Fig. S1: *Weighted UniFrac clustering of HIV dataset.*
Each point represents a single fecal sample. Panels A and B represent the same PCoA analysis but for clarity, Panel A only shows the HIV positive individuals who were not on ART at the time of initial sample collection and HIV-negative controls, and Panel B shows only the HIV positive individuals on ART. Points are labeled with ART duration at the time of sample collection in months. Samples from the same individual at two different timepoints are joined with a line. Bacterial families are plotted as a weighted average of the coordinates of all samples, where the weights are the relative abundances of the taxon in the samples (grey circles) using the Biplot functionality of QIIME. The size of the sphere representing a taxon is proportional to the mean relative abundance of the taxon across all samples. Related to Fig. 1.

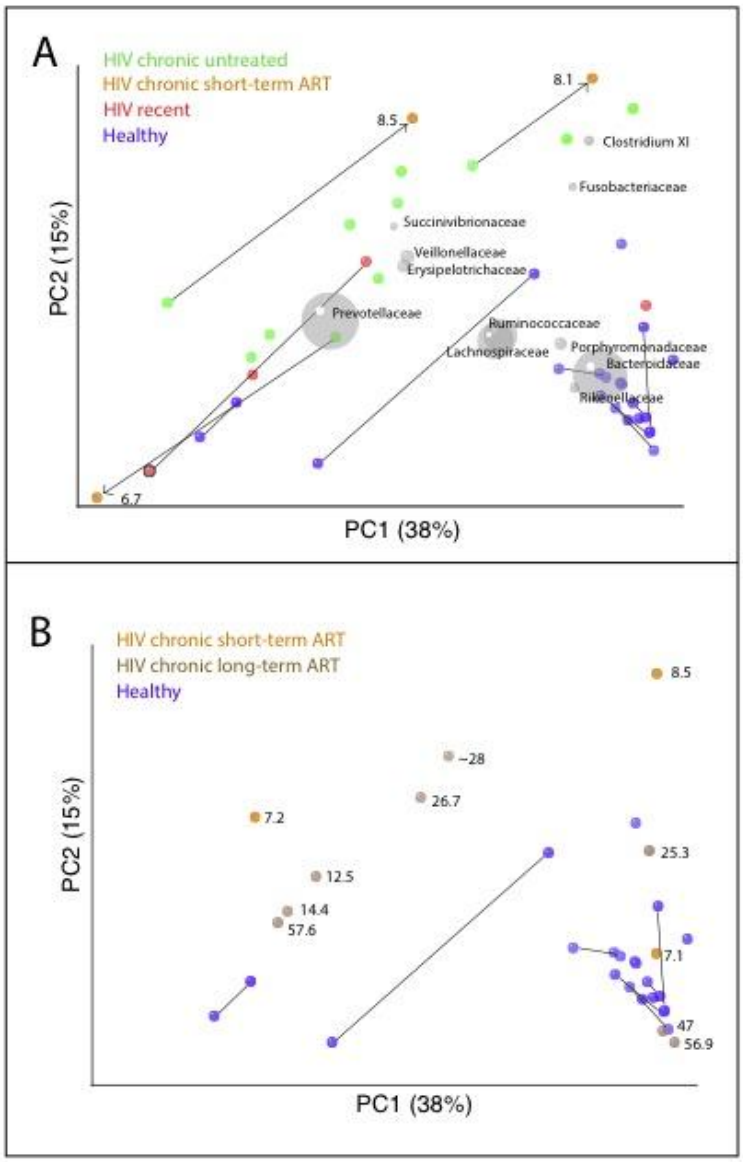


Fig. S2: *Phylogenetic tree showing how OTUs that discriminate by HIV-status and by culture relate to each other and to bacterial species.*

Trees show the relationship between OTUs that 1) Discriminated between individuals chronically infected with HIV (HIVpos; red) and healthy controls (HIVneg; blue) or 2) Discriminated between adults from Malawi and the Amazonas State of Venezuela (Agrarian; orange) and from the US (Western; green). Related bacterial species are in black text. Trees were made by inserting the representative sequence of the OTUs and of related bacterial species into a reference tree using the parsimony insertion algorithm in ARB (see methods). Panels show A) Bacteroidetes, B) Clostridiales, C) Erysipelotrichales, D) Proteobacteria, and E) Veillonellaceae.

This figure allows for evaluation of whether discriminating OTUs are parts of deeper phylogenetic lineages, their relationship to types species, and the lineages that are driving factors of the observed relationship between the microbiota in HIV positive individuals and healthy individuals in agrarian cultures. In some cases the same greengenes reference OTU was important both for differentiating adults from agrarian cultures from the US and HIV-positive individuals within the US population from HIV-negative. These include OTU168718 (related to *Clostridium viride*; Ruminococcaceae), OTU76547 (related to *Mogibacterium neglectum*; Clostridium XIII), OTU562244 (related to *Sutterella wadsworthia*; Beta Proteobacteria), OTU268604 (related to Paraprevotella) and 3 OTUs closely related to *P. stercoreae*. Different Agrarian- and HIV-positive associated greengenes OTUs were also sometimes highly related, such as in a lineage related to the acetogen *Sporobacter termitidis* (Ruminococcaceae) and one containing *E. biforme* (Erysipelotrichales). Related to Fig. 2.

