Supporting Information for "Uncovering hidden members and functions of the soil microbiome using *de novo* metaproteomics"

Joon-Yong Lee[1], Hugh D. Mitchell[1], Meagan C. Burnet[1], Ruonan Wu[1], Sarah C. Jenson[2], Eric D. Merkley[2], Ernesto S. Nakayasu[1], Carrie D. Nicora[1], Janet K. Jansson[1], Kristin E. Burnum-Johnson[3*], Samuel H. Payne[4*]*

[1] Biological Sciences Division, Pacific Northwest National Laboratory, Richland WA

[2] Signature Sciences and Technology Division, Pacific Northwest National Laboratory, Richland WA

[3] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland WA

[4] Biology Department, Brigham Young University, Provo UT

* Corresponding authors

Table of Contents

Supplementary Figures

Supplementary Tables

Supplementary Figures

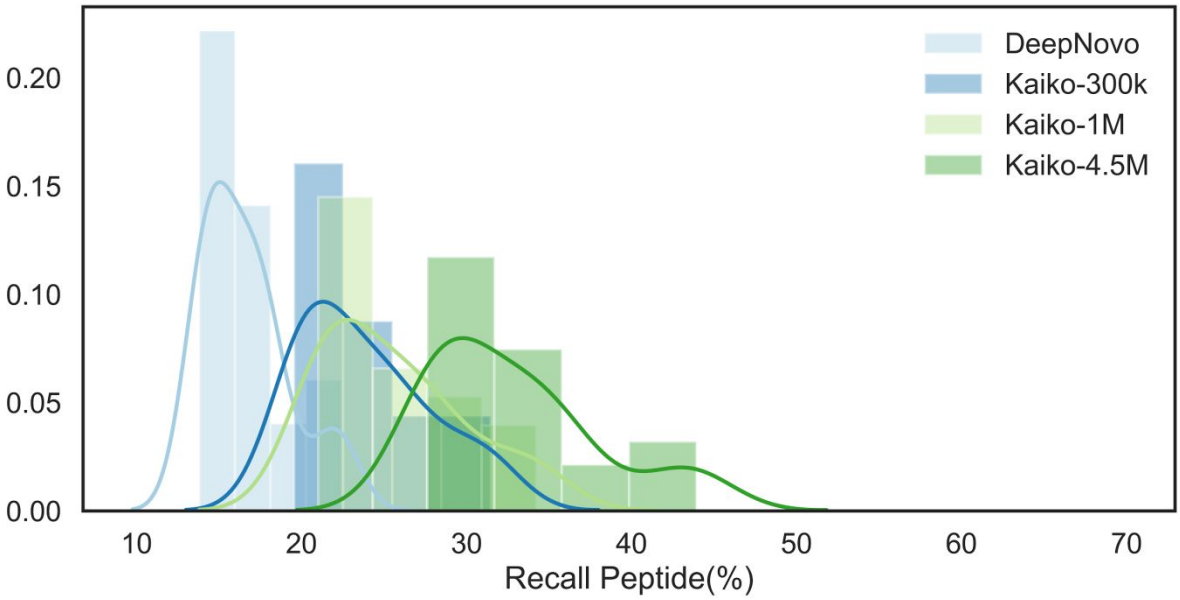

Figure S1. Improving deep neural networks with more training data. The accuracy of peptide/spectrum matching is shown for four deep neural network models. DeepNovo is a pre-trained publicly available model trained on 50,000 spectra. Kaiko was trained with varying numbers of spectra. The final model was trained with 4.5 million spectra. A significant improvement is seen in model performance with increased training data.

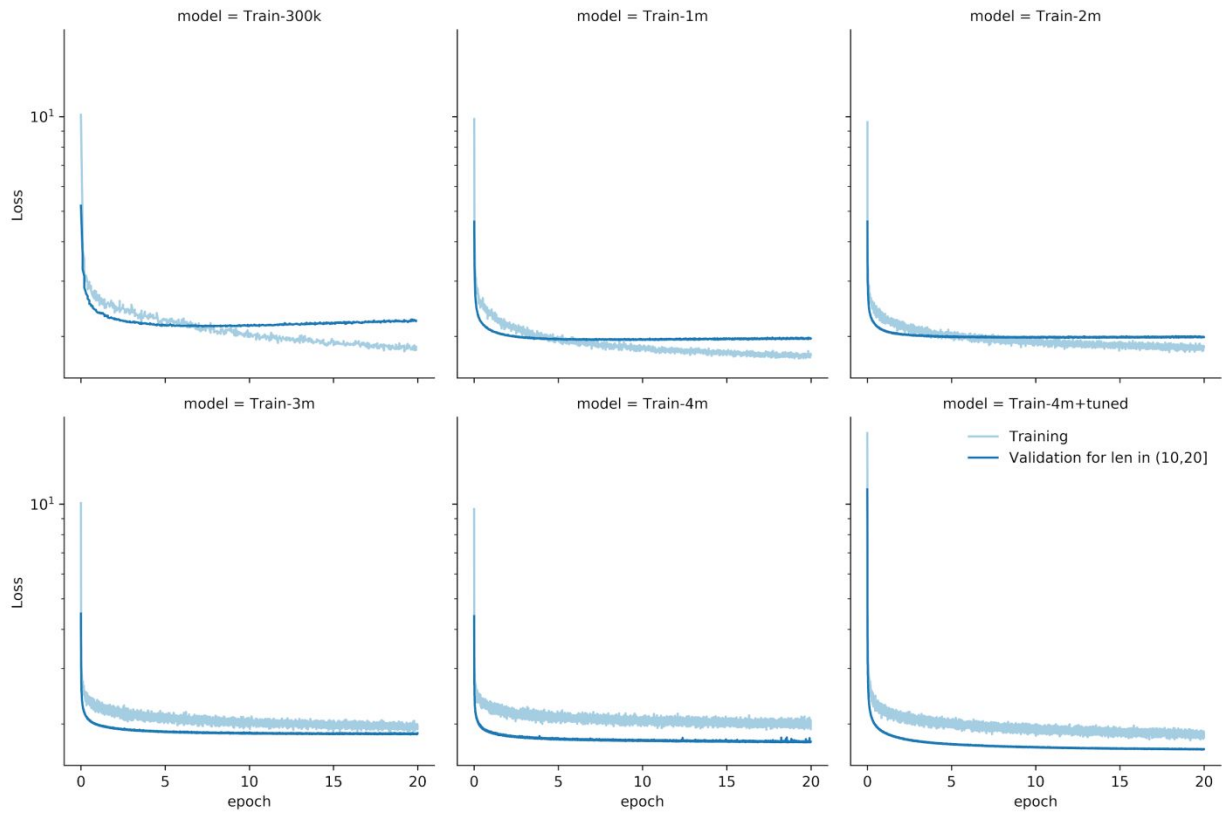

Figure S2. Training and validation error. During the epochs of learning for the deep neural network, progress is measured by evaluating the accuracy of spectrum annotation. We employ a cross-entropy loss function, which represents how well the algorithm is being trained, with small numbers being better. The light blue line represents the error on batches of training data. The dark blue lines represent the error on the random samples of the validation data. When the training error improves beyond the validation error, the model is likely to be overfitting.

