A.H., Muenzer, J.T., Rasche, D., Boomer, J.S., Sato, B., Brownstein, B.H., Pachot, A., Brooks, T.L., Deych, E., Shannon, W.D.*, et al.* (2014). Reactivation of multiple viruses in patients with sepsis. PloS one *9*, e98819.
Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S.*, et al.* (2015). gplots: Various R Programming Tools for Plotting Data. . R package *version 2.17.0*.