

1.0 INTRODUCTION

1.1 Background and Prior Research

Microbiomics is an exciting field with tons of potential for preventative/proactive health care and the development of new biomarkers. We, humans, have thousands of microbes that live inside of us that potentially affect our health. In this study, we will be using a publicly available dataset provided by Junjie Qin and their team through the Knights Lab. Significant findings can potentially lead to developing therapies that help prevent or help treat Type 2 Diabetes (T2D) and its symptoms.

The microbiome is a vast, diverse aspect for every person, and the technology and methods we have cannot completely and accurately measure any person's microbiota. Even defining different microbial species is an enormous challenge. While bioinformaticians and biostatisticians can come up with ways to supplement this, I want to see if I can find anything significant with what we have so far.

In reality, if I wanted to determine whether I had T2D or not, I would take some form of a blood sugar test instead of having to bring in my stool then having that analyzed for signs of microbiota that could maybe indicate T2D. Microbiome research also constitutes defining microbiome differentiation in terms of composition, in other words, what makes a healthy microbiome as compared to what makes up an unhealthy one. But there can be differences between individuals and subgroup populations in the realm of microbiota as there are a variety of factors that influence an individual's microbiome. So, my scientific interest is motivated not by finding complex relationships between microbiota and T2D but identifying universal, individual strains that are associated with T2D. I want to adjust for age and sex because these are factors that can heavily affect microbiome composition. For now, we will ignore the role microbiome plays in potentially affecting sex, specifically hormones. We will focus on the notion that different reproductive organs and hormones can affect the microbiome. Age is especially important to adjust for because aging is driving the prevalence of T2D. To reiterate, my goal is not to define microbial compositions but determine if individual strains affect diabetes or its mechanisms, primarily blood sugar

1.2 Scientific Aims

1. For each OTU, is it associated with hba1c (blood sugar), adjusting for age and sex?
2. For each OTU, is it associated with T2D status, adjusting for age and sex?

2.0 DATA SET DESCRIPTION

Whole-genome shotgun (WGS) sequencing was carried on 145 Chinese individuals, (71 Cases and 74 Controls) living in the southern China. 11 individuals were omitted from the public repository (but their results can be found in supplementary material). So, overall, there were 134 individuals (64 cases and 70 controls). In the original study, 200 more Chinese individuals were sampled and analyzed, but we will not be considering that information. However, through the University of Washington, the additional details about our current participants and more can be found in the supplementary material of the

The study design is case control as individuals were identified as either healthy or diabetic then samples were collected from them. Note, individuals were trusted to provide their own stool by freezing potential samples at home.

3.0 STATISTICAL METHODS

3.1 Exploratory Methods

First, we will assess whether if age is linearly associated with log odds of T2D status by plotting the log odds of T2D by age. Then, we will provide a box plot of T2D by age and a bar plot of T2D by sex.

3.2 Data Preparation

- We will remove all features that are not OTUs, age, or sex (which is labelled as “Gender”) because background research indicates that OTUs could affect them (blood sugar level, choline levels, etc.).
- Next, we will filter for complete genome OTUs only. We are limited on the amount of expertise on how to deal with non-complete genome sequences.
- Then we will filter OTUs for those with counts higher than the total number of OTUs. This is a common standard as it helps reduce possible errors from sequencing.
- We will filter for OTUs with less zeros than the total number of samples (patients). Another practice to possibly reduce the errors from sequencing.
- We will apply a log transformation the counts of OTUs.

3.3 First Aim

For each microbial, we will use a trivariate linear regression model for each OTU and adjust for age and sex to find evidence of association between each OTU on the odds of T2D status. We will provide inference for significant OTUs.

3.4 Second Aim

For each microbial, we will use a trivariate logistic regression model for each OTU and adjust for age and sex to find evidence of association between each OTU on the odds of T2D status. We provide inference for significant OTUs.

3.4.1 Model Assumptions

Each observation represented one stool sample from each patient, so no correlation between patients is expected. Though, it is an interesting discussion whether eating the same food could affect the microbiome of stool samples in similar ways. We attempt to remove potential outliers through our filtering process. With three independent variables in each tri-variable model, we assumed that there would be no multicollinearity as we assume age and sex will potentially be confounders. Finally, we assumed there was a linear relationship between the log odds of T2D and log count of an OTU since our hope is that more or less of a specified OTU affects T2D. We will also assume linear relationship between blood sugar and log OTU count.

