

In the format provided by the authors and unedited.

Culture-enriched metagenomic sequencing enables in-depth profiling of the cystic fibrosis lung microbiota

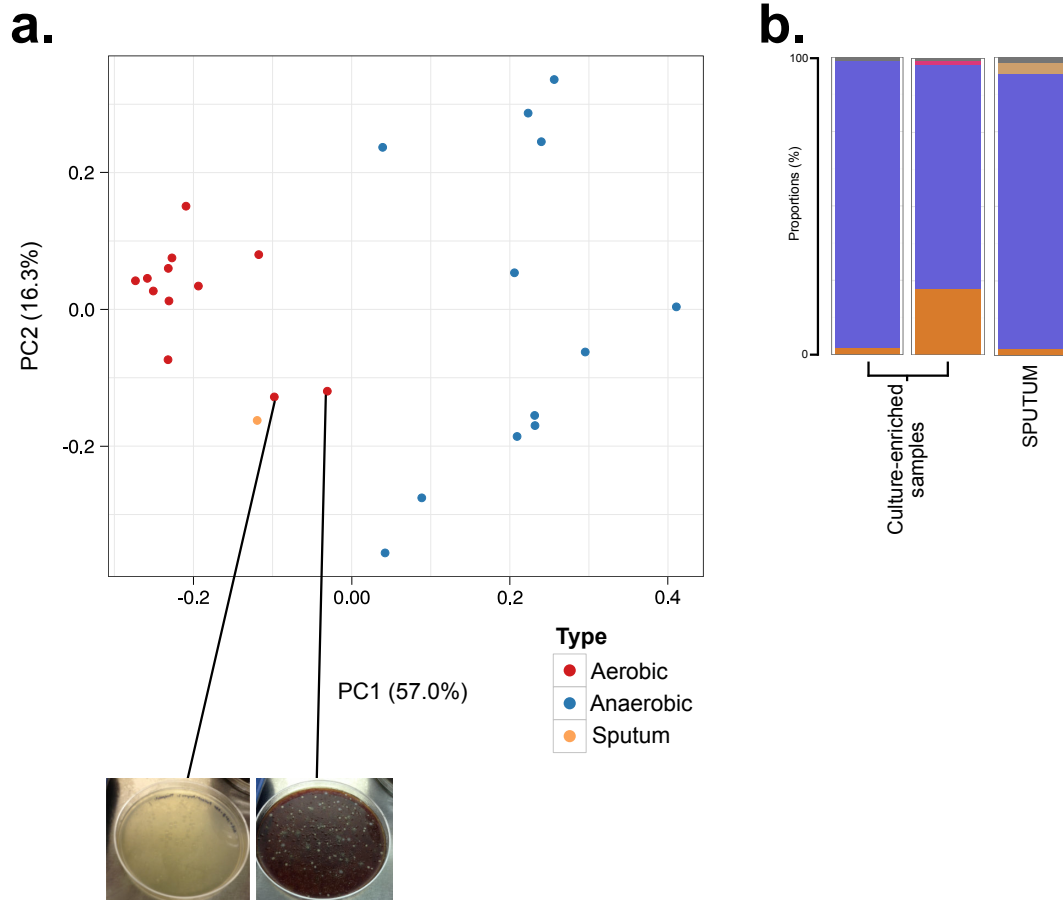
Fiona J. Whelan¹ , Barbara Waddell² , Saad A. Syed³, Shahrokh Shekariz³, Harvey R. Rabin^{2,4}, Michael D. Parkins^{2,4} and Michael G. Surette^{2,3,5*} 

¹School of Life Sciences, University of Nottingham, Nottingham, UK. ²Department of Microbiology, Immunology and Infectious Diseases, University of Calgary, Calgary, Alberta, Canada. ³Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, Canada.