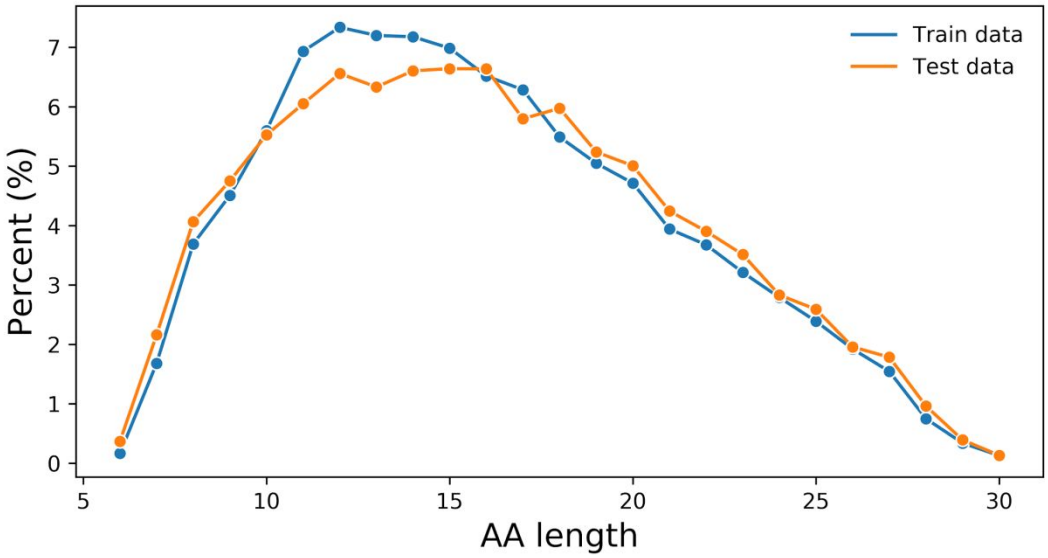

Figure S3. Distribution of peptide lengths used for training and testing the Kaiko model.

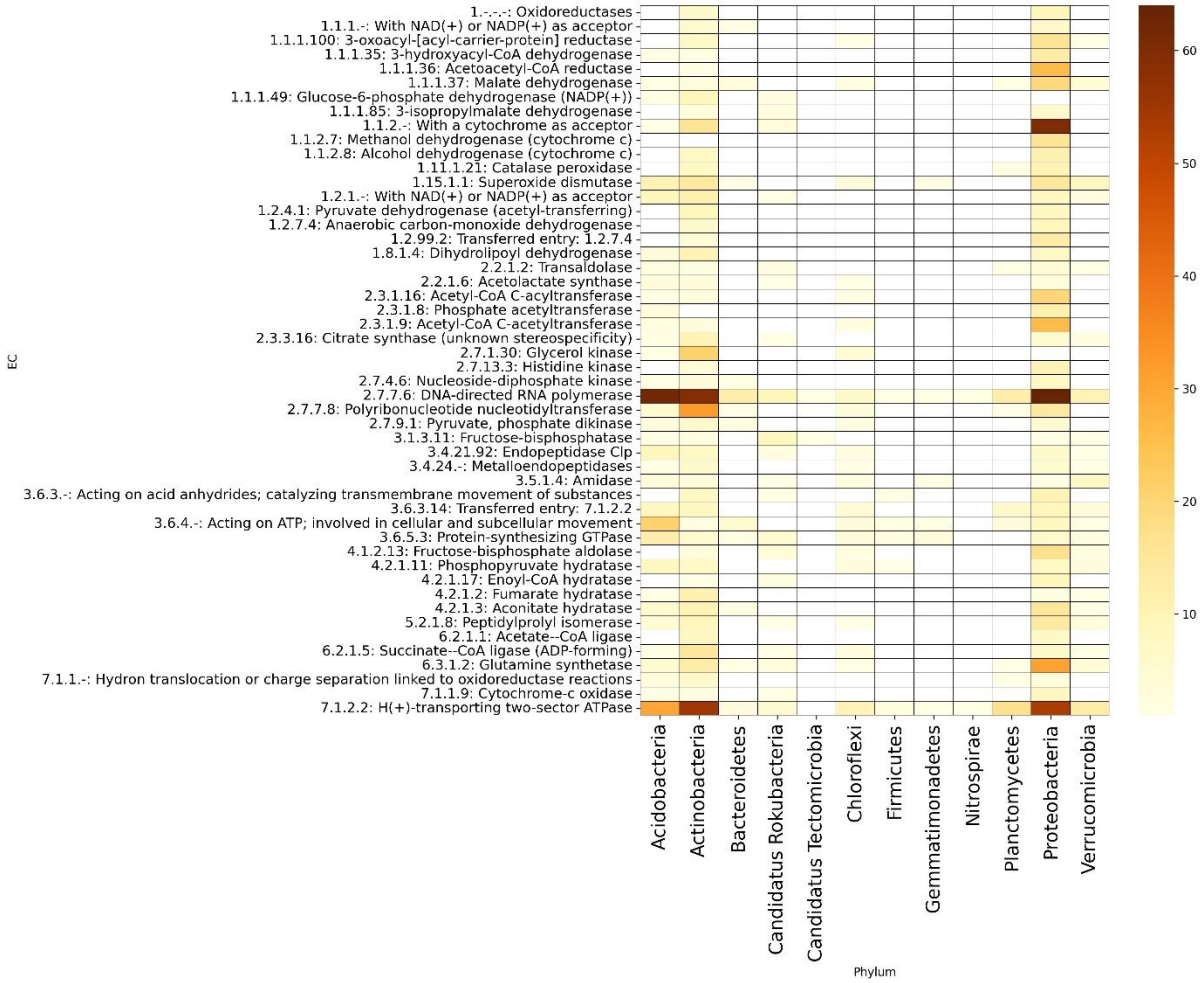

Figure S4. Heatmap of the peptide counts for the most common functions over the diverse phyla. Columns and rows in the heatmap represent the phyla and EC numbers, respectively. Cell colors indicate the number of phyla-affiliated peptides corresponding to a specific phylum and function.

