

Supplementary Table S1. Raw and rarefied sequence read data for sewage and animal samples.

Host group	Bacterial sequence reads (total for group)	# samples in group	Average reads^b per sample after subsampling
Sewage	13,565,089	18	143,361 ^c
Cat	7,878,286	9	500,000
Chicken	4,547,099	9	462,212
Cow	6,884,544	9	500,000
Deer	6,188,212	8	500,000
Dog	6,132,510	9	485,038
Swine	7,032,645	9	500,000

^aSewage was used as a proxy for human fecal inputs, but it also contains a large proportion of water- and pipe-associated bacteria. We bioinformatically subsampled sewage data to obtain only the fecal signature, as described by Newton et al. 2015, that is predominantly composed of six families: Bacteroidaceae, Rikenellaceae, Porphyromonadaceae, Prevotellaceae, Lachnospiraceae, and Ruminococcaceae.

^bIndividual animal samples were randomly subsampled 500,000 reads if the total was greater; all sequence counts were kept if the total was <500,000.

^cSewage reads represent only the fecal portion of sewage, not the total community.

Supplementary Table S2. Oligotyping analysis: total sequences analyzed, sequences kept after noise filtering, oligotypes generated, nucleotide positions selected for decomposition.

Taxon	Total sequences	Sequences after noise filtering ^a	Oligotypes after noise filtering ^b	Nucleotide positions selected ^c
<i>Alistipes</i>	436,909	434,380	52	15,16,20,21,23,24,26,32,33,34,37,40,43,44,50,52
<i>Bacteroides</i>	2,574,691	2,504,099	187	0,1,7, 12, 13,14, 15, 16, 20, 21, 22,23, 24, 25, 32, 33, 34, 35, 36, 37, 38, 40, 43, 48, 49, 50, 52, 53
<i>Blautia</i>	610,105	586,640	152	1,2,11,14,15,16,19,20,21,23,24,25,26,27,31,35,36,37,38,40,41, 43,44,49,50,53,54,56
<i>Coprococcus</i>	173,044	164,401	242	0,2,14,15,16,19,20,21,23,24,25,26,31,33,35,36,37,38,40,41,42,44,50, 52,53,56
<i>Dorea</i>	95,431	94,057	89	1,14,16,19,21,22,23,26,31,34,35,36,37,41,49,53,56
<i>Faecalibacterium</i>	173,618	171,276	101	3,14,15,18,19,20,21,22,23,24,26,29,31,32,34,35,36,37,44,52
<i>Lachnospiraceae</i> (unclassified)	1,775,165	1,514,466	613	0, 1, 2, 3, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 31, 32, 34, 37, 38, 40, 41, 42, 43, 44, 48, 49, 50, 51, 53, 56
<i>Parabacteroides</i>	522,886	512,607	103	0,4,15,16,17,21,22,23,24,34,35,36,37,38,42,44,45,50,51,52,53,55
<i>Prevotella</i>	1,902,562	1,869,883	121	0, 11, 13, 14, 17, 19, 20, 21, 22, 24, 26, 27, 29, 34, 35, 37, 38, 41, 45, 46, 49, 50, 52, 53, 55, 56, 58
<i>Roseburia</i>	228,352	219,861	186	1,2,9,14,15,16,20,21,22,24,26,31,32,35,36,37,38,40,41,50,51,53,56

^aFinal number of sequences used in analysis after noise filtering

^bNoise filtering removed oligotypes present in less than 3 samples or with a minimum substantive abundance equal to [total sequences/5000]