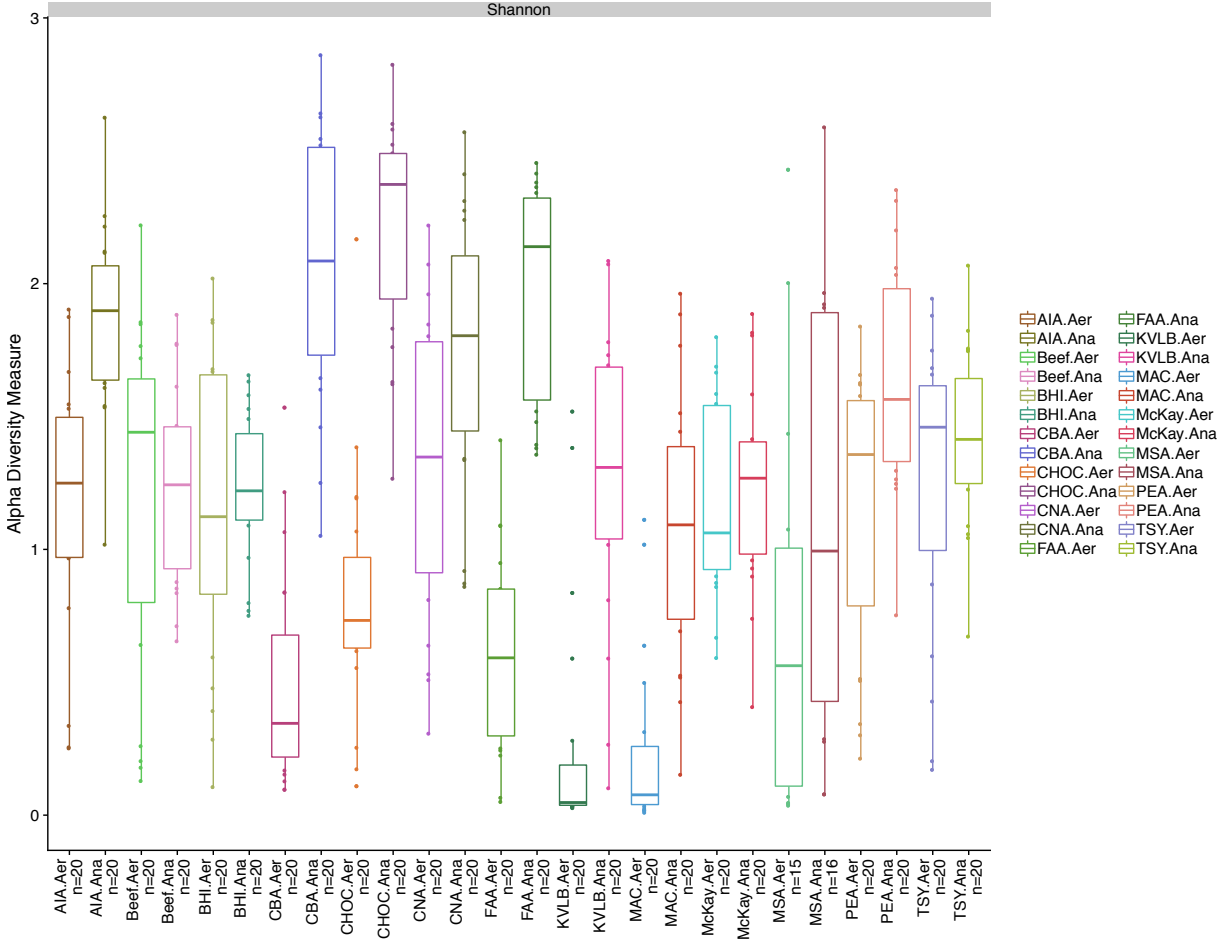
⁴Department of Medicine, University of Calgary, Calgary, Alberta, Canada. ⁵Department of Medicine, McMaster University, Hamilton, Ontario, Canada.

*e-mail: surette@mcmaster.ca

Supplementary Information



Supplementary Figure 1: **Clustering of aerobic culture-enriched samples with the direct sequencing is not due to non-viable DNA.** **a.** Two culture-enriched samples sharing similar communities with the direct sequencing result, as measured by Unweighted Unifrac, have visible microbial growth as indicated by images captured following culture-enrichment, n=25. **b.** Instead, these samples have similar taxonomic profiles due to an abundance of *Pseudomonas sp.*



Supplementary Figure 2: The α -diversity of each culture condition varies across samples. The Shannon diversity of each cultured plate varies across and within culture conditions. Boxplots display the 1st and 3rd quartiles, a horizontal line to indicate the median, and whiskers extending to 1.5 times the interquartile range. For media abbreviations, see **Extended Data Table 1**; Aer = aerobic, Ana = anaerobic.

a.

```

pre:
abund = abundance cutoff
plateList = ()
allOTUs = all cultured OTUs

for each OTU a in allOTUs
  count the # of plates with OTU a at abundance > abund
  if count == 1
    plate p = plate w/ OTU a on it
    plateList += plate p
    allOTUs -= any OTU on plate p
  while allOTUs is not empty
    find plate q with the greatest # of OTUs remaining in all OTUs
    plateList += q
    allOTUs -= any OTU on plate q

post:
plateList = list of plates for metagen seqing
allOTUs = ()