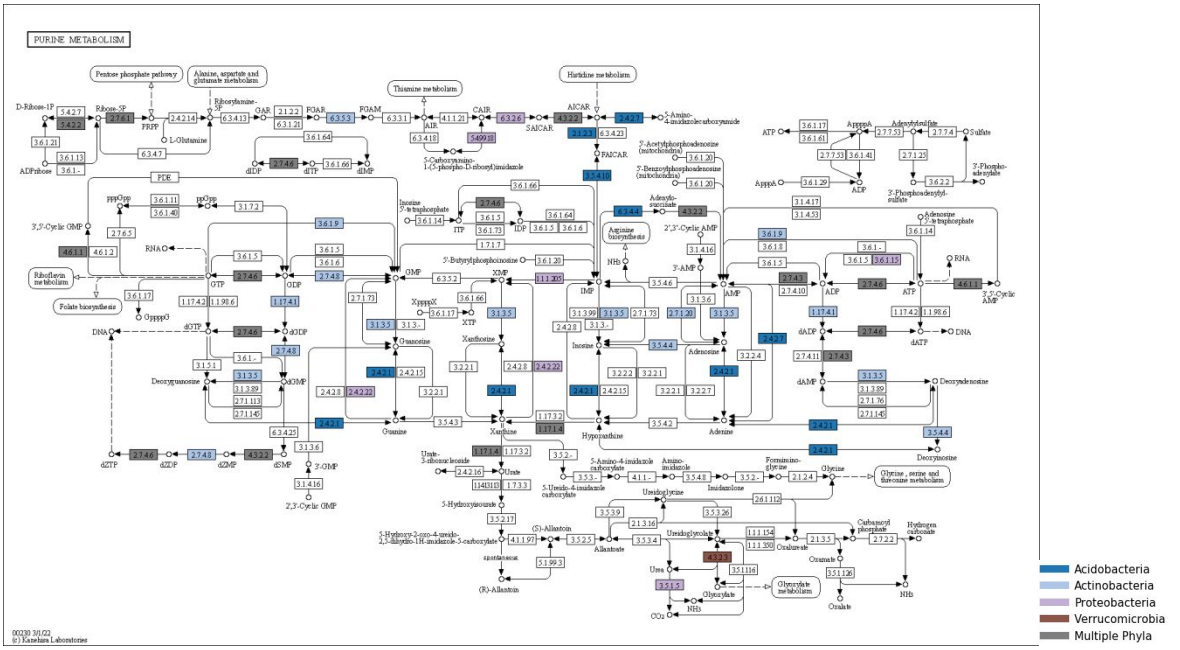

Figure S5. Taxa-specific peptides for enzymes in purine metabolism at the phylum level. Many EC numbers were uniquely detected in four phyla (Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia).

Supplementary Tables

Table S1. LC/MS data files for training and testing the Kaiko model.

Index	Files	Species	# PSM	# Cumulative PSM
0	Biodiversity_A_cryptum_FeTSB_anaerobic_1_01Jun16_Pippin_16-03-39	Acidiphilium_cryptum_JF-5	6659	6659
1	Biodiversity_A_cryptum_FeTSB_anaerobic_2_01Jun16_Pippin_16-03-39	Acidiphilium_cryptum_JF-5	8532	15191
2	Biodiversity_A_cryptum_FeTSB_anaerobic_3_01Jun16_Pippin_16-03-39	Acidiphilium_cryptum_JF-5	7379	22570
3	Biodiversity_A_faecalis_LB_aerobic_01_26Feb16_Arwen_16-01-01	Alcaligenes_faecalis	15496	38066
4	Biodiversity_A_faecalis_LB_aerobic_02_26Feb16_Arwen_16-01-01	Alcaligenes_faecalis	15367	53433
5	Biodiversity_A_faecalis_LB_aerobic_03_26Feb16_Arwen_16-01-01	Alcaligenes_faecalis	15035	68468
6	Biodiversity_A_tumefaciens_R2A_aerobic_1_23Nov16_Pippin_16-09-11	Agrobacterium_tumefaciens_IAM_12048	12994	81462
7	Biodiversity_A_tumefaciens_R2A_aerobic_2_23Nov16_Pippin_16-09-11	Agrobacterium_tumefaciens_IAM_12048	12442	93904
8	Biodiversity_A_tumefaciens_R2A_aerobic_3_23Nov16_Pippin_16-09-11	Agrobacterium_tumefaciens_IAM_12048	11916	105820
9	Biodiversity_B_bifidum_CMcarb_anaerobic_01_26Feb16_Arwen_16-01-01	Bifidobacterium_bifidum_ATCC29521	14409	120229
10	Biodiversity_B_bifidum_CMcarb_anaerobic_02_26Feb16_Arwen_16-01-01	Bifidobacterium_bifidum_ATCC29521	13731	133960
11	Biodiversity_B_bifidum_CMcarb_anaerobic_03_26Feb16_Arwen_16-01-01	Bifidobacterium_bifidum_ATCC29521	13854	147814
12	Biodiversity_B_cereus_ATCC14579_LB_aerobic_1_17July16_Samwise_16-04-10	Bacillus_cereus_ATCC14579	23828	171642
13	Biodiversity_B_cereus_ATCC14579_LB_aerobic_2_17July16_Samwise_16-04-10	Bacillus_cereus_ATCC14579	23693	195335
14	Biodiversity_B_cereus_ATCC14579_LB_aerobic_3_17July16_Samwise_16-04-10	Bacillus_cereus_ATCC14579	22460	217795
15	Biodiversity_B_cereus_PN_L_CL_1_09Oct16_Pippin_16-05-06	Bacillus_cereus_ATCC14579	22349	240144
16	Biodiversity_B_cereus_PN_L_CL_2_09Oct16_Pippin_16-05-06	Bacillus_cereus_ATCC14579	22572	262716

17	Biodiversity_B_cereus_PN_L_CL_3_09Oct16_Pippin_16-05-06	Bacillus_cereus_ATCC14579	23153	285869
18	Biodiversity_B_fragilis_01_28Jul15_Arwen_14-12-03	Bacteroides_fragilis_638R	19454	305323
19	Biodiversity_B_fragilis_Carb_01_28Oct15_Arwen_15-07-13	Bacteroides_fragilis_638R	17656	322979
20	Biodiversity_B_fragilis_CMcarb_anaerobic_01_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	21410	344389
21	Biodiversity_B_fragilis_CMcarb_anaerobic_02_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	21703	366092
22	Biodiversity_B_fragilis_CMcarb_anaerobic_03_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	22366	388458
23	Biodiversity_B_fragilis_CMgluc_anaerobic_01_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	19770	408228
24	Biodiversity_B_fragilis_CMgluc_anaerobic_02_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	20803	429031
25	Biodiversity_B_fragilis_CMgluc_anaerobic_03_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	20515	449546
26	Biodiversity_B_fragilis_LB_anaerobic_01_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	21122	470668
27	Biodiversity_B_fragilis_LB_anaerobic_02_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	21756	492424
28	Biodiversity_B_fragilis_LB_anaerobic_03_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	22228	514652
29	Biodiversity_B_fragilis_LIB_aerobic_01_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	20523	535175
30	Biodiversity_B_fragilis_LIB_aerobic_02_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	20037	555212
31	Biodiversity_B_fragilis_LIB_aerobic_03_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	21645	576857
32	Biodiversity_B_fragilis_LIB_anaerobic_01_08Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	1055	577912
33	Biodiversity_B_fragilis_LIB_anaerobic_02_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	22114	600026
34	Biodiversity_B_fragilis_LIB_anaerobic_03_01Feb16_Arwen_15-07-13	Bacteroides_fragilis_638R	21766	621792
35	Biodiversity_B_infantis_CMcarb_anaerobic_01_26Feb16_Arwen_16-01-01	Bifidobacterium_longum_infantis_ATCC15697	11900	633692
36	Biodiversity_B_infantis_CMcarb_anaerobic_02_26Feb16_Arwen_16-01-01	Bifidobacterium_longum_infantis_ATCC15697	12737	646429
37	Biodiversity_B_infantis_CMcarb_anaerobic_03_26Feb16_Arwen_16-01-01	Bifidobacterium_longum_infantis_ATCC15697	11620	658049