Fig. S2

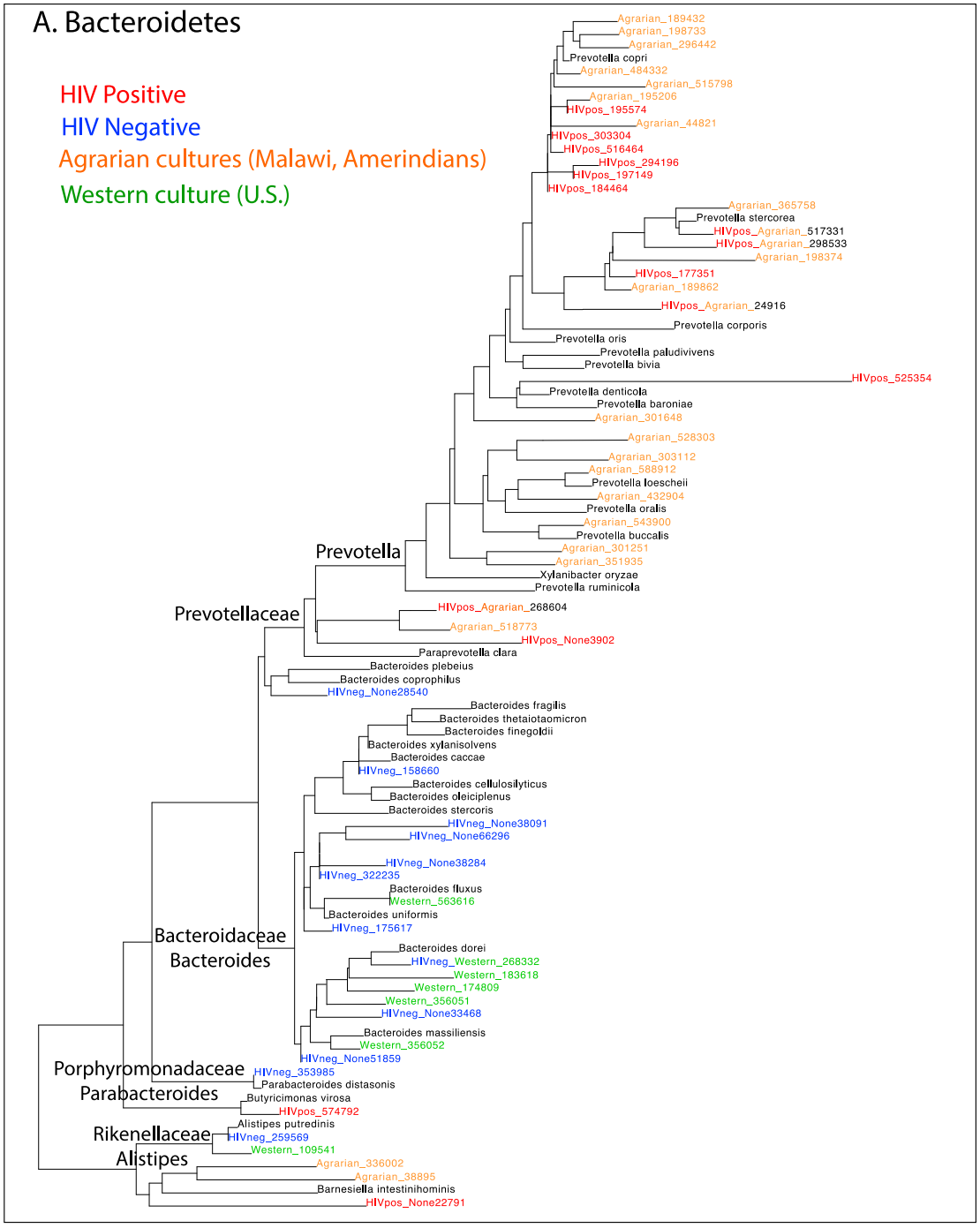
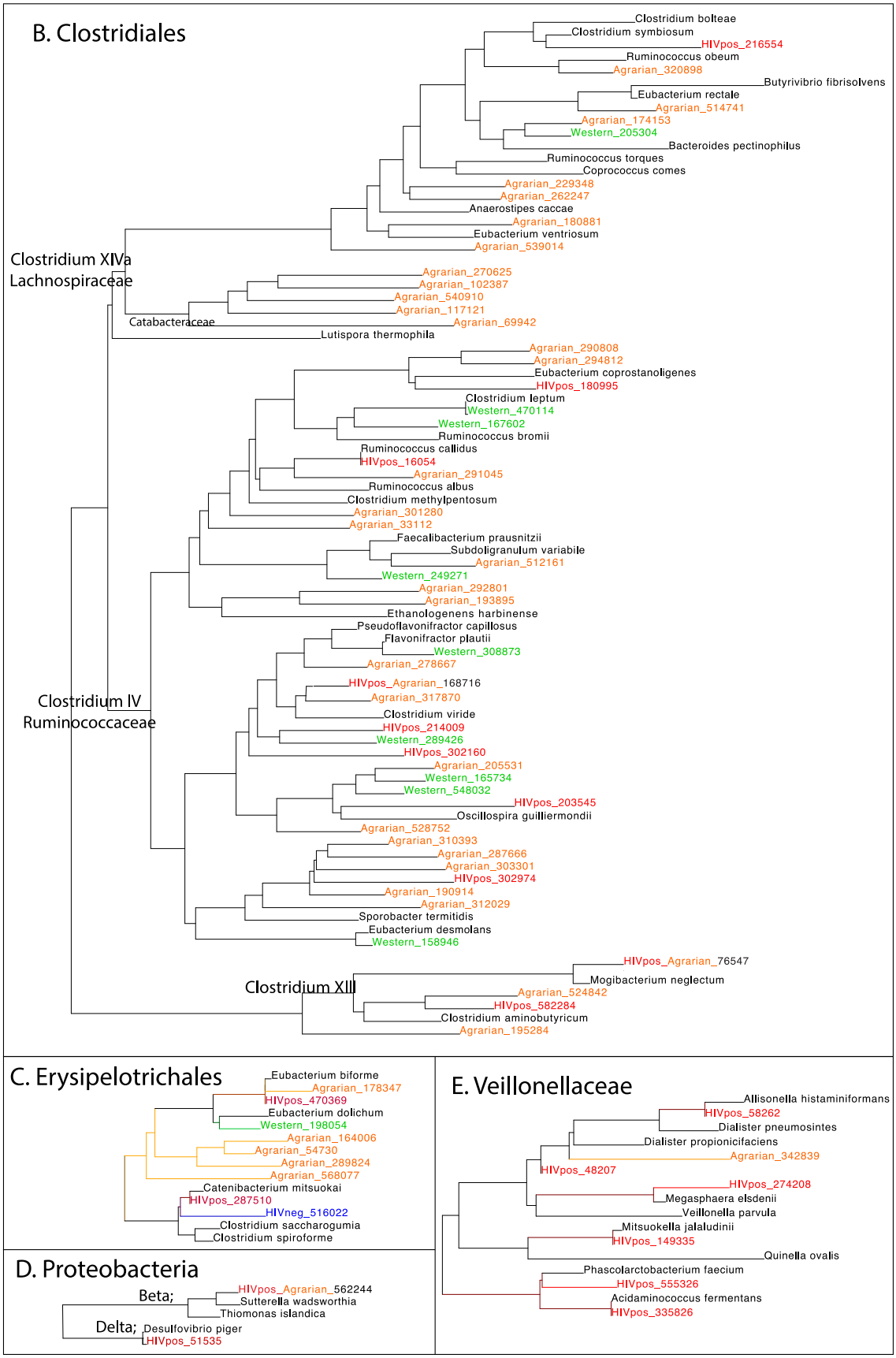