For both methods, we will use a Bonferroni and test at a 0.01 alpha level to control the false discovery rate. We will conduct all analysis using R, version 4.0.5.

4.0 RESULTS

4.1 Exploratory Results

From Figure 1, we see that age is linearly associated with the log odds of T2D. We seem from Figure 2, that there appears to be more males than females in our dataset, and that there is a slightly higher number of diabetic individuals for each sex. In Figure 3, we can see that the diabetic box plot is much higher visually than the non-diabetic boxplot, suggesting that age plays a role in T2D status.

4.2 First Aim Results

Our primary analysis results can be found in Appendix 3.

We do not have sufficient evidence, at a 0.01 alpha level and using a Bonferroni correction, that any of the OTUs in each of their specified models, are associated with blood sugar (hba1c) status using Wald tests (respectively; all $p > .01/127$), adjusting for age and sex.

4.3 Second Aim Results

The result of our secondary analysis can be found in Appendix 2. We have significant evidence, at a 0.01 alpha level and using a Bonferroni correction, that *Enterococcus faecalis*, *Lactobacillus acidophilus*, and *Acidaminococcus intestini*, and *Streptococcus gallolyticus* in each of their specified models, are associated with Type 2 Diabetes (T2D) status using Wald tests (respectively; all $p < .001$), adjusting for age and sex. Under the same testing conditions and using the same methods, we do not have significant evidence all other OTUs that they are associated with T2D status, adjusting for age and sex.

We conclude that *Enterococcus faecalis*, *Lactobacillus acidophilus*, and *Acidaminococcus intestini*, have an effect on the odds of T2D, adjusting for age and sex but not for each other.

Adjusting for age and sex, we estimate that between two groups that differ in one order of logarithmic magnitude in count of *Enterococcus faecalis*, the odds of T2D for the group with more of this OTU is 1.79 times higher than the odds of T2D for the group with less.

Adjusting for age and sex, we estimate that between two groups that differ in one order of logarithmic magnitude in count of *Lactobacillus acidophilus*, the odds of T2D for the group with more of this OTU is 2.08 times higher than the odds of T2D for the group with less.

Adjusting for age and sex, we estimate that between two groups that differ in one order of logarithmic magnitude in count of *Acidaminococcus intestini*, the odds of T2D for the group with more of this OTU is 1.67 times higher than the odds of T2D for the group with less.

Adjusting for age and sex, we estimate that between two groups that differ in one order of logarithmic magnitude in count of *Streptococcus gallolyticus*, the odds of T2D for the group with more of this OTU is 0.63 times the odds of T2D for the group with less.

5.0 DIAGNOSTICS & LIMITATIONS

5.1 Model Diagnostics

For accessibility, we will only assess the significant findings. Looking at Figure 4, we see that our significant OTUs were linearly associated with the log odds of T2D. Observing at Figure 5, we see that all our significant OTUs had some correlation with Age. Interestingly, the significant OTUs that were positively associated with log odds of T2D status were also slightly correlated with each other.

5.2 Sources of Error

Any inferences or estimates made can only be generalized to the mainland Chinese population. There were 3 main sources of potential error from the dataset that I identified. One, the sequencing technology could have made mistakes. Two, our OTU classifications could also be incorrect as an arbitrary threshold is used to determine what species a certain organism is. And three, we do not know how accurately human stool represents someone's microbiome. A main source of error on my part is inexperience. Several components were overlooked such as the difference between a complete genome vs a complete sequence.

5.3 Assumption Violations

Through filtering, I believed we could reduce some error stemming from the big 3 above. It is possible that age and an OTU were highly correlated, but we included age because we believe it could be a confounder. It is also possible that outliers still existed after the data preparation stage. We focused on removing OTUs, but we did not filter samples, which could be a source of great variability.

5.4 Limitations

For the purpose of this study, our goal was to determine key OTU species that had strong effects on the odds of T2D. We ignored potential interaction among OTUs. For example, we removed many more "prominent" biomarkers like choline and blood sugar level because microorganisms could affect them. It could be the case, and there most definitely is a case, where microorganisms affect each other or modify each other's effect on the odds of T2D. A way we could have approached this is by analyzing the microbial composition of T2D vs no T2D. Our analysis is useful in a sense that we could potentially determine "bad" or "good" microorganisms, but we could not capture of what a "good" or "bad" microbiome looks like. One reason I decided that this was an appropriate philosophy for this analysis is that in the context of our current medical and scientific affairs and technology, I believe it is more meaningful to focus on identifying a few key OTUs instead of trying to paint a picture of what a healthy microbiome should look like. I think that in reality, the approach of identifying key OTUs is more accessible and achievable for more people. From Figure 5 once more, it indicates that perhaps clustering OTUs then trying to find significant clusters would be a strong approach.