38	Biodiversity_B_subtilis_NCIB3610_24h_plates_1_13Jun16_Pippin_16-03-39	Bacillus_subtilis_NCIB3610	14159	672208
39	Biodiversity_B_subtilis_NCIB3610_24h_plates_2_13Jun16_Pippin_16-03-39	Bacillus_subtilis_NCIB3610	12880	685088
40	Biodiversity_B_subtilis_NCIB3610_24h_plates_3_13Jun16_Pippin_16-03-39	Bacillus_subtilis_NCIB3610	12518	697606
41	Biodiversity_B_subtilis_NCIB3610_48h_plates_1_13Jun16_Pippin_16-03-39	Bacillus_subtilis_NCIB3610	9087	706693
42	Biodiversity_B_subtilis_NCIB3610_48h_plates_2_13Jun16_Pippin_16-03-39	Bacillus_subtilis_NCIB3610	8258	714951
43	Biodiversity_B_subtilis_NCIB3610_48h_plates_3_13Jun16_Pippin_16-03-39	Bacillus_subtilis_NCIB3610	5163	720114
44	Biodiversity_B_subtilis_NCIB3610_pellet_1_03May16_Samwise_16-03-32	Bacillus_subtilis_NCIB3610	20922	741036
45	Biodiversity_B_subtilis_NCIB3610_pellet_2_03May16_Samwise_16-03-32	Bacillus_subtilis_NCIB3610	21034	762070
46	Biodiversity_B_subtilis_NCIB3610_plates_1_03May16_Samwise_16-03-32	Bacillus_subtilis_NCIB3610	12240	774310
47	Biodiversity_B_subtilis_NCIB3610_plates_2_03May16_Samwise_16-03-32	Bacillus_subtilis_NCIB3610	13306	787616
48	Biodiversity_B_subtilis_pellet_set2_1_13Jun16_Pippin_16-03-39	Bacillus_subtilis_NCIB3610	17709	805325
49	Biodiversity_B_subtilis_pellet_set2_2_13Jun16_Pippin_16-03-39	Bacillus_subtilis_NCIB3610	17532	822857
50	Biodiversity_B_subtilis_pellet_set2_3_13Jun16_Pippin_16-03-39	Bacillus_subtilis_NCIB3610	18214	841071
51	Biodiversity_B_thet_CMcarb_anaerobic_01_01Feb16_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	22586	863657
52	Biodiversity_B_thet_CMcarb_anaerobic_02_01Feb16_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	25220	888877
53	Biodiversity_B_thet_CMcarb_anaerobic_03_01Feb16_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	22535	911412
54	Biodiversity_B_thet_CMgluc_anaerobic_01_01Feb16_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	20596	932008
55	Biodiversity_B_thet_CMgluc_anaerobic_02_01Feb16_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	20725	952733
56	Biodiversity_B_thet_CMgluc_anaerobic_03_01Feb16_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	20639	973372
57	Biodiversity_B_thet_LB_anaerobic_01_01Feb16_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	22310	995682
58	Biodiversity_B_thet_LB_anaerobic_02_01Feb16_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	20736	1016418

59	Biodiversity_B_thet_LB_anaerobic_03_01Feb16_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	17178	1033596
60	Biodiversity_B_thet_LIB_anaerobic_01_01Feb16_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	23175	1056771
61	Biodiversity_B_thet_LIB_anaerobic_02_01Feb16_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	21920	1078691
62	Biodiversity_B_thet_LIB_anaerobic_03_01Feb16_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	22215	1100906
63	Biodiversity_B_thetaiotaomicron_Carb_01_26Aug15_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	22781	1123687
64	Biodiversity_B_thetaiotaomicron_Glc_01_26Aug15_Arwen_15-07-13	Bacteroides_thetaiotaomicron_VPI-5482	25625	1149312
65	Biodiversity_Bacillus_subtilis_LB_01_27Dec15_Arwen_15-07-13	Bacillus_subtilis_168	23891	1173203
66	Biodiversity_Bacillus_subtilis_LB_02_27Dec15_Arwen_15-07-13	Bacillus_subtilis_168	23513	1196716
67	Biodiversity_Bacillus_subtilis_LB_03_27Dec15_Arwen_15-07-13	Bacillus_subtilis_168	25596	1222312
68	Biodiversity_C_Baltica_T240_R1_C_27Jan16_Arwen_15-07-13	Cellulophaga_baltica_18	23983	1246295
69	Biodiversity_C_Baltica_T240_R1_Inf_27Jan16_Arwen_15-07-13	Cellulophaga_baltica_18	26844	1273139
70	Biodiversity_C_Baltica_T240_R2_C_27Jan16_Arwen_15-07-13	Cellulophaga_baltica_18	26240	1299379
71	Biodiversity_C_Baltica_T240_R2_Inf_27Jan16_Arwen_15-07-13	Cellulophaga_baltica_18	26536	1325915
72	Biodiversity_C_Baltica_T240_R3_C_27Jan16_Arwen_15-07-13	Cellulophaga_baltica_18	27084	1352999
73	Biodiversity_C_Baltica_T240_R3_Inf_27Jan16_Arwen_15-07-13	Cellulophaga_baltica_18	27658	1380657
74	Biodiversity_C_comes_Carb_01_14Sep15_Arwen_15-07-13	Coprococcus_comes_ATCC27758	20528	1401185
75	Biodiversity_C_comes_Glc_01_28Oct15_Arwen_15-07-13	Coprococcus_comes_ATCC27758	21095	1422280
76	Biodiversity_C_comes_LIB_01_28Oct15_Arwen_15-07-13	Coprococcus_comes_ATCC27758	21532	1443812
77	Biodiversity_C_freundii_LB_01_14Sep15_Arwen_15-07-13	Citrobacter_freundii	23455	1467267
78	Biodiversity_C_freundii_LIB_01_28Oct15_Arwen_15-07-13	Citrobacter_freundii	22562	1489829
79	Biodiversity_C_gilvus_GS2_anaerobic_01_01Feb16_Arwen_15-07-13	Cellulomonas_gilvus_ATCC13127	25544	1515373