A. Bacteroidetes
HIV Positive
HIV Negative
Agrarian cultures (Malawi, Amerindians)
Western culture (U.S.)
Agrarian_189432
Agrarian_198733
Agrarian_296442
Prevotella copri
Agrarian_484332
Agrarian_515798
Agrarian_195206
HIVpos_195574
Agrarian_44821
HIVpos_303304
HIVpos_516464
HIVpos_294196
HIVpos_197149
HIVpos_184464
Agrarian_365758
Prevotella stercorea
HIVpos_Agrarian_517331
HIVpos_Agrarian_298533
Agrarian_198374
HIVpos_177351
Agrarian_189862
HIVpos_Agrarian_24916
Prevotella corporis
Prevotella oris
Prevotella paludivivens
Prevotella bivia
HIVpos_525354
Prevotella denticola
Prevotella baroniae
Agrarian_301648
Agrarian_528303
Agrarian_303112
Agrarian_588912
Prevotella loescheii
Agrarian_432904
Prevotella oralis
Agrarian_543900
Prevotella buccalis
Agrarian_301251
Agrarian_351935
Xylanibacter oryzae
Prevotella ruminicola
Prevotella
Prevotellaceae
HIVpos_Agrarian_268604
Agrarian_518773
HIVpos_None3902
Paraprevotella clara
Bacteroides plebeius
Bacteroides coprophilus
HIVneg_None28540
Bacteroides fragilis
Bacteroides thetaiotaomicron
Bacteroides finegoldii
Bacteroides xylanisolvens
Bacteroides caccae
HIVneg_158660
Bacteroides cellulosilyticus
Bacteroides oleiciplenus
Bacteroides stercoris
HIVneg_None38091
HIVneg_None66296
HIVneg_None38284
HIVneg_322235
Bacteroides fluxus
Western_563616
Bacteroides uniformis
HIVneg_175617
Bacteroidaceae
Bacteroides
Bacteroides dorei
HIVneg_Western_268332
Western_183618
Western_174809
Western_356051
HIVneg_None33468
Bacteroides massiliensis
Western_356052
HIVneg_None51859
Porphyromonadaceae
Parabacteroides
HIVneg_353985
Parabacteroides distasonis
Butyricimonas virosa
HIVpos_574792
Rikenellaceae
Alistipes
Alistipes putredinis
HIVneg_259569
Western_109541
Agrarian_336002
Agrarian_38895
Barnesiella intestinihominis
HIVpos_None22791

Fig. S2 (cont.)