6.0 DISCUSSION

To summarize our findings, we found that that *Enterococcus faecalis*, *Lactobacillus acidophilus*, and *Acidaminococcus intestini*, and *Streptococcus gallolyticus* were associated with the log odds of T2D status, adjusting for age and sex. A quick background research indicates some inconsistency. For example, *Streptococcus gallolyticus* is potentially a main causative agent of septicemia and infective endocarditis but was negatively associated with T2D status. This leads me to believe that is that it could be incorrect to think that a microbial lead to T2D status when it could be that T2D status that affects the microbials. Perhaps, because someone is diabetic, there needs to be an increase of “good” microbials. I do think there is still merit to finding specific strains that affect T2D status because this would allow biomedical researchers to create drugs that can target these strains instead of shifting the course of the microbiome when it is unnecessary. But, after this study, I think the impact of the microbiome on T2D may not be important as other factors such as proper exercise, nutrition, and sleep. But let’s see what the future brings.

7.0 TABLES AND FIGURES

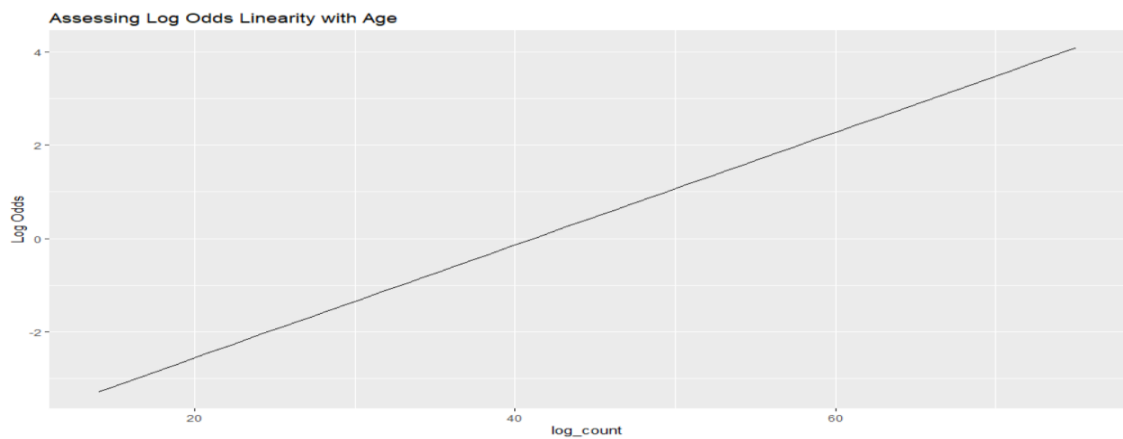


Figure 1

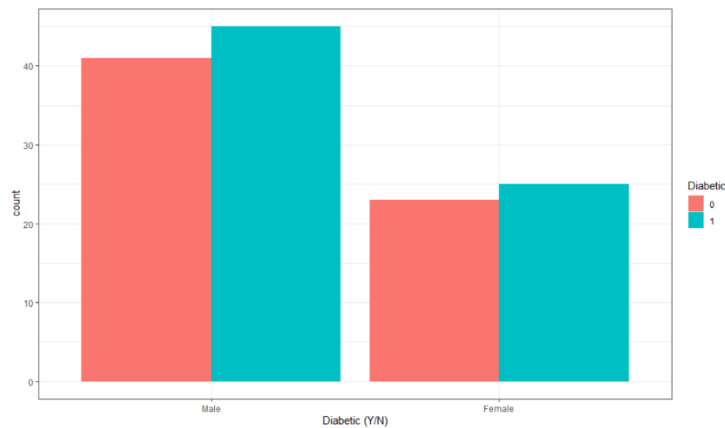


Figure 2

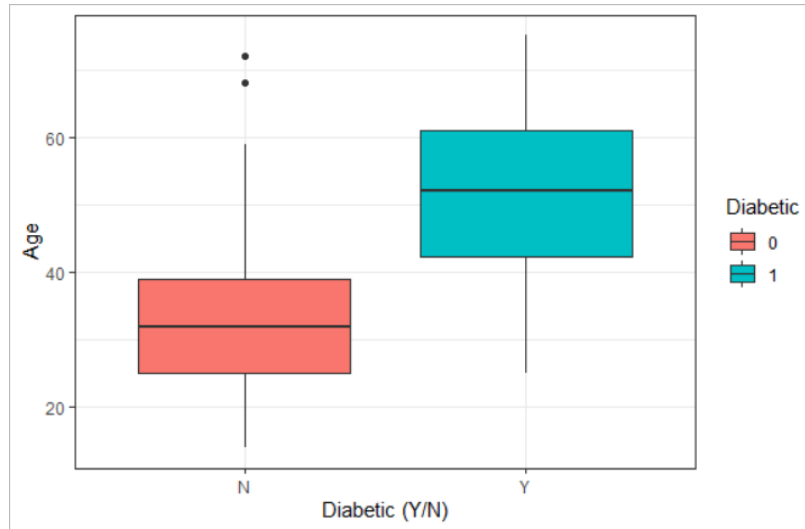


Figure 3

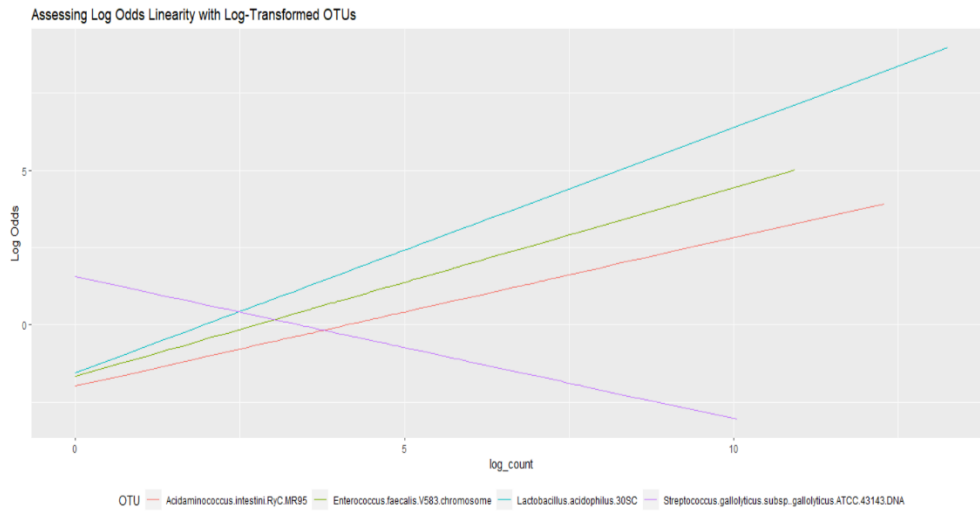


Figure 4

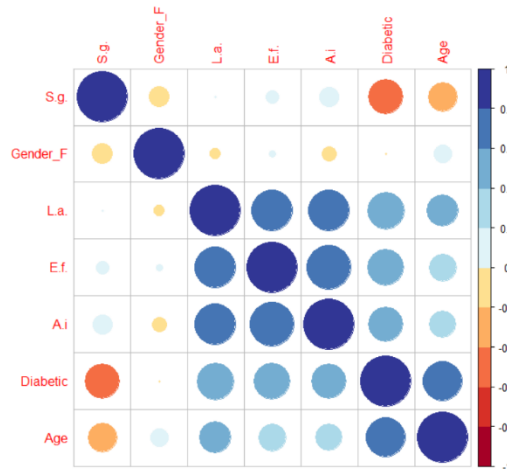


Figure 5

8.0 APPENDICES

8.1 References

- <https://www.sciencedirect.com/science/article/pii/S2352304217300351#bib35>
- <https://www.nature.com/articles/nbt1209-1135>
- <https://www.healthline.com/health/enterococcus-faecalis>
- <https://www.healthline.com/nutrition/lactobacillus-acidophilus>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5891635/>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6920072/>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6477058/>