80	Biodiversity_C_gilvus_GS2_anaerobic_02_01Feb16_Arwen_15-07-13	Cellulomonas_gilvus_ATCC13127	24443	1539816
81	Biodiversity_C_gilvus_GS2_anaerobic_03_01Feb16_Arwen_15-07-13	Cellulomonas_gilvus_ATCC13127	24651	1564467
82	Biodiversity_C_indologenes_LIB_aerobic_01_03May16_Samwise_16-03-32	Chryseobacterium_indologenes	12314	1576781
83	Biodiversity_C_indologenes_LIB_aerobic_02_03May16_Samwise_16-03-32	Chryseobacterium_indologenes	12289	1589070
84	Biodiversity_C_indologenes_LIB_aerobic_03_03May16_Samwise_16-03-32	Chryseobacterium_indologenes	12315	1601385
85	Biodiversity_C_ljungdahlii_CO_anaerobic_1_04Oct16_Pippin_16-05-06	Clostridium_ljungdahlii_DMS_13528	20363	1621748
86	Biodiversity_C_ljungdahlii_CO_anaerobic_2_04Oct16_Pippin_16-05-06	Clostridium_ljungdahlii_DMS_13528	21268	1643016
87	Biodiversity_C_ljungdahlii_CO_anaerobic_3_04Oct16_Pippin_16-05-06	Clostridium_ljungdahlii_DMS_13528	20785	1663801
88	Biodiversity_C_ljungdahlii_Fructose_anaerobic_1_04Oct16_Pippin_16-05-06	Clostridium_ljungdahlii_DMS_13528	21315	1685116
89	Biodiversity_C_ljungdahlii_Fructose_anaerobic_2_04Oct16_Pippin_16-05-06	Clostridium_ljungdahlii_DMS_13528	21903	1707019
90	Biodiversity_C_ljungdahlii_Fructose_anaerobic_3_04Oct16_Pippin_16-05-06	Clostridium_ljungdahlii_DMS_13528	22204	1729223
91	Biodiversity_C_necator_R2A_aerobic_1_23Nov16_Pippin_16-09-11	Cupriavidus_necator_N-1	17734	1746957
92	Biodiversity_C_necator_R2A_aerobic_2_23Nov16_Pippin_16-09-11	Cupriavidus_necator_N-1	16630	1763587
93	Biodiversity_C_necator_R2A_aerobic_3_23Nov16_Pippin_16-09-11	Cupriavidus_necator_N-1	16005	1779592
94	Biodiversity_Cellulomonas_gilvus_GS2_01_27Dec15_Arwen_15-07-13	Cellulomonas_gilvus_ATCC13127	24408	1804000
95	Biodiversity_Cellulomonas_gilvus_GS2_02_27Dec15_Arwen_15-07-13	Cellulomonas_gilvus_ATCC13127	24056	1828056
96	Biodiversity_Cellulomonas_gilvus_GS2_03_27Dec15_Arwen_15-07-13	Cellulomonas_gilvus_ATCC13127	23653	1851709
97	Biodiversity_Cibrobacter_freundii_LB_aerobic_01_01Feb16_Arwen_15-07-13	Citrobacter_freundii	23581	1875290
98	Biodiversity_Cibrobacter_freundii_LB_aerobic_02_01Feb16_Arwen_15-07-13	Citrobacter_freundii	22579	1897869
99	Biodiversity_Cibrobacter_freundii_LB_aerobic_03_01Feb16_Arwen_15-07-13	Citrobacter_freundii	23273	1921142
100	Biodiversity_D_acidovorans_TGY_aerobic_01_29Apr16_Samwise_16-03-32_renamed	Delftia_acidovorans_SPH1	21871	1943013

101	Biodiversity_D_acidovorans_TGY_aerobic_02_29Apr16_Samwise_16-03-32_renamed	Delftia_acidovorans_SPH1	20549	1963562
102	Biodiversity_D_acidovorans_TGY_aerobic_03_29Apr16_Samwise_16-03-32_renamed	Delftia_acidovorans_SPH1	19125	1982687
103	Biodiversity_D_longicatena_CarbI_01_26Aug15_Arwen_15-07-13	Dorea_longicatena_DSM13814	23197	2005884
104	Biodiversity_D_longicatena_CarbII_01_26Aug15_Arwen_15-07-13	Dorea_longicatena_DSM13814	22441	2028325
105	Biodiversity_D_longicatena_Glc_01_28Oct15_Arwen_15-07-13	Dorea_longicatena_DSM13814	19551	2047876
106	Biodiversity_F_novicida_TSB_aerobic_01_01Feb16_Arwen_15-07-13	Francisella_novicida_U112	25900	2073776
107	Biodiversity_F_novicida_TSB_aerobic_02_01Feb16_Arwen_15-07-13	Francisella_novicida_U112	23556	2097332
108	Biodiversity_F_novicida_TSB_aerobic_03_01Feb16_Arwen_15-07-13	Francisella_novicida_U112	23189	2120521
109	Biodiversity_F_prausnitzii_Carb_01_28Oct15_Arwen_15-07-13	Faecalibacterium_prausnitzii	11204	2131725
110	Biodiversity_F_prausnitzii_Glc_01_28Oct15_Arwen_15-07-13	Faecalibacterium_prausnitzii	13901	2145626
111	Biodiversity_F_prausnitzii_LIB_01_28Oct15_Arwen_15-07-13	Faecalibacterium_prausnitzii	12858	2158484
112	Biodiversity_F_succinogenes_MDM_01_27Dec15_Arwen_15-07-13	Fibrobacter_succinogenes_S85	20923	2179407
113	Biodiversity_F_succinogenes_MDM_02_27Dec15_Arwen_15-07-13	Fibrobacter_succinogenes_S85	23278	2202685
114	Biodiversity_F_succinogenes_MDM_03_27Dec15_Arwen_15-07-13	Fibrobacter_succinogenes_S85	21665	2224350
115	Biodiversity_HL111_HLHglutamate_aerobic_1_14July16_Pippin_16-05-01	Erythrobacter_HL-111	16707	2241057
116	Biodiversity_HL111_HLHglutamate_aerobic_2_14July16_Pippin_16-05-01	Erythrobacter_HL-111	15709	2256766
117	Biodiversity_HL111_HLHglutamate_aerobic_3_14July16_Pippin_16-05-01	Erythrobacter_HL-111	17636	2274402
118	Biodiversity_HL48_HLHxylose_aerobic_1_09Jun16_Pippin_16-03-39	Halomonas_HL-48	19564	2293966
119	Biodiversity_HL48_HLHxylose_aerobic_2_09Jun16_Pippin_16-03-39	Halomonas_HL-48	11419	2305385
120	Biodiversity_HL48_HLHxylose_aerobic_3_09Jun16_Pippin_16-03-39	Halomonas_HL-48	19558	2324943
121	Biodiversity_HL49_HLHYE_aerobic_1_05Oct16_Pippin_16-05-06	Algoriphagus_marincola_HL-49	20745	2345688