B. Clostridiales
Clostridium bolteae
Clostridium symbiosum
HIVpos_216554
Ruminococcus obeum
Agrarian_320898
Butyrivibrio fibrisolvens
Eubacterium rectale
Agrarian_514741
Agrarian_174153
Western_205304
Bacteroides pectinophilus
Ruminococcus torques
Coprococcus comes
Agrarian_229348
Agrarian_262247
Anaerostipes caccae
Agrarian_180881
Eubacterium ventriosum
Agrarian_539014
Clostridium XIVa
Lachnospiraceae
Agrarian_270625
Agrarian_102387
Agrarian_540910
Agrarian_117121
Catabacteraceae
Agrarian_69942
Lutispora thermophila
Agrarian_290808
Agrarian_294812
Eubacterium coprostanoligenes
HIVpos_180995
Clostridium leptum
Western_470114
Western_167602
Ruminococcus bromii
Ruminococcus callidus
HIVpos_16054
Agrarian_291045
Ruminococcus albus
Clostridium methylpentosum
Agrarian_301280
Agrarian_33112
Faecalibacterium prausnitzii
Subdoligranulum variabile
Agrarian_512161
Western_249271
Agrarian_292801
Agrarian_193895
Ethanologenens harbinense
Pseudoflavonifractor capillosus
Flavonifractor plautii
Western_308873
Agrarian_278667
Clostridium IV
Ruminococcaceae
HIVpos_Agrarian_168716
Agrarian_317870
Clostridium viride
HIVpos_214009
Western_289426
HIVpos_302160
Agrarian_205531
Western_165734
Western_548032
HIVpos_203545
Oscillospira guilliermondii
Agrarian_528752
Agrarian_310393
Agrarian_287666
Agrarian_303301
HIVpos_302974
Agrarian_190914
Agrarian_312029
Sporobacter termitidis
Eubacterium desmolans
Western_158946
HIVpos_Agrarian_76547
Clostridium XIII
Mogibacterium neglectum
Agrarian_524842
HIVpos_582284
Clostridium aminobutyricum
Agrarian_195284
C. Erysipelotrichales
Eubacterium biforme
Agrarian_178347
HIVpos_470369
Eubacterium dolichum
Western_198054
Agrarian_164006
Agrarian_54730
Agrarian_289824
Agrarian_568077
Catenibacterium mitsuokai
HIVpos_287510
HIVneg_516022
Clostridium saccharogumia
Clostridium spiroforme
D. Proteobacteria
Beta;
HIVpos_Agrarian_562244
Sutterella wadsworthia
Thiomonas islandica
Delta;
Desulfovibrio piger
HIVpos_51535
E. Veillonellaceae
Allisonella histaminiformans
HIVpos_58262
Dialister pneumosintes
Dialister propionicifaciens
Agrarian_342839
HIVpos_48207
HIVpos_274208
Megasphaera elsdenii
Veillonella parvula
Mitsuokella jalaludinii
HIVpos_149335
Quinella ovalis
Phascolarctobacterium faecium
HIVpos_555326
Acidaminococcus fermentans
HIVpos_335826

45

Fig. S3: *Metaanalysis of HIV samples with a global survey of gut microbiota diversity and a study of Inflammatory Bowel Disease.*

Since HIV has been described to produce a chronic inflammatory state in the gut (Bjarnason et al., 1996; Brenchley and Douek, 2008), we also used metaanalysis to determine whether there was a relationship between the microbiota in HIV-infection and a chronic inflammatory disease of the gut. We combined this study's data with fecal samples from individuals ages 10-66 from 1) A survey of individuals from the US, Malawi and the Amazonas State of Venezuela (Global gut)(Yatsunenko et al., 2012) and 2) a study that detected differences between fecal microbiota of individuals with ileal Crohn's disease and healthy controls (IBD study)(Willing et al., 2010). Note that the latter study also included individuals with colonic Crohn's disease and ulcerative colitis, but we only show microbiota from ileal Crohn's disease here because it had the strongest divergence from healthy controls (Willing et al., 2010).

PCoA clustering with unweighted UniFrac was largely driven by technical variation between studies (Panel A), as might be expected since they varied methodologically (Table S3). Since all 3 studies sequenced fecal samples from healthy US adults however, we can partially correct for this study effect by rotating the axes so that the healthy samples from each of the three studies overlap (Panel B). Samples from individuals with ileal Crohn's disease now separate from healthy controls on one axis (left to right Panel C), and individuals from the US separate from those from agrarian cultures on a distinct axis of variation (top to bottom in Panel C). If we now show the position of fecal samples from individuals from the HIV study in this same plot (Panel D), individuals with chronic HIV infection (treated and untreated colored the same) and with recent HIV infection, all generally position further left (closer to IBD samples along

this axis) than healthy controls from the same study. As already shown (Fig. 3), the majority of individuals with chronic HIV infection cluster closer to agrarian cultures. Related to Fig. 3.

