

**COMPARING SPECIES DIVERSITY OF MICROBIAL COMMUNITIES  
DERIVED FROM 16S rRNA SEQUENCING AND SHOTGUN  
METAGENOMICS  
Supplementary Material**

Neethu Shah, Haixu Tang, Thomas G. Doak, Yuzhen Ye

Table 1. Details of 32 metagenomic datasets of gut-associated microbial communities.

Data ID	Total reads	Average length(bp)	Source
TS1	217386	238	Gordon
TS2	443526	178	Gordon
TS3	510972	201	Gordon
TS4	414754	229	Gordon
TS5	490776	205	Gordon
TS6	535763	221	Gordon
TS7	555853	243	Gordon
TS8	414497	243	Gordon
TS9	499499	250	Gordon
TS19	498880	165	Gordon
TS20	495039	198	Gordon
TS21	413772	215	Gordon
TS28	302772	335	Gordon
TS29	502399	345	Gordon
TS30	495865	190	Gordon
TS49	519072	177	Gordon
TS50	549700	204	Gordon
TS51	434187	187	Gordon
DISTAL	146194	1050	Gill
F1S	28900	1315	Kurokawa
F1T	36326	1191	Kurokawa
F1U	16539	1474	Kurokawa
F2V	36455	1260	Kurokawa
F2W	30198	1327	Kurokawa
F2X	31237	1251	Kurokawa
F2Y	35177	1293	Kurokawa
InA	20266	1448	Kurokawa
InB	9958	1456	Kurokawa
InD	37296	1244	Kurokawa
InE	20532	1325	Kurokawa
InM	16164	1605	Kurokawa
InR	34797	1249	Kurokawa

Note: Data IDs are used as in the respective studies: Gordon (Nature, 457:480-484, 2009), Gill (Science, 312:1355-1359, 2006), Kurokawa (DNA Research, 14:169-181, 2006).

Table 2. Summary of 33 16S rRNA gene sequencing datasets of gut samples

---

Data ID	Total reads	Average length(bp)
TS1	25,140	126
TS2	42,186	126
TS3	17,726	126
TS4	25,705	126
TS5	26,608	126
TS6	27,007	126
TS7	17,469	126
TS8	17,170	126
TS9	14,787	126
TS13	15,296	126
TS14	14,220	126
TS15	14,244	126
TS19	43,635	126
TS20	13,476	126
TS21	23,714	126
TS25	20,491	126
TS26	27,626	126
TS27	25,494	126
TS28	20,905	126
TS29	15,698	126
TS30	32,083	126
TS31	16,530	126
TS32	31,690	126
TS33	28,962	126
TS49	22,201	126
TS50	30,498	126
TS51	22,691	126
TS55	37,027	126
TS56	31,512	126
TS57	30,392	126
TS148	26,458	126
TS149	35,838	126
TS150	23,463	126

---

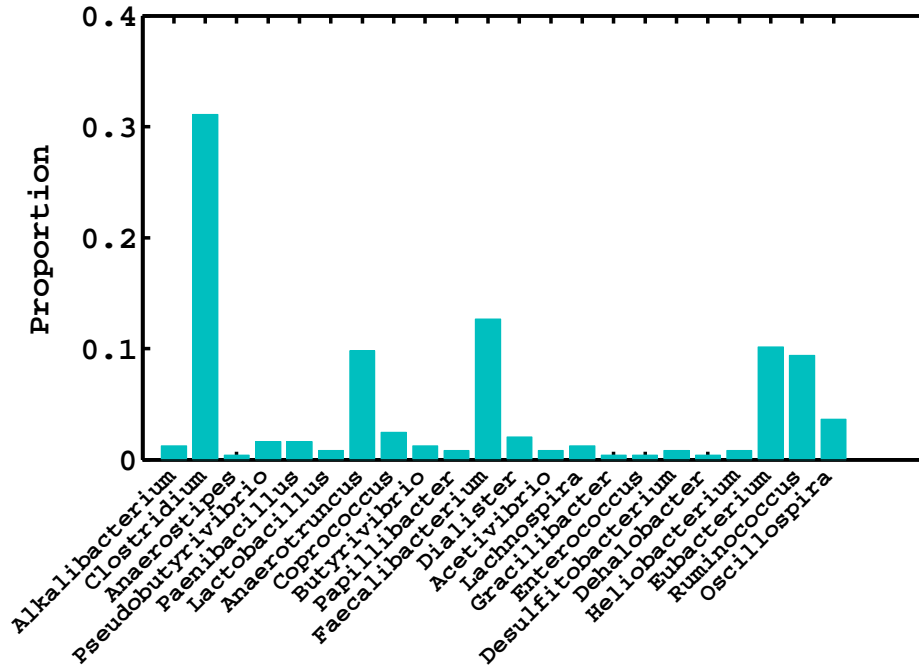
Table 3. Summary of 16S rRNA gene sequencing datasets of gut samples, submitted by Arizona State University to NIH SRA