```

b.

```

pre:
sputumAbund = abundance cutoff in the originating sample
abund = abundance cutoff (corresponding to that in Figure 7)
plateList = ()
allOTUs = all cultured OTUs

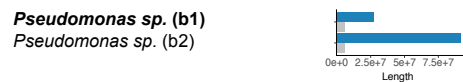
abundOTUs = a subset of allOTUs where abundance > sputumAbund
for each OTU a in abundOTUs
  count the # of plates with OTU a at abundance > abund
  if count == 1
    plate p = plate w/ OTU a on it
    plateList += plate p
    allOTUs -= any OTU on plate p
  while abundOTUs is not empty
    find plate q with the greatest # of OTUs remaining in all OTUs
    plateList += q
    abundOTUs -= any OTU on plate q

post:
plateList = list of plates for metagen seqing
abundOTUs = ()

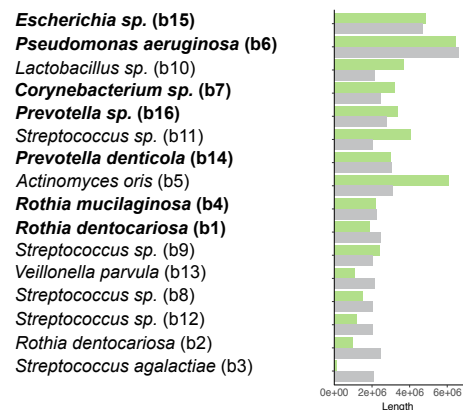
```

Supplementary Figure 3: Pseudocode for the denovo (a) and adjusted (b) PLCAs. Implementations are available under a GNU license at <https://github.com/fwhelan/PLCA>.

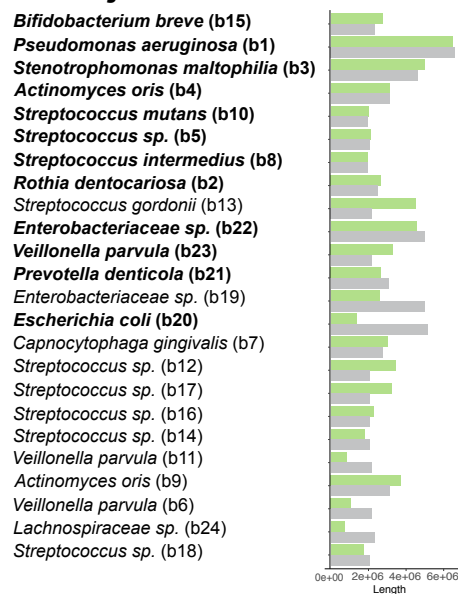
a. Direct



b. denovo PLCA



c. adjusted PLCA



Supplementary Figure 4: Direct and culture-enriched metagenomic bin lengths (coloured) and genome length of the closest reference genome (gray). Bin lengths for the direct (a), denovo (b), and adjusted PLCA metagenomics (c). Bold taxonomic names indicates the bins as MAGs.

Supplementary Table 1: Genus-level assignments of OTUs not previously reported in culture-enrichment studies.