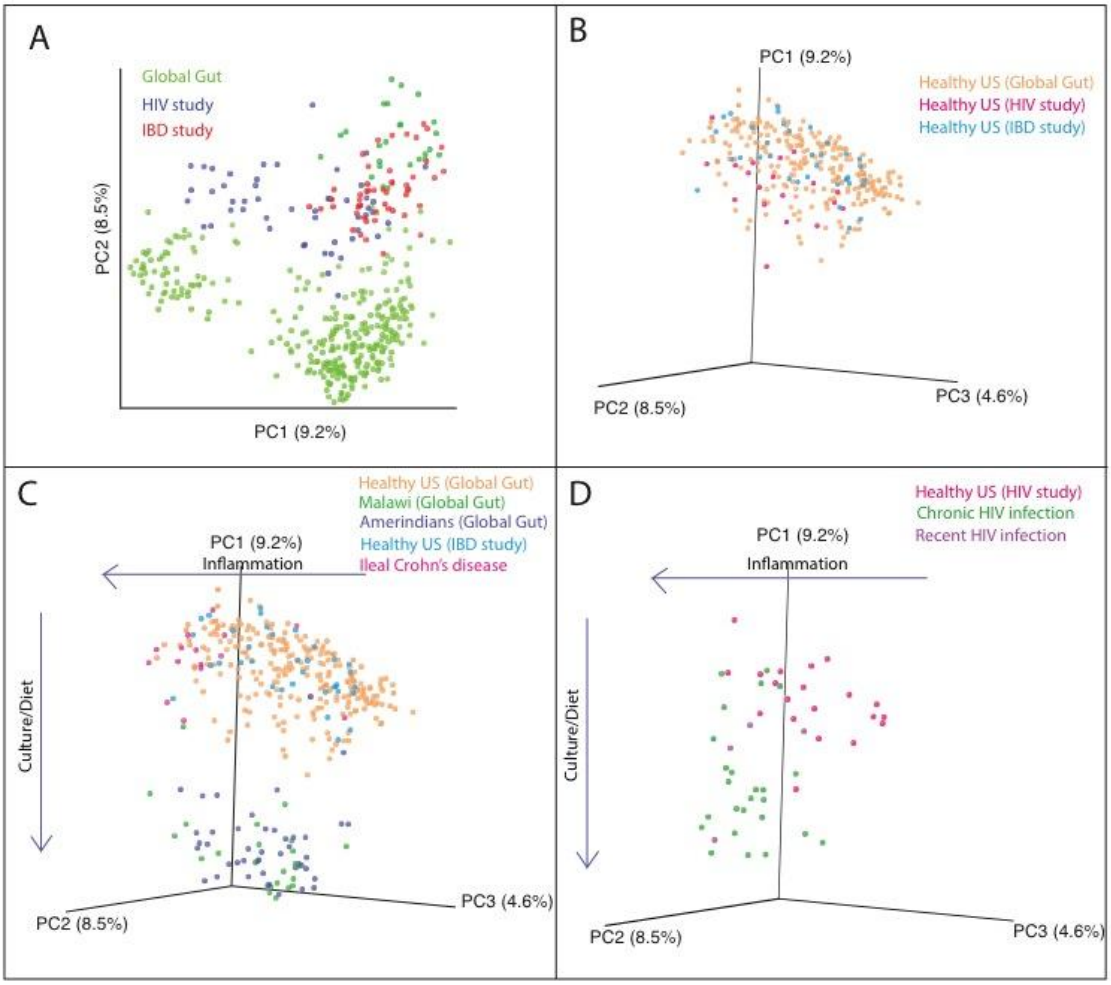


Fig. S4. *TNF-α and IL-10 production in response to lysates from selected bacteria is higher in HIV positive than HIV negative subjects.*

PBMC were stimulated with bacterial lysates and LPS. A) TNF- α and B) IL-10 as measured using a bead array assay after a 24 hour stimulation. An asterisk indicates that the responses was significantly higher in HIV-positive individuals compared to controls with Mann-Whitney t-tests ($p<0.05$). Related to Fig. 4.