Data ID	Total reads	Average length(bp)
SRX012481	31,835	102
SRX012482	16,260	103
SRX012483	16,940	102
SRX012484	13,963	102
SRX012485	26,915	102
SRX012486	17,334	102
SRX012487	17,544	103
SRX012488	24,794	102
SRX012489	18,509	102

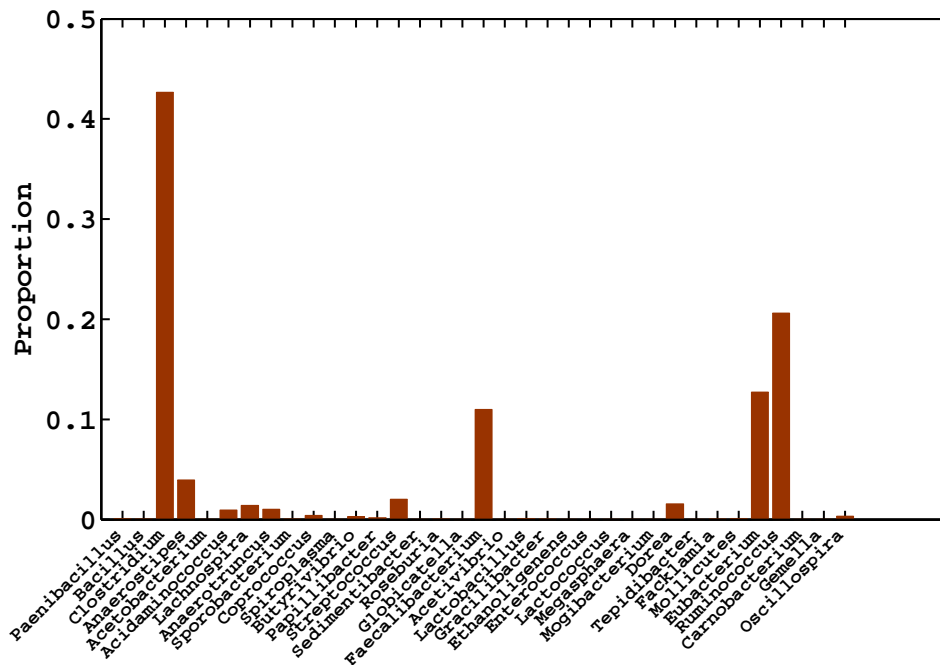
Table 4. Summary of 30 metagenomic datasets of oral samples

Data ID	Total reads	Average length(bp)
SRX011821	7,315	262
SRX011822	21,870	264
SRX011823	20,788	257
SRX011824	19,453	262
SRX011825	14,579	268
SRX011826	18,297	259
SRX011827	18,721	260
SRX011828	19,439	259
SRX011829	19,362	266
SRX011830	14,172	248
SRX011831	16,076	254
SRX011832	30,070	259
SRX011833	10,685	244
SRX011834	14,254	240
SRX011835	13,571	241
SRX011836	8,105	248
SRX011837	12,703	247
SRX011838	15,517	258
SRX011839	16,190	256
SRX011840	20,191	257
SRX011841	19,046	256
SRX011842	21,929	254
SRX011843	21,775	258
SRX011844	19,923	255
SRX011845	13,944	233
SRX011846	21,397	248
SRX011847	13,691	258
SRX011848	10,553	242
SRX011849	17,909	235
Oral(shotgun)	377275	59

4



(a) Genus distribution for metagenomics data



(b) Genus distribution for 16S rRNA sequencing

Fig. 1. Genus distribution of Firmicutes genera for a gut-associated microbial community.