122	Biodiversity_HL49_HLHYE_aerobic_2_05Oct16_Pippin_16-05-06	Algoriphagus_marincola_HL-49	25153	2370841
123	Biodiversity_HL49_HLHYE_aerobic_3_05Oct16_Pippin_16-05-06	Algoriphagus_marincola_HL-49	24520	2395361
124	Biodiversity_HL69_HLA_aerobic_1_05Oct16_Pippin_16-05-06	Cyanobacterium_stanieri	15059	2410420
125	Biodiversity_HL69_HLA_aerobic_2_05Oct16_Pippin_16-05-06	Cyanobacterium_stanieri	14529	2424949
126	Biodiversity_HL69_HLA_aerobic_3_05Oct16_Pippin_16-05-06	Cyanobacterium_stanieri	16531	2441480
127	Biodiversity_HL91_HLHsucrose_aerobic_1_09Jun16_Pippin_16-03-39	Rhodobacteraceae_bacterium_HL-91	18271	2459751
128	Biodiversity_HL91_HLHsucrose_aerobic_2_09Jun16_Pippin_16-03-39	Rhodobacteraceae_bacterium_HL-91	18280	2478031
129	Biodiversity_HL91_HLHsucrose_aerobic_3_09Jun16_Pippin_16-03-39	Rhodobacteraceae_bacterium_HL-91	19857	2497888
130	Biodiversity_HL93_HLHfructose_aerobic_1_09Jun16_Pippin_16-03-39	Halomonas_HL-93	9976	2507864
131	Biodiversity_HL93_HLHfructose_aerobic_2_09Jun16_Pippin_16-03-39	Halomonas_HL-93	9277	2517141
132	Biodiversity_HL93_HLHfructose_aerobic_3_09Jun16_Pippin_16-03-39	Halomonas_HL-93	8961	2526102
133	Biodiversity_L_monocytogenes_BHI_aerobic_01_27Feb17_Pippin_16-11-03	Listeria_monocytogenes_10403S	27172	2553274
134	Biodiversity_L_monocytogenes_BHI_aerobic_02_27Feb17_Pippin_16-11-03	Listeria_monocytogenes_10403S	25972	2579246
135	Biodiversity_L_monocytogenes_BHI_aerobic_03_27Feb17_Pippin_16-11-03	Listeria_monocytogenes_10403S	26591	2605837
136	Biodiversity_Lactobacillus_casei_MRS_01_27Dec15_Arwen_15-07-13	Lactobacillales_casei	11555	2617392
137	Biodiversity_Lactobacillus_casei_MRS_02_27Dec15_Arwen_15-07-13	Lactobacillales_casei	10984	2628376
138	Biodiversity_Lactobacillus_casei_MRS_03_27Dec15_Arwen_15-07-13	Lactobacillales_casei	12133	2640509
139	Biodiversity_M_luteus_LIB_aerobic_01_26Feb16_Arwen_16-01-01	Micrococcus_luteus	14623	2655132
140	Biodiversity_M_luteus_LIB_aerobic_02_26Feb16_Arwen_16-01-01	Micrococcus_luteus	14694	2669826
141	Biodiversity_M_luteus_LIB_aerobic_03_26Feb16_Arwen_16-01-01	Micrococcus_luteus	14865	2684691
142	Biodiversity_M_smegmatis_BHI_aerobic_1_05Oct16_Pippin_16-05-06	Mycobacterium_smegmatis	22302	2706993

143	Biodiversity_M_smegmatis_BHI_aerobic_2_05Oct16_Pippin_16-05-06	Mycobacterium_smegmatis	23937	2730930
144	Biodiversity_M_smegmatis_BHI_aerobic_3_05Oct16_Pippin_16-05-06	Mycobacterium_smegmatis	23123	2754053
145	Biodiversity_M_xanthus_DZ2_24h_plates_1_13Jun16_Pippin_16-03-39	Myxococcus_xanthus_DZ2	17676	2771729
146	Biodiversity_M_xanthus_DZ2_24h_plates_2_13Jun16_Pippin_16-03-39	Myxococcus_xanthus_DZ2	18291	2790020
147	Biodiversity_M_xanthus_DZ2_24h_plates_3_13Jun16_Pippin_16-03-39	Myxococcus_xanthus_DZ2	17586	2807606
148	Biodiversity_M_xanthus_DZ2_48h_plates_1_13Jun16_Pippin_16-03-39	Myxococcus_xanthus_DZ2	20435	2828041
149	Biodiversity_M_xanthus_DZ2_48h_plates_2_13Jun16_Pippin_16-03-39	Myxococcus_xanthus_DZ2	18715	2846756
150	Biodiversity_M_xanthus_DZ2_48h_plates_3_13Jun16_Pippin_16-03-39	Myxococcus_xanthus_DZ2	19998	2866754
151	Biodiversity_M_xanthus_DZ2_pellet_1_03May16_Samwise_16-03-32	Myxococcus_xanthus_DZ2	24459	2891213
152	Biodiversity_M_xanthus_DZ2_pellet_2_03May16_Samwise_16-03-32	Myxococcus_xanthus_DZ2	24181	2915394
153	Biodiversity_M_xanthus_DZ2_plates_1_03May16_Samwise_16-03-32	Myxococcus_xanthus_DZ2	18520	2933914
154	Biodiversity_M_xanthus_DZ2_plates_2_03May16_Samwise_16-03-32	Myxococcus_xanthus_DZ2	17909	2951823
155	Biodiversity_M_xanthus_pellet_set2_1_13Jun16_Pippin_16-03-39	Myxococcus_xanthus_DZ2	23354	2975177
156	Biodiversity_M_xanthus_pellet_set2_2_13Jun16_Pippin_16-03-39	Myxococcus_xanthus_DZ2	23884	2999061
157	Biodiversity_M_xanthus_pellet_set2_3_13Jun16_Pippin_16-03-39	Myxococcus_xanthus_DZ2	22772	3021833
158	Biodiversity_P_denitrificans_LIB_aerobic_01_29Apr16_Samwise_16-03-32_renamed	Paracoccus_denitrificans	22165	3043998
159	Biodiversity_P_denitrificans_LIB_aerobic_02_29Apr16_Samwise_16-03-32_renamed	Paracoccus_denitrificans	20888	3064886
160	Biodiversity_P_denitrificans_LIB_aerobic_03_29Apr16_Samwise_16-03-32_renamed	Paracoccus_denitrificans	23115	3088001
161	Biodiversity_P_hydrogenalis_01_28Jul15_Arwen_14-12-03	Anaerococcus_hydrogenalis_DSM_7454	18519	3106520
162	Biodiversity_P_hydrogenalis_CMgluc_anaerobic_01_26Feb16_Arwen_16-01-01	Anaerococcus_hydrogenalis_DSM_7454	12813	3119333
163	Biodiversity_P_hydrogenalis_CMgluc_anaerobic_02_26Feb16_Arwen_16-01-01	Anaerococcus_hydrogenalis_DSM_7454	13371	3132704

164	Biodiversity_P_hydrogenalis_CMgluc_anaerobic_03_26Feb16_Arwen_16-01-01	Anaerococcus_hydrogenalis_DSM_7454	12649	3145353
165	Biodiversity_P_polymyxa_TBS_aerobic_1_17July16_Samwise_16-04-10	Paenibacillus_polymyxa_ATCC842	25623	3170976
166	Biodiversity_P_polymyxa_TBS_aerobic_2_17July16_Samwise_16-04-10	Paenibacillus_polymyxa_ATCC842	25057	3196033
167	Biodiversity_P_polymyxa_TBS_aerobic_3_17July16_Samwise_16-04-10	Paenibacillus_polymyxa_ATCC842	24268	3220301
168	Biodiversity_P_ruminicola_MDM_anaerobic_1_09Jun16_Pippin_16-03-39	Prevotella_ruminicola_23_ATCC_19189	17277	3237578
169	Biodiversity_P_ruminicola_MDM_anaerobic_2_09Jun16_Pippin_16-03-39	Prevotella_ruminicola_23_ATCC_19189	17543	3255121
170	Biodiversity_R_gnavus_01_28Jul15_Arwen_14-12-03	Ruminococcus_gnavus	20132	3275253
171	Biodiversity_R_gnavus_Carb_01_28Oct15_Arwen_15-07-13	Ruminococcus_gnavus	22004	3297257
172	Biodiversity_R_jostii_R2A_aerobic_1_23Nov16_Pippin_16-09-11	Rhodococcus_jostii_RHA1	24374	3321631
173	Biodiversity_R_jostii_R2A_aerobic_2_23Nov16_Pippin_16-09-11	Rhodococcus_jostii_RHA1	23736	3345367
174	Biodiversity_R_jostii_R2A_aerobic_3_23Nov16_Pippin_16-09-11	Rhodococcus_jostii_RHA1	22296	3367663
175	Biodiversity_R_palustris_PM_aerobic_1_01Jun16_Pippin_16-03-39	Rhodopseudomonas_palustris	21988	3389651
176	Biodiversity_R_palustris_PM_aerobic_2_01Jun16_Pippin_16-03-39	Rhodopseudomonas_palustris	21998	3411649
177	Biodiversity_R_palustris_PM_aerobic_3_01Jun16_Pippin_16-03-39	Rhodopseudomonas_palustris	20740	3432389
178	Biodiversity_R_palustris_PMnitro_anaerobic_1_01Jun16_Pippin_16-03-39	Rhodopseudomonas_palustris	20820	3453209
179	Biodiversity_R_palustris_PMnitro_anaerobic_2_01Jun16_Pippin_16-03-39	Rhodopseudomonas_palustris	20800	3474009
180	Biodiversity_R_palustris_PMnitro_anaerobic_3_01Jun16_Pippin_16-03-39	Rhodopseudomonas_palustris	19401	3493410
181	Biodiversity_R_palustris_PMnonnitro_anaerobic_1_01Jun16_Pippin_16-03-39	Rhodopseudomonas_palustris	21220	3514630
182	Biodiversity_R_palustris_PMnonnitro_anaerobic_2_01Jun16_Pippin_16-03-39	Rhodopseudomonas_palustris	21947	3536577
183	Biodiversity_R_palustris_PMnonnitro_anaerobic_3_01Jun16_Pippin_16-03-39	Rhodopseudomonas_palustris	20793	3557370
184	Biodiversity_S_agalactiae_LIB_aerobic_01_26Feb16_Arwen_16-01-01	Streptococcus_agalactiae	12558	3569928