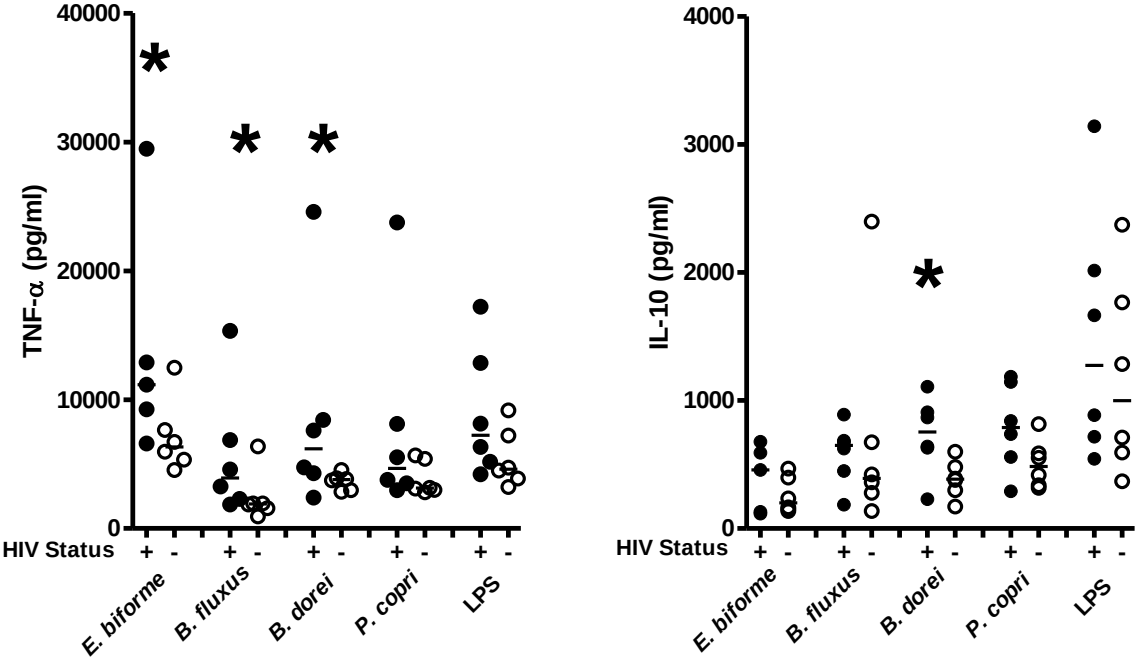


Table S1: *Detailed characteristics of subject cohorts summarized in Table 1*

Subject ID	HIV Status	Cohort	Age Years	Sex	Hispanic	Race	ART Treated	Time on ART (months)	Visits (months apart)	CD4	Viral Load	Antibiotics past 6 months	BMI
LHM002	Recent	1	26	F	No	White	No/No	0/0	2(13.2)	1342/1071	418/11,100	Yes/No	32.4/38.8
LHM026	Recent	1	34	M	No	White	No	0	1	452	449,000	Yes	38.8
LHM032	Recent	1	33	M	No	White	No	0	1	555	199,000	Yes	23.1
LHM011	Chronic	2	27	M	No	White	No	0	1	493	55,495	Yes	37
LHM025	Chronic	2	25	M	No	African American	No	0	1	291	72,300	Yes	21.6
LHM028	Chronic	2	35	M	Yes	African American	No/Yes	0/8.5	2(8.5)	447/600	301,000/39	Yes/no	26.2/24.4
LHM030	Chronic	2	30	M	Yes	White	No	0	1	1095	50,200	No	19.2
LHM031	Chronic	2	25	M	No	White	No/Yes	0/8.1	2(8.1)	614/681	8,925/40	Yes/No	25.9/24.0
LHM034	Chronic	2	33	M	No	White	No/Yes	0/6.7	2(6.7)	537/594	17,815/40	No	27.2/25.0
LHM035	Chronic	2	34	M	Yes	American Indian	No	0	1	509	1350	No	26.4
LHM047	Chronic	2	42	M	No	White	No	0	1	594	7010	No	28.5
LHM052	Chronic	2	28	M	No	White	No	0	1	658	72,500	No	21.3
LHM062	Chronic	2	27	M	No	White	No	0	1	270	869,000	No	23.7
LHM069	Chronic	2	23	M	Yes	White	No	0	1	551	74,600	No	21.3
LHM004	Chronic	2*	31	M	Yes	White	Yes	7.2	1	547	40	No	30
LHM005	Chronic	2*	54	F	No	African American	Yes	8.5	1	658	297	No	33.1
LHM017	Chronic	2*	28	M	No	White	Yes	7.1	1	576	40	No	25.9
LHM021	Chronic	3	51	M	No	White	Yes	12.5	1	311	44	No	26.8

LHM036	Chronic	3	48	M	No	White	Yes	57.6	1	443	31	No	26.3
LHM037	Chronic	3	34	M	Yes	White	Yes	47	1	364	21	No	21.7
LHM042	Chronic	3	54	M	No	White	Yes	26.7	1	876	20	No	25.8
LHM045	Chronic	3	43	M	No	African American	Yes	14.4	1	204	20	Yes	20.4
LHM048	Chronic	3	53	F	No	White	Yes	25.3	1	818	819	No	38.4
LHM063	Chronic	3	32	M	Yes	White	Yes	~28	1	611	<20	No	19.2
LHM065	Chronic	3	44	M	No	White	Yes	56.9	1	239	<20	No	18.6
LHM001	Neg	4	24	M	No	White	NA	NA	2(11.7)	NA	NA	No/No	25.4/25.4
LHM003	Neg	4	26	F	No	White	NA	NA	2(13.6)	NA	NA	No	24.9/24.9
LHM008	Neg	4	51	M	Refused	White	NA	NA	1	NA	NA	No	24.9
LHM009	Neg	4	37	M	No	White	NA	NA	2(13.6)	NA	NA	No/No	30.1/30.2
LHM010	Neg	4	22	F	No	White	NA	NA	1	NA	NA	Yes	22.3
LHM012	Neg	4	53	M	No	White	NA	NA	1	NA	NA	No	25
LHM013	Neg	4	49	M	No	African American	NA	NA	2(13.6)	NA	NA	No/No	30.5/29.2
LHM014	Neg	4	46	F	No	White	NA	NA	2(15)	NA	NA	No/Yes	30.4/30.6
LHM015	Neg	4	44	M	No	White	NA	NA	2(15.2)	NA	NA	No/No	29.4/24.9
LHM016	Neg	4	21	M	No	White	NA	NA	1	NA	NA	Yes	21.2
LHM019	Neg	4	30	F	No	White	NA	NA	2(13.6)	NA	NA	No/No	27.4/25.8
LHM027	Neg	4	52	F	Yes	White	NA	NA	2(12.8)	NA	NA	No/No	19.3/24.4
LHM049	Neg	4	31	M	No	White	NA	NA	1	NA	NA	No	25.2

Table S2: *Microbial taxa that differ between untreated individuals with chronic HIV infection and HIV-seronegative individuals*

Panels A) and B) show genera and families that significantly differ with HIV-infection. Those with higher relative abundance with HIV-infection are highlighted in yellow and in HIV-negative individuals in blue. All genera with a raw p-value <0.05 with ANOVA are listed and those with a p-value <0.05 after correcting for multiple tests with the FDR technique are marked with a star..Panel C) shows 97% ID OTUs that discriminated between the fecal microbiota of individuals chronically infected with HIV (untreated) and HIV-negative in a Random Forest analysis. All OTUs with an importance score >0.001 are listed. The direction of change was determined based on their means. OTUs that increased in relative abundance with HIV infection are in yellow and decreased are in blue. Those which Random Forests analysis determined to be important for discrimination but for which direction of change with HIV infection could not be reliably determined (ANOVA p>0.1) are listed as unresolved and not shaded. The Mean and standard error of the importance score for 100 rarefactions at 4600 sequences per sample is listed. A representative sequence for each OTU was BLASTed against a database of 16S rRNA gene sequences from type strains and the top hit, % ID to this top hit, and alignment and sequence length for the top hit are listed. OTUs for which there was a %ID greater than 97% ID over the near full length are in bold. The mean relative abundance in individuals chronically infected with HIV and HIV negative are listed as well as the probability that the relative abundance is significantly different between the two groups as calculated with ANOVA. Related to Fig. 2. **See Excel file.**