Genus taxonomic assignment	Maximum relative abundance observed (%)
Root; p_Proteobacteria; c_Gammaproteobacteria; o_Pasteurellales; f_Pasteurellaceae; g_Actinobacillus	33.268
Root; p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_.; g_Aquabacterium	0.014
Root; p_Proteobacteria; c_Gammaproteobacteria; o_Pseudomonadales; f_Pseudomonadaceae; g_Azorhizophilus	0.387
Root; p_Firmicutes; c_Clostridia; o_Thermoanaerobacterales; f_Caldicellulosiruptoraceae; g_Caldicellulosiruptor	0.025
Root; p_Firmicutes; c_Clostridia; o_Clostridiales; f_Lachnospiraceae; g_Catonella	1.171
Root; p_Firmicutes; c_Clostridia; o_Clostridiales; f_Lachnospiraceae; g_Epuloniscium	0.014
Root; p_Proteobacteria; c_Alphaproteobacteria; o_Rhodospirillales; f_Acetobacteraceae; g_Gluconobacter	0.020
Root; p_Proteobacteria; c_Alphaproteobacteria; o_Rhodospirillales; f_Rhodospirillaceae; g_Magnetospirillum	0.032
Root; p_Firmicutes; c_Clostridia; o_Clostridiales; f_ClostridialesFamilyXIII.IncertaeSedis; g_Mogibacterium	6.939
Root; p_Firmicutes; c_Clostridia; o_Clostridiales; f_Lachnospiraceae; g_Moryella	0.240
Root; p_Tenericutes; c_Mollicutes; o_Mycoplasmatales; f_Mycoplasmataceae; g_Mycoplasma	0.013
Root; p_Actinobacteria; c_Actinobacteria; o_Bifidobacteriales; f_Bifidobacteriaceae; g_Parascardovia	1.218
Root; p_Firmicutes; c_Clostridia; o_Clostridiales; f_Lachnospiraceae; g_Pseudobutyrvibrio	0.089
Root; p_Proteobacteria; c_Alphaproteobacteria; o_Rhizobiales; f_Rhizobiaceae; g_Rhizobium	0.027
Root; p_Firmicutes; c_Clostridia; o_Clostridiales; f_Lachnospiraceae; g_Shuttleworthia	3.375
Root; p_Proteobacteria; c_Betaproteobacteria; o_Neisseriales; f_Neisseriaceae; g_Simoniella	3.6646
Root; p_Bacteroidetes; c_Bacteroidia; o_Bacteroidales; f_Porphyromonadaceae; g_Tannerella	1.260
Root; p_Synergistetes; c_Synergistia; o_Synergistales; f_Dethiosulfovibrionaceae; g-TG5	0.025

Supplementary Table 2: Full-length 16S rRNA gene sequencing results for colonies isolated from *Stenotrophomonas* re-growth. Of the 10 isolates picked from the *Stenotrophomonas* isolation procedure, 8 were identified as *Stenotrophomonas maltophilia* strains by comparison to the Human Oral Microbiome Database (HOMD) and to the 16S ribosomal RNA sequences (Bacteria and Archaea) NCBI BLAST Database.

ID	Query length (nt)	HOMD Top Hit	% Identity	BLAST Top Hit	Max Score	Query cover	E Value	Identity
1	855	S. maltophilia—HOT.663—Strain.LMG958—X95923—Named	98.7	NR_041577.1—S. maltophilia str in IAM 12423 16S rRNA gene	1487	97%	0	99%
2	867	S. maltophilia—HOT.663—Strain.LMG958—X95923—Named	98.5	NR_041577.1—S. maltophilia str in IAM 12423 16S rRNA gene	1483	97%	0	99%
3	919	S. maltophilia—HOT.663—Strain.LMG958—X95923—Named	98.5	NR_041577.1—S. maltophilia str in IAM 12423 16S rRNA gene	1485	89%	0	99%
4	924	S. maltophilia—HOT.663—Strain.LMG958—X95923—Named	98.7	NR_041577.1—S. maltophilia str in IAM 12423 16S rRNA gene	1496	90%	0	99%
5	926	S. maltophilia—HOT.663—Strain.LMG958—X95923—Named	98.6	NR_041577.1—S. maltophilia str in IAM 12423 16S rRNA gene	1478	89%	0	99%
6	919	Pseudomonas aeruginosa—HOT.536—Strain.LMG1242—Z76651—Named	99.2	NR_074828.1—Pseudomonas aeruginosa PAO1 strain PAO1 1 6S rRNA	1506	90%	0	99%
7	910	S. maltophilia—HOT.663—Strain.LMG958—X95923—Named	98.3	NR_041577.1—S. maltophilia str in IAM 12423 16S rRNA gene	1500	93%	0	98%
8	928	S. maltophilia—HOT.663—Strain.LMG958—X95923—Named	98.5	NR_041577.1—S. maltophilia str in IAM 12423 16S rRNA gene	1513	91%	0	99%
9	868	Pseudomonas aeruginosa—HOT.536—Strain.LMG1242—Z76651—Named	98.8	NR_074828.1—Pseudomonas aeruginosa PAO1 strain PAO1 1 6S rRNA	1347	85%	0	99%
10	914	S. maltophilia—HOT.663—Strain.LMG958—X95923—Named	98.1	NR_041577.1—S. maltophilia str in IAM 12423 16S rRNA gene	1504	93%	0	98%

Supplementary Table 3: Taxon recovery of OTUs from denovo PLCA-assisted culture-enriched metagenomics.