185	Biodiversity_S_agalactiae_LIB_aerobic_02_26Feb16_Arwen_16-01-01	Streptococcus_agalactiae	11366	3581294
186	Biodiversity_S_agalactiae_LIB_aerobic_03_26Feb16_Arwen_16-01-01	Streptococcus_agalactiae	11845	3593139
187	Biodiversity_S_aurantiaca_CYE_aerobic_1_17July16_Samwise_16-04-10	Stigmatella_aurantiaca_DW431	26687	3619826
188	Biodiversity_S_aurantiaca_CYE_aerobic_2_17July16_Samwise_16-04-10	Stigmatella_aurantiaca_DW431	25198	3645024
189	Biodiversity_S_aurantiaca_CYE_aerobic_3_17July16_Samwise_16-04-10	Stigmatella_aurantiaca_DW431	28243	3673267
190	Biodiversity_S_elongatus_BG11_aerobic_1_14July16_Pippin_16-05-01	Synechococcus_elongatus_PCC 7942	16601	3689868
191	Biodiversity_S_elongatus_BG11_aerobic_2_14July16_Pippin_16-05-01	Synechococcus_elongatus_PCC 7942	16512	3706380
192	Biodiversity_S_elongatus_BG11_aerobic_3_14July16_Pippin_16-05-01	Synechococcus_elongatus_PCC 7942	16383	3722763
193	Biodiversity_S_elongatus_BG11NaCl_aerobic_1_05Oct16_Pippin_16-05-06	Synechococcus_elongatus_PCC 7942	18618	3741381
194	Biodiversity_S_elongatus_BG11NaCl_aerobic_2_05Oct16_Pippin_16-05-06	Synechococcus_elongatus_PCC 7942	19025	3760406
195	Biodiversity_S_elongatus_BG11NaCl_aerobic_3_05Oct16_Pippin_16-05-06	Synechococcus_elongatus_PCC 7942	18419	3778825
196	Biodiversity_S_griseorubens_HSM_aerobic_1_23Nov16_Pippin_16-09-11	Streptomyces_griseorubens	7798	3786623
197	Biodiversity_S_griseorubens_HSM_aerobic_2_23Nov16_Pippin_16-09-11	Streptomyces_griseorubens	7869	3794492
198	Biodiversity_S_griseorubens_HSM_aerobic_3_23Nov16_Pippin_16-09-11	Streptomyces_griseorubens	5960	3800452
199	Biodiversity_S_thermosulf_FeYE_anaerobic_1_01Jun16_Pippin_16-03-39	Sulfobacillus_thermosulfidooxidans	14607	3815059
200	Biodiversity_S_thermosulf_FeYE_anaerobic_2_01Jun16_Pippin_16-03-39	Sulfobacillus_thermosulfidooxidans	14762	3829821
201	Biodiversity_S_thermosulf_FeYE_anaerobic_3_01Jun16_Pippin_16-03-39	Sulfobacillus_thermosulfidooxidans	15862	3845683
202	Cj_media_MH_R1_23Feb15_Arwen_14-12-03	Campylobacter_jejuni	24941	3870624
203	Cj_media_MH_R2_23Feb15_Arwen_14-12-03	Campylobacter_jejuni	24931	3895555
204	Cj_media_MH_R3_23Feb15_Arwen_14-12-03	Campylobacter_jejuni	21400	3916955
205	Cj_media_MH_R4_23Feb15_Arwen_14-12-03	Campylobacter_jejuni	25037	3941992

206	Cj_media_MH_R5_23Feb15_Arwen_14-12-03	Campylobacter_jejuni	20168	3962160
207	LP_LS_Phi_Stat_R1_30Sep14_Pippin_13-04-12	Legionella_pneumophila	26779	3988939
208	LP_LS_Phi_Stat_R2_30Sep14_Pippin_13-04-12	Legionella_pneumophila	27771	4016710
209	LP_LS_Phi_Stat_R3_30Sep14_Pippin_13-04-12	Legionella_pneumophila	25938	4042648
210	P_putida_01Dec15_1_21Mar16_Arwen_16-01-03	Pseudomonas_putida_KT2440	23697	4066345
211	P_putida_01Dec15_2_21Mar16_Arwen_16-01-03	Pseudomonas_putida_KT2440	24322	4090667
212	P_putida_17Nov15_1_21Mar16_Arwen_16-01-03	Pseudomonas_putida_KT2440	22904	4113571
213	P_putida_17Nov15_2_21Mar16_Arwen_16-01-03	Pseudomonas_putida_KT2440	22395	4135966
214	P_putida_18Nov15_1_21Mar16_Arwen_16-01-03	Pseudomonas_putida_KT2440	22246	4158212
215	P_putida_18Nov15_2_21Mar16_Arwen_16-01-03	Pseudomonas_putida_KT2440	22937	4181149
216	S_venezuelae_GYM_1_21Mar16_Arwen_16-01-03	Streptomyces_venezuelae	12030	4193179
217	S_venezuelae_GYM_2_21Mar16_Arwen_16-01-03	Streptomyces_venezuelae	11099	4204278
218	S_venezuelae_MYM_1_21Mar16_Arwen_16-01-03	Streptomyces_venezuelae	14276	4218554
219	S_venezuelae_MYM_2_21Mar16_Arwen_16-01-03	Streptomyces_venezuelae	14322	4232876
220	QC_Shew_13_05_500ng_2_100uL_5hr_30Mar14_Samwise_13-07-17	Shewanella_oneidensis_MR-1	49123	4281999
221	QC_Shew_13_05_500ng_2_5hr_19Mar14_Samwise_13-07-17	Shewanella_oneidensis_MR-1	50274	4332273
222	QC_Shew_13_05_500ng_2_5hr_24Mar14_Samwise_13-07-17	Shewanella_oneidensis_MR-1	50273	4382546
223	M_alcali_copp_CH4_B1_T1_07_QE_23Mar18_Oak_18-01-07	Methylomicrobium_alcaliphilum	21112	4403658
224	M_alcali_copp_CH4_B1_T2_08_QE_23Mar18_Oak_18-01-07	Methylomicrobium_alcaliphilum	21074	4424732
225	M_alcali_copp_CH4_B2_T1_09_QE_23Mar18_Oak_18-01-07	Methylomicrobium_alcaliphilum	18470	4443202
226	M_alcali_copp_CH4_B2_T2_10_QE_23Mar18_Oak_18-01-07	Methylomicrobium_alcaliphilum	18257	4461459
227	M_alcali_copp_CH4_B3_T1_11_QE_23Mar18_Oak_18-01-07	Methylomicrobium_alcaliphilum	20464	4481923