Table S3: *Studies included in the meta-analyses shown in Fig. 3 and Fig. S3*

Study	Target region/ primer	Sequencing platform	General description	Result	Reference
Global gut	**V4** CCGGACTACHV GGGTWTCTAAT	Illumina Hi-Seq 2000	Fecal samples obtained at a single timepoint from 531 individuals, aged 0-83 years, from the USA, the Amazonas state of Venezuela, and Malawi.	Stratification by age and country, with USA particularly distinct.	(Yatsunenko et al., 2012)
Diet_enterotypes	**V2** CTGCTGCCTYCC GTA	454 Titanium	Fecal samples from 98 individuals with long-term diet inventories .	Strongly association of fecal microbiota with long-term diets, particularly protein and animal fat (Bacteroides) versus carbohydrates (Prevotella).	(Wu et al., 2011)
IBD_twins	**V1-3** GGTACCGCGGCKGC TGGCAC	454 FLX	Fecal samples from a cohort of 40 twin pairs concordant or discordant for Crohn's disease or ulcerative colitis.	A particularly large divergence noted with Ileal Crohn's disease.	(Willing et al., 2010)
HIV	**V4** CCGGACTACHV GGGTWTCTAAT	MiSeq personal Sequencer	This study.		

Supplemental Experimental Procedures

Subject Recruitment

All study participants were residents of the Denver metropolitan area and were recruited and studied at the Clinical Translational Research Center of the University of Colorado Hospital. Four cohorts were prospectively recruited based on HIV-1 status: ***1) Recent HIV-1 infection*:** Individuals with a high risk HIV-1 exposure in the past six months with a positive ELISA or rapid test and Western Blot obtained within 21 days prior to study entry and one of the following: A. A non-reactive detuned ELISA (either a #a11-LS or a LgG-Capture BED-EIA assay). B. Documented negative ELISA or rapid test within 6 months prior to study entry or C. Documented negative or indeterminate Western blot within 6 months prior to study entry. ***2) Chronic HIV-1 infection untreated***: Individuals with a positive antibody or PCR test at least 6 months prior to enrollment and either ART drug-naïve (defined as < 10 days of ART treatment at any time prior to entry), or previously on ART but off treatment for the previous 6 months prior to screening. ***3) Chronic HIV-1 infection on long-term ART***: ART for ≥12 months with a minimum of three ART drugs prior to study entry and <50 copies HIV RNA/mL within 30 days prior to study entry and no plasma HIV-1 RNA ≥50 copies/mL in the past 6 months. ***4) Healthy controls***: HIV negative individuals matched to HIV-infected participants for sex, age and smoking status.

Sequence Data Analysis

Raw sequences were processed using the default parameters of QIIME version 1.5.0 (Caporaso et al., 2010b). Specifically, we truncated the reads after runs of >3

consecutive low quality base calls and excluded reads with <0.75 of the original read length after truncation. We excluded reads with ambiguous bases after quality trimming. We demultiplexed the samples using error-correcting Golay bar-codes to prevent mis-assignment. OTUs with only one sequence in them were removed from the analysis as likely noise.

To produce the phylogenetic tree for input into UniFrac we aligned a representative sequence for each OTU using PyNAST (Caporaso et al., 2010a) and constructed a phylogenetic tree from this alignment using FastTree2 (Price et al., 2010). Random Forests classification on the HIV data was carried out using SourceTracker software (http://sourceforge.net/projects/sourcetracker/)(Knights et al., 2011) with 500 trees and all default settings. We included only one sample per individual so that each microbiota could be considered an independent observation and rarefied the number of sequences per sample at 4,600. Random Forests analysis was performed for each comparison on 100 rarefied versions of the data, and the average cross-validation error estimates and OTU importance estimates were calculated using custom code. Random Forests assigns an importance score to each OTU by estimating the increase in error caused by removing that OTU from the set of predictors. We considered an OTU to be highly predictive if its average importance score was at least 0.001, as previously done in (Yatsunenko et al., 2012). Additionally, we determined the direction of change with HIV status as the difference in the mean relative abundance in samples from healthy individuals and individuals chronically infected with HIV. OTUs for which the direction of change was not supported by an ANOVA p-value of less than 0.1 were excluded from further consideration.

Bacterial families and genera represented in each sample were determined using the RDP classifier retrained on the greengenes taxonomy (McDonald et al., 2012). For statistical comparisons of healthy individuals to those chronically infected with HIV, only a single sample per person was used again so that each sample could be considered an independent observation. Significant differences in taxonomic groups and OTUs were evaluated using the ANOVA test. Comparisons were only performed on OTUs/taxa found in at least 4 individuals, and p-value were corrected for multiple comparisons using the false discovery rate of Benjamini and Hochberg (Benjamini and Hochberg, 1995), using the otu_category_significance.py script of QIIME.

To determine whether any of the discriminating OTUs were highly related to bacterial type strains, we compared the sequences to near full-length 16S rRNA gene sequences from 9177 bacterial type strains downloaded from the Ribosomal Database Project II database (Maidak et al., 2001) using the blastall program. We further evaluated the relationships between discriminating OTUs and type species by inserting them into a reference phylogenetic tree using the parsimony insertion algorithm in ARB (Ludwig et al., 2004). Sequences were first aligned using NAST (DeSantis et al., 2006), and then inserted into the reference tree provided in the 16S rRNA ARB database of Philip Hugenholz (Hugenholtz, 2002), which can be downloaded under the resources tab of the RDP website (http://rdp.cme.msu.edu), and used the filter supplied in this database (lanemaskPH) to exclude hypervariable positions that are not well aligned with NAST. Trees were exported from ARB and loaded into TopiaryExplorer to color the node names and branches (Pirrung et al., 2011).