- <https://dalspace.library.dal.ca/bitstream/handle/10222/72141/Chen-Chang-Msc-STAT-Aug-2016.pdf?sequence=1&isAllowed=y>
- <https://www.bmj.com/content/369/bmj.m997>
- <https://www.webmd.com/diabetes/qa/what-is-candida-albicans-and-where-does-it-occur-in-people-with-diabetes>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6535288/>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4348024/>
- <https://www.thelancet.com/journals/ebiom/article/PIIS235239641930800-X/fulltext>
- <https://www.medicalnewstoday.com/articles/meet-the-bacteria-that-might-help-treat-diabetes#Impacting-metabolism>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6363653/>
- https://knights-lab.github.io/MLRepo/docs/qin_healthy_diabetes.html
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6532624/>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4578629/>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5552927/>
- <https://news.weill.cornell.edu/news/2020/02/gut-microbiome-controls-blood-glucose-levels-through-the-liver>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6445422/>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3705322/>
- <https://www.researchgate.net/post/What-statistical-test-to-use-to-find-out-what-taxa-significantly-differ-between-several-treatments>
- <https://www.nicholas-ollberding.com/post/introduction-to-the-statistical-analysis-of-microbiome-data-in-r/>
- <https://peterlangfelder.com/2018/11/25/filtering-and-collapsing-data/>
- <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0237779>

Appendix II

	Estimate	Std. Error	z value	P_Value
<i>Lactobacillus.acidophilus</i> .30SC	0.732	0.169	4.326	1.52e-05
<i>Enterococcus.faecalis</i> .V583.chromosome	0.584	0.138	4.229	2.34e-05
<i>Acidaminococcus.intestini</i> .RyC.MR95	0.511	0.125	4.085	4.41e-05
<i>Streptococcus.gallolyticus</i> .subsp..gallolyticus.ATCC.43143.DNA	-0.467	0.117	-3.999	6.37e-05
<i>Enterococcus.faecium</i> .DO.chromosome	0.661	0.180	3.682	0.000232