OTU taxon assignment	Corresponding metagenomic bin taxon assignment	Status of metagenomic bin	Above/below threshold
p_Proteobacteria;c_Gammaproteobacteria; o_Pseudomonadales;f_Pseudomonadaceae;g_Pseudomonas	<i>Pseudomonas aeruginosa</i>	MAG	Above
p_Proteobacteria;c_Gammaproteobacteria; o_Enterobacteriales;f_Enterobacteriaceae;g_Escherichia	<i>Escherichia sp.</i>	MAG	Above
p_Firmicutes;c_Bacilli;o_Lactobacillales; f_Streptococcaceae;g_Streptococcus	<i>Streptococcus sp.</i>	Metagenomic bin	Above
p_Firmicutes;c_Clostridia;o_Clostridiales; f_Veillonellaceae;g_Veillonella	<i>Veillonella parvula</i>	Metagenomic bin	Above
p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales; f_Prevotellaceae;g_Prevotella	<i>Prevotella sp.</i>	MAG	Above
p_Actinobacteria;c_Actinobacteria;o_Actinomycetales; f_Micrococcaceae;g_Rothia	<i>Rothia mucilaginosa</i>	MAG	Above
p_Firmicutes;c_Bacilli;o_Lactobacillales; f_Streptococcaceae;g_Streptococcus	<i>Streptococcus sp.</i>	Metagenomic bin	Above
p_Firmicutes;c_Bacilli;o_Lactobacillales; f_Streptococcaceae;g_Streptococcus	<i>Streptococcus sp.</i>	Metagenomic bin	Above
p_Firmicutes;c_Bacilli;o_Lactobacillales; f_Streptococcaceae;g_Streptococcus	<i>Streptococcus sp.</i>	Metagenomic bin	Above
p_Firmicutes;c_Bacilli;o_Lactobacillales; f_Streptococcaceae;g_Streptococcus	<i>Streptococcus agalactiae</i>	Metagenomic bin	Above
-	<i>Lactobacillus sp.</i>	Metagenomic bin	Below
-	<i>Corynebacterium sp.</i>	MAG	Below
-	<i>Prevotella denticola</i>	MAG	Below
-	<i>Actinomyces oris</i>	Metagenomic bin	Below
-	<i>Rothia dentocariosa</i>	MAG	Below
-	<i>Rothia dentocariosa</i>	Metagenomic bin	Below

Supplementary Table 4: Taxon recovery of OTUs adjusted PLCA-assisted culture-enriched metagenomics.

OTU taxon assignment	Corresponding metagenomic bin taxon assignment	Status of metagenomic bin	Above/below threshold
p_Proteobacteria;c_Gammaproteobacteria; o_Pseudomonadales;f_Pseudomonadaceae;g_Pseudomonas	<i>Pseudomonas aeruginosa</i>	MAG	Above
p_Actinobacteria;c_Actinobacteria; o_Bifidobacteriales;f_Bifidobacteriaceae;g_Bifidobacterium	<i>Bifidobacterium breve</i>	MAG	Above
p_Firmicutes;c_Bacilli;o_Lactobacillales; f_Streptococcaceae;g_Streptococcus	<i>Streptococcus mutans</i>	MAG	Above
p_Firmicutes;c_Bacilli;o_Lactobacillales; f_Streptococcaceae;g_Streptococcus	<i>Streptococcus sp.</i>	MAG	Above
p_Firmicutes;c_Bacilli;o_Lactobacillales; f_Streptococcaceae;g_Streptococcus	<i>Streptococcus intermedius</i>	MAG	Above
p_Firmicutes;c_Clostridia;o_Clostridiales; f_Lachnospiraceae;g_Oribacterium		Cultured below threshold (0.23% relative abundance)	Above
p_Proteobacteria;c_Gammaproteobacteria; o_Pseudomonadales;f_Moraxellaceae;g_Acinetobacter		Cultured below threshold (0.003% relative abundance)	Above
p_Firmicutes;c_Clostridia;o_Clostridiales; f_Lachnospiraceae	<i>Lachnospiraceae sp.</i>	Metagenomic bin	Above
p_Actinobacteria;c_Actinobacteria; o_Bifidobacteriales;f_Bifidobacteriaceae;g_Bifidobacterium		Cultured below threshold (0.07% relative abundance)	Above
p_Proteobacteria;c_Gammaproteobacteria; o_Enterobacteriales;f_Enterobacteriaceae;g_Escherichia	<i>Escherichia coli</i>	MAG	Above
p_Firmicutes;c_Bacilli;o_Bacillales; f_Staphylococcaceae;g_Staphylococcus		Cultured below threshold (0.35% relative abundance)	Above
p_Firmicutes;c_Clostridia;o_Clostridiales; f_Lachnospiraceae;g_Blautia		Cultured below threshold (0.081% relative abundance)	Above
p_Firmicutes;c_Bacilli;o_Lactobacillales; f_Streptococcaceae;g_Streptococcus	<i>Streptococcus gordonii</i>	Metagenomic bin	Above
p_Firmicutes;c_Bacilli;o_Lactobacillales; f_Carnobacteriaceae;g_Granulicatella		Not identified	Above
p_Firmicutes;c_Clostridia;o_Clostridiales; f_Veillonellaceae;g_Veillonella	<i>Veillonella parvula</i>	MAG	Above
p_Proteobacteria;c_Gammaproteobacteria; o_Xanthomonadales;f_Xanthomonadaceae; g_Stenotrophomonas	<i>Stenotrophomonas maltophilia</i>	MAG	Above
-	<i>Actinomyces oris</i>	MAG	Below
-	<i>Rothia dentocariosa</i>	MAG	Below
-	<i>Prevotella denticola</i>	MAG	Below
-	<i>Enterobacteriaceae sp.</i>	Metagenomic bin	Below
-	<i>Enterobacteriaceae sp.</i>	MAG	Below
-	<i>Capnocytophaga gingivalis</i>	Metagenomic bin	Below
-	<i>Streptococcus sp.</i>	Metagenomic bin	Below
-	<i>Streptococcus sp.</i>	Metagenomic bin	Below
-	<i>Streptococcus sp.</i>	Metagenomic bin	Below
-	<i>Streptococcus sp.</i>	Metagenomic bin	Below
-	<i>Veillonella parvula</i>	Metagenomic bin	Below
-	<i>Actinomyces oris</i>	Metagenomic bin	Below
-	<i>Veillonella parvula</i>	Metagenomic bin	Below
-	<i>Streptococcus sp.</i>	Metagenomic bin	Below