228	M_alcali_copp_CH4_B3_T2_12_QE_23Mar18_Oak_18-01-07	Methylomicrobium_alcaliphilum	20368	4502291
229	M_alcali_copp_MeOH_B1_T1_01_QE_23Mar18_Oak_18-01-07	Methylomicrobium_alcaliphilum	18067	4520358
230	M_alcali_copp_MeOH_B1_T2_02_QE_23Mar18_Oak_18-01-07	Methylomicrobium_alcaliphilum	18229	4538587
231	M_alcali_copp_MeOH_B2_T1_03_QE_23Mar18_Oak_18-01-07	Methylomicrobium_alcaliphilum	17445	4556032
232	M_alcali_copp_MeOH_B2_T2_04_QE_23Mar18_Oak_18-01-07	Methylomicrobium_alcaliphilum	18191	4574223
233	M_alcali_copp_MeOH_B3_T1_05_QE_23Mar18_Oak_18-01-07	Methylomicrobium_alcaliphilum	15781	4590004
234	M_alcali_copp_MeOH_B3_T2_06_QE_23Mar18_Oak_18-01-07	Methylomicrobium_alcaliphilum	14536	4604540
235	Alverdy_Efae_1A_lys_13Jul13_Pippin_12-12-39	Enterococcus_faecalis	16181	4620721
236	Alverdy_Efae_1B_lys_13Jul13_Pippin_12-12-39	Enterococcus_faecalis	16055	4636776
237	Alverdy_Efae_1C_lys_13Jul13_Pippin_12-12-39	Enterococcus_faecalis	15814	4652590
238	Biodiversity_A_muciniphila_test_27Feb17_Pippin_16-11-03	Akkermansia_muciniphila_ATCC_BAA-835	21214	4673804
239	Ha_150NaCl_1_13_QE_12Aug15_Arwen_14-12-03	Halanaerobium_congolense	27902	4701706
240	Ha_150NaCl_2_14_QE_12Aug15_Arwen_14-12-03	Halanaerobium_congolense	29961	4731667
241	Ha_150NaCl_3_15_QE_12Aug15_Arwen_14-12-03	Halanaerobium_congolense	27892	4759559
242	Ha_200NaCl_1_22_QE_21Jan16_Arwen_15-07-13	Halanaerobium_congolense	29705	4789264
243	Ha_200NaCl_2_23_QE_21Jan16_Arwen_15-07-13	Halanaerobium_congolense	28807	4818071
244	Ha_200NaCl_3_24_QE_21Jan16_Arwen_15-07-13	Halanaerobium_congolense	29544	4847615
245	Ha_250NaCl_1_16_QE_12Aug15_Arwen_14-12-03	Halanaerobium_congolense	15326	4862941
246	Ha_250NaCl_2_17_QE_12Aug15_Arwen_14-12-03	Halanaerobium_congolense	19355	4882296
247	Ha_250NaCl_3_18_QE_12Aug15_Arwen_14-12-03	Halanaerobium_congolense	22461	4904757
248	YJ_Cc_WT1_C_P_9Jan17_Pippin_16-09-11	Caulobacter_crescentus_NA1000	22820	4927577
249	YJ_Cc_WT1_IM_P_9Jan17_Pippin_16-09-11	Caulobacter_crescentus_NA1000	19081	4946658

250	YJ_Cc_WT1_OM_P_9Jan17_Pippin_16-09-11	Caulobacter_crescentus_NA1000	26060	4972718
251	YJ_Cc_WT1_P_Prot_9Jan17_Pippin_16-09-11	Caulobacter_crescentus_NA1000	21323	4994041
252	YJ_Cc_WT1_WC_P_9Jan17_Pippin_16-09-11	Caulobacter_crescentus_NA1000	17975	5012016
253	YJ_Cc_WT2_C_P_9Jan17_Pippin_16-09-11	Caulobacter_crescentus_NA1000	22620	5034636
254	YJ_Cc_WT2_IM_P_9Jan17_Pippin_16-09-11	Caulobacter_crescentus_NA1000	19881	5054517
255	YJ_Cc_WT2_OM_P_9Jan17_Pippin_16-09-11	Caulobacter_crescentus_NA1000	20632	5075149
256	YJ_Cc_WT2_P_Prot_9Jan17_Pippin_16-09-11	Caulobacter_crescentus_NA1000	24577	5099726
257	YJ_Cc_WT2_WC_P_9Jan17_Pippin_16-09-11	Caulobacter_crescentus_NA1000	16579	5116305

Table S2. For each of the six natural isolates, replicate proteomics data was annotated with Kaiko and high-rank taxa were identified from the sequence alignment of identified peptides against the UniRef database. Each sheet in the Supplementary file has the number of hits for each taxon id for 6 natural isolates (2~3 reps for each). It will be attached as an Excel file.

Table S3. Top 20 of EC numbers most frequently matched from the unique peptides using Unipept 4.6.3 with the identified peptide sequences.

EC number	Name	PepCounts
EC:7.1.2.2	H(+)-transporting two-sector ATPase	563
EC:2.7.7.6	DNA-directed RNA polymerase	498
EC:3.6.5.3	Protein-synthesizing GTPase	198
EC:6.3.1.2	Glutamine synthetase	149
EC:1.2.1.-	With NAD(+) or NADP(+) as acceptor	144
EC:1.1.2.-	With a cytochrome as acceptor	110
EC:3.6.3.-	Acting on acid anhydrides; catalyzing transmembrane movement of substances	107
EC:3.6.3.14	Transferred entry: 7.1.2.2	105

EC:2.7.7.8	Polyribonucleotide nucleotidyltransferase	101
EC:3.6.4.-	Acting on ATP; involved in cellular and subcellular movement	96
EC:1.11.1.21	Catalase peroxidase	90
EC:1.1.1.37	Malate dehydrogenase	89
EC:5.2.1.8	Peptidylprolyl isomerase	84
EC:4.2.1.11	Phosphopyruvate hydratase	82
EC:4.2.1.3	Aconitate hydratase	82
EC:1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase	80
EC:2.7.13.3	Histidine kinase	80
EC:1.15.1.1	Superoxide dismutase	71
EC:1.-.-.-	Oxidoreductases	70
EC:1.1.1.-	With NAD(+) or NADP(+) as acceptor	67