<i>Salmonella.enterica</i> .subsp..enterica.serovar.Typhi.str..CT18.plasmid.pHCM1	-0.414	0.133	-3.102	0.00192
<i>Streptococcus.suis</i> .BM407.chromosome	0.351	0.115	3.063	0.00219
<i>Aggregatibacter.aphrophilus</i> .strain.W10433	-0.363	0.150	-2.418	0.0156
<i>Haemophilus.influenzae</i> .Rd.KW20.chromosome	-0.353	0.147	-2.405	0.0162
<i>Shigella.dysenteriae</i> .Sd197.chromosome	-0.197	0.084	-2.353	0.0186
<i>Bifidobacterium.thermophilum</i> .RBL67	-0.298	0.138	-2.154	0.0313
<i>Lactobacillus.mucosae</i> .LM1	0.210	0.104	2.017	0.0437
<i>Enterobacteria</i> .phage.P1	-0.192	0.108	-1.769	0.0768
<i>Clostridium.phytofermentans</i> .ISDg	-0.271	0.178	-1.518	0.129
<i>Kluyvera.intermedia</i> .strain.CAV1151.plasmid.pCAV1151.215	-0.194	0.131	-1.480	0.139
<i>Obesumbacterium.proteus</i> .strain.DSM.2777	-0.184	0.127	-1.445	0.148
<i>Porphyromonas.gingivalis</i> .ATCC.33277.DNA	0.238	0.167	1.423	0.155
<i>Akkermansia.muciniphila</i> .ATCC.BAA.835	0.089	0.076	1.177	0.239
<i>Salmonella.enterica</i> .subsp..enterica.serovar.Typhimurium.str..LT2	-0.196	0.178	-1.107	0.268
<i>Streptococcus.parasanguinis</i> .ATCC.15912	-0.104	0.099	-1.059	0.29
<i>Streptococcus.gordonii</i> .str..Challis.substr..CH1	0.122	0.115	1.056	0.291
<i>Streptococcus.anginosus</i> .C238	0.112	0.106	1.052	0.293
<i>Yersinia.pestis</i> .CO92.chromosome	-0.152	0.148	-1.029	0.304
<i>Lachnospiraceae.bacterium</i> .2_1_46FAA.genomic.scaffold.acIXT.supercont	-0.096	0.097	-0.990	0.322
<i>Escherichia.coli</i> .O157.H7.str..Sakai.plasmid.pO157	-0.102	0.105	-0.971	0.332
<i>Enterobacter.cloacae</i> .subsp..cloacae.ATCC.13047.plasmid.pECL_A	-0.139	0.148	-0.937	0.349
<i>Barnesiella.viscericola</i> .DSM.18177	-0.087	0.094	-0.931	0.352
<i>Klebsiella.pneumoniae</i> .subsp..pneumoniae.HS11286.plasmid.pKPHS3	0.112	0.121	0.929	0.353
<i>Bifidobacterium.breve</i> .DSM.20213...JCM.1192.DNA	0.073	0.079	0.928	0.353
<i>Escherichia.coli</i> .IAI39.chromosome	-0.070	0.076	-0.923	0.356
<i>Salmonella.enterica</i> .subsp..enterica.serovar.Typhi.str..CT18	-0.107	0.121	-0.882	0.378