Supplementary Table 5: Numbers of reads and OTUs identified in the direct sequencing in the raw, processed, and >0.01% tables.

Sample	Raw number of reads	Raw number of OTUs	Processed number of reads	Processed number of OTUs	Number of OTUs >0.01%	Number of reads within the OTUs at >0.01%
1	30502	299	23666	126	51	23569
2	56838	336	49050	93	40	48957
3	18555	540	11090	144	97	10750
4	90439	212	88208	98	35	88041
5	67476	322	64733	211	65	64410
6	85285	290	76250	98	33	76117
7	64452	231	61360	133	48	61185
8	75229	202	71213	117	32	71041
9	11516	165	10562	48	30	10544
10	41036	357	34458	211	93	34280
11	67355	292	65910	218	73	65633
12	73518	606	64080	173	52	63839
13	104562	303	87287	148	56	87067
14	108555	624	95600	122	39	95395
15	73402	410	59000	98	32	58853
16	18735	269	14561	117	72	14516
17	44143	688	33331	104	43	33237
18	61067	411	53836	85	33	53732
19	80622	312	79322	199	76	79023
20	66241	318	64612	246	96	64319

Supplementary Table 6: Assembly statistics for the denovo- and adjusted-PLCA assisted culture-enriched metagenomics and direct metagenomic sequencing datasets.

Type of assembly	Mean contig length	Maximum contig length	N50
denovo PLCA	1,644.74	142,691	4504
adjusted PLCA	1,877.28	670,683	4423
Direct MG	548.36	34,449	565

Supplementary Table 7: Sequencing statistics, including the proportion of reads lost to host contamination, from the direct and culture-enriched metagenomic sequencing. Read loss from the AIA anaerobic sample was due to FASTQ file corruption of the R1 file provided.

Sample	Number of raw reads	Number of processed reads	Read Loss (%)
Sputum	29,229,781	2,207,904	92.45
denovo PLCA: AIA anaerobic	5,708,137	238,081	95.8
denovo PLCA: Beef aerobic	23,669,048	21,717,755	8.24
denovo PLCA: KVLB anaerobic	18,866,927	16,636,966	11.82
denovo PLCA: MAC anaerobic	21,628,034	18,290,293	15.43
denovo PLCA: McKay anaerobic	415,513	332,612	19.96
adjusted PLCA: CHOC anaerobic	21,878,343	21,192,362	3.14
adjusted PLCA: McKay aerobic	12,561,265	7,943,779	36.8
adjusted PLCA: TSY aerobic	12,783,676	11,509,050	10.0