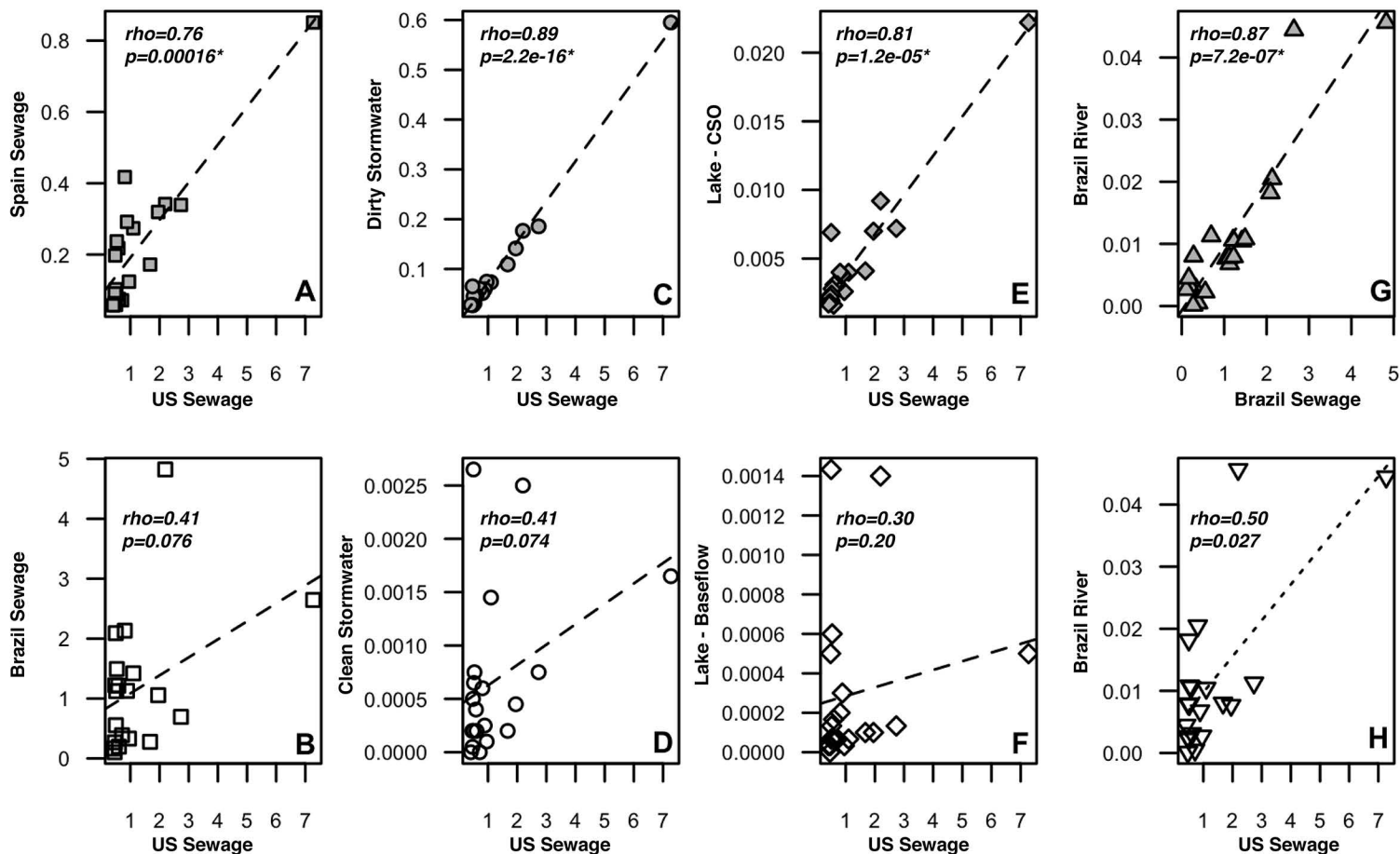
^cNumbers represent the nucleotide position from 0-*N*, where *N* is the length of the padded sequences in each analysis (*N*≈60 for all taxa).

Supplementary Table S3. Proportion of variability in oligotype distributions of ten fecal taxa that can be attributed to differences among host groups. Statistics are based on a non-parametric equivalent of the MANOVA (*adonis*) for community datasets. A paired t-test of *adonis* R² values confirmed that pooled values were significantly lower than correlations based on individual animal hosts ($p=2.0 \times 10^{-10}$), and analysis of variance of indicated that this was due to higher variance within the pooled animal group ($p=2 \times 10^{-16}$).

Taxon	Individual Host Groups		Sewage vs. Pooled Animals	
	R²	p	R²	p
<i>Alistipes</i>	0.553	<0.001	0.214	<0.001
<i>Bacteroides</i>	0.579	<0.001	0.184	<0.001
<i>Blautia</i>	0.603	<0.001	0.169	<0.001
<i>Coprococcus</i>	0.585	<0.001	0.179	<0.001
<i>Dorea</i>	0.625	<0.001	0.160	<0.001
<i>Faecalibacterium</i>	0.557	<0.001	0.184	<0.001
<i>Lachnospiraceae</i> (unclassified)	0.620	<0.001	0.178	<0.001
<i>Parabacteroides</i>	0.516	<0.001	0.170	<0.001
<i>Prevotella</i>	0.565	<0.001	0.149	<0.001
<i>Roseburia</i>	0.549	<0.001	0.191	<0.001

Supplementary Figure S2. Sewage signature correlations between sanitary sewage and environmental samples.

Spearman's rank coefficients (*rho*) and significance values (*p*) are given for each data pair; significant values ($p < 0.0071$) based on corrected alpha are noted (*). A) Spain sanitary sewage, B) Brazil sanitary sewage, C) sewage-contaminated stormwater, D) clean stormwater, E) Lake Michigan after a CSO, F) Lake Michigan during baseflow compared to the US sewage signature; and Brazil river with human fecal contamination versus Brazil sewage signature (G) or US sewage signature (H).



REFERENCES CITED

Newton RJ, McLellan SL, Dila DK, Vineis JH, Morrison HG, Eren AM, Sogin ML.
2015. Sewage reflects the microbiomes of human populations. *mBio* **6**(2):e02574-14.