	Estimate	Std. Error	z value	P_Value
<i>Prevotella.enoea</i> .strain.F0113	0.127	0.145	0.876	0.381
<i>Bifidobacterium.adolescentis</i> .ATCC.15703.DNA	-0.060	0.070	-0.853	0.394
<i>Adlercreutzia.equolifaciens</i> .DSM.19450.DNA	-0.073	0.086	-0.847	0.397
<i>Citrobacter.freundii</i> .CFNIH1.plasmid.pKEC.a3c	0.105	0.125	0.841	0.401
<i>Clostridium.sp.</i> .SY8519.DNA	-0.087	0.105	-0.828	0.407
<i>Citrobacter.freundii</i> .CFNIH1	-0.086	0.105	-0.824	0.41
<i>Staphylococcus.haemolyticus</i> .JCSC1435.DNA	-0.093	0.113	-0.822	0.411
<i>Veillonella.parvula</i> .DSM.2008	-0.074	0.090	-0.816	0.415
<i>Rothia.mucilaginosa</i> .DY.18.DNA	-0.105	0.130	-0.808	0.419
<i>Shigella.flexneri</i> .2a.str..301.chromosome	-0.063	0.078	-0.802	0.423
<i>Filifactor.alocis</i> .ATCC.35896	-0.111	0.144	-0.773	0.44
<i>Cronobacter.sakazakii</i> .strain.ATCC.29544	-0.079	0.107	-0.740	0.459
<i>Escherichia.coli</i> .UMN026	0.055	0.075	0.739	0.46
<i>Escherichia.coli</i> .O104.H4.str..2011C.3493.plasmid.pAA.EA11	-0.077	0.107	-0.720	0.472
<i>Shigella.phage</i> .SfV	-0.089	0.127	-0.698	0.485
<i>Bifidobacterium.bifidum</i> .PRL2010.chromosome	-0.059	0.084	-0.697	0.486
<i>Escherichia.coli</i> .O104.H4.str..2011C.3493.chromosome	-0.050	0.072	-0.696	0.487
<i>Escherichia.coli</i> .O83.H1.str..NRG.857C.chromosome	-0.053	0.078	-0.680	0.497
<i>Ruminococcus.bicirculans</i> .chromosome.II	0.046	0.068	0.670	0.503
<i>Enterococcus.faecium</i> .DO.plasmid.2	-0.101	0.152	-0.665	0.506
<i>Escherichia.coli</i> .O104.H4.str..2011C.3493.plasmid.pESBL.EA11	-0.058	0.090	-0.644	0.52
<i>Enterobacteria.phage</i> .lambda	-0.087	0.137	-0.640	0.522
<i>Clostridium.saccharolyticum</i> .WM1	-0.072	0.116	-0.624	0.532
<i>Acidaminococcus.fermentans</i> .DSM.20731	-0.083	0.133	-0.619	0.536
<i>Pseudoalteromonas.translucida</i> .KMM.520.chromosome.I	0.071	0.116	0.607	0.544
<i>Shigella.flexneri</i> .2a.str..301.plasmid.pCP301	-0.075	0.129	-0.580	0.562
<i>Klebsiella.oxytoca</i> .strain.CAV1374.plasmid.pCAV1374.150	-0.076	0.133	-0.573	0.567
<i>Streptococcus.mitis</i> .B6	-0.068	0.120	-0.572	0.567
<i>Clostridium.perfringens</i> .ATCC.13124	0.056	0.098	0.571	0.568
<i>Escherichia.coli</i> .O83.H1.str..NRG.857C.plasmid.pO83_CORR	-0.050	0.090	-0.557	0.577
<i>Tannerella.forsythia</i> .92A2	-0.047	0.088	-0.535	0.593
<i>Odoribacter.splanchnicus</i> .DSM.20712	0.039	0.078	0.497	0.619

	Estimate	Std. Error	z value	P_Value
Kluyvera.intermedia.strain.CAV1151.plasmid.pCAV1151.296	-0.057	0.118	-0.485	0.628
Escherichia.coli.str..K.12.substr..MG1655	-0.032	0.067	-0.480	0.631
Peptoclostridium.difficile.630	-0.048	0.101	-0.476	0.634
Pluralibacter.gergoviae.strain.FB2	-0.076	0.167	-0.455	0.649
Streptococcus.thermophilus.strain.S9	-0.055	0.125	-0.443	0.658
Clostridium.bolteae.90A9.genomic.scaffold.acBQZ.supercont	0.036	0.084	0.428	0.669
Flavonifractor.plautii.strain.YL31	-0.033	0.079	-0.414	0.679
Streptococcus.salivarius.strain.NCTC.8618	-0.032	0.080	-0.405	0.686
Lactococcus.lactis.subsp..lactis.II1403.chromosome	-0.056	0.141	-0.399	0.69
Desulfovibrio.fairfieldensis.strain.CCUG.45958	-0.040	0.106	-0.380	0.704
Bacteroides.fragilis.YCH46.DNA	0.028	0.074	0.377	0.706
Enterobacter.aerogenes.KCTC.2190.chromosome	-0.040	0.108	-0.365	0.715
Enterococcus.faecalis.V583.plasmid.pTEF1	-0.040	0.110	-0.364	0.716
Citrobacter.amalonaticus.strain.FDAARGOS_165	-0.040	0.112	-0.353	0.724
Enterobacteria.phage.mEp460	-0.046	0.132	-0.352	0.725
Parabacteroides.distasonis.ATCC.8503	0.024	0.075	0.328	0.743
Lactobacillus.salivarius.UCC118.chromosome	-0.032	0.097	-0.326	0.745
Fusobacterium.nucleatum.subsp..animalis.7_1	-0.057	0.177	-0.321	0.748
Parvimonas.micra.strain.KCOM.1535	-0.038	0.127	-0.301	0.764
Bacteroides.ovatus.strain.ATCC.8483	0.023	0.076	0.299	0.765
Bibersteinia.trehalosi.USDA.ARS.USMARC.192	-0.036	0.131	-0.272	0.786
Salmonella.phage.SJ46	0.027	0.103	0.264	0.792
Eubacterium.limosum.KIST612	0.027	0.104	0.260	0.795
Bifidobacterium.longum.NCC2705.chromosome	-0.016	0.064	-0.256	0.798
Klebsiella.pneumoniae.subsp..pneumoniae.HS11286.plasmid.pKPHS4	-0.028	0.112	-0.253	0.8
Klebsiella.oxytoca.strain.CAV1374.plasmid.pCAV1374.228	-0.031	0.123	-0.251	0.802
Bacteroides.thetaiotaomicron.strain.7330	-0.021	0.083	-0.249	0.803
Bacteroides.thetaiotaomicron.VPI.5482.plasmid.p5482	-0.019	0.079	-0.237	0.812
Streptococcus.pneumoniae.R6.chromosome	-0.032	0.139	-0.232	0.816
Bacteroides.thetaiotaomicron.VPI.5482.chromosome	-0.017	0.075	-0.231	0.817
Bacteroides.salanitronis.DSM.18170	-0.022	0.094	-0.228	0.819
Bacteroides.fragilis.YCH46.plasmid.pBFY46	0.034	0.155	0.220	0.826

	Estimate	Std. Error	z value	P_Value
Eubacterium.eligens.ATCC.27750.plasmid	0.018	0.083	0.217	0.828
Enterobacter.cloacae.subsp..cloacae.ATCC.13047.chromosome	-0.021	0.100	-0.212	0.832
Streptococcus.oligofermentans.AS.1.3089	-0.024	0.122	-0.200	0.841
Klebsiella.pneumoniae.subsp..pneumoniae.HS11286.chromosome	-0.014	0.073	-0.191	0.849
Roseburia.hominis.A2.183	-0.014	0.079	-0.183	0.855
Eubacterium.eligens.ATCC.27750	0.013	0.074	0.176	0.86
Escherichia.coli.O157.H7.str..Sakai	-0.013	0.080	-0.169	0.866
Streptococcus.sanguinis.SK36.chromosome	-0.020	0.121	-0.164	0.87
Alistipes.finegoldii.DSM.17242	-0.013	0.081	-0.158	0.874
Campylobacter.coli.RM4661.plasmid.pRM4661_48kbp	-