Bacterial cultures and lysates

Type strains for *B. dorei* (DSM 17855), *B. fluxus* (DSM 22534), *P. copri* (DSM18205) and *E. biforme* (DSM 3989) were purchased from the DSMZ and incubated anaerobically in balch tubes with butyl rubber stoppers at 37^{o}C until stationary phase. All 4 isolates were cultured in rich gut microbiota medium described in (Goodman et al., 2011). To avoid denaturing potentially critical proteins and because these bacteria are all non spore-forming strict anaerobes, bacterial lysates were made without a "heat-kill" step. Bacterial cultures were washed, resuspended in water and subjected to multiple rounds of freeze/thaw to disrupt the membrane. In addition, *E. coli* serotype O55:K59(B5) was purchased from the ATCC. It was grown to log phase in nutrient broth (BD 234000) and because it is an aerobic bacteria it was heat killed by autoclaving for 20 minutes. The lysates were then quantified using a BCA protein assay kit (Pierce Rocford, Ill) and adjusted to 1 mg/ml and stored at -80^{o}C. To confirm that the bacterial lysates were not able to grow, 100ug of lysate was cultured at 37^{o}C in tissue culture media (RPMI+ 10% human serum) without antibiotics for 7 days. Cultures were carefully inspected and no bacterial growth was noted.

T Cell Proliferation assay

Blood was collected from 10 untreated HIV+ subjects with chronic infection and 11 HIV seronegative subjects in Vacutainer tubes containing sodium heparin (BD Vacutainer) and PBMCs were isolated from whole blood by density gradient centrifugation on Ficoll (Amersham Biosciences). Labeled Cells were analyzed using a

LSR-II flow cytometer (BD Immunocytometry Systems). Between 250000 and 500000 events were collected. Electronic compensation was performed with Ab capture beads (BD Biosciences) stained separately with individual mAbs used in the test samples. To ensure the accuracy and precision of the measurements taken from day to day quality control was performed on the LSR-II daily using the Cytometer Setup & Tracking (CS&T) feature within BD FACSDiva software. The data files were analyzed using Diva software (BD) and FlowJo Software (Treestar, Co). Lymphocytes were gated by their forward and side scatter profile. $CD3^{+}$, CD4+ T cells were then selected and their Cell Trace profile was analyzed. Unstimulated PBMC labeled with Cell Trace were used to determine where the gate for proliferating cells was set and the percentage of divided cells (Cell Trace low) were determined.

Supplemental References

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B *57*, 289-300.

Bjarnason, I., Sharpstone, D.R., Francis, N., Marker, A., Taylor, C., Barrett, M., Macpherson, A., Baldwin, C., Menzies, I.S., Crane, R.C., *et al.* (1996). Intestinal inflammation, ileal structure and function in HIV. AIDS *10*, 1385-1391.

Brenchley, J.M., and Douek, D.C. (2008). HIV infection and the gastrointestinal immune system. Mucosal immunology *1*, 23-30.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., and Knight, R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics *26*, 266-267.

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

DeSantis, T.Z., Hugenholtz, P., Keller, K., Brodie, E.L., Larsen, N., Piceno, Y.M., Phan, R., and Andersen, G.L. (2006). NAST: A multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res *34*, 394-399.

Goodman, A.L., Kallstrom, G., Faith, J.J., Reyes, A., Moore, A., Dantas, G., and Gordon, J.I. (2011). Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. Proc Natl Acad Sci U S A *108*, 6252-6257.

Hugenholtz, P. (2002). Exploring prokaryotic diversity in the genomic era. Genome Biol *3*, 1-8.

Knights, D., Kuczynski, J., Koren, O., Ley, R.E., Field, D., Knight, R., DeSantis, T.Z., and Kelley, S.T. (2011). Supervised classification of microbiota mitigates mislabeling errors. The ISME journal *5*, 570-573.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., *et al.* (2004). ARB: a software environment for sequence data. Nucleic Acids Res *32*, 1363-1371.

Maidak, B.L., Cole, J.R., Lilburn, T.G., Parker, C.T., Jr., Saxman, P.R., Farris, R.J., Garrity, G.M., Olsen, G.J., Schmidt, T.M., and Tiedje, J.M. (2001). The RDP-II (Ribosomal Database Project). Nucleic Acids Res *29*, 173-174.

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. Isme Journal *6*, 610-618.

Pirrung, M., Kennedy, R., Caporaso, J.G., Stombaugh, J., Wendel, D., and Knight, R. (2011). TopiaryExplorer: visualizing large phylogenetic trees with environmental metadata. Bioinformatics *27*, 3067-3069.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. PloS one *5*, e9490.

Willing, B.P., Dicksved, J., Halfvarson, J., Andersson, A.F., Lucio, M., Zheng, Z., Jarnerot, G., Tysk, C., Jansson, J.K., and Engstrand, L. (2010). A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. Gastroenterology *139*, 1844-1854 e1841.

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., *et al.* (2011). Linking long-term dietary patterns with gut microbial enterotypes. Science *334*, 105-108.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., *et al.* (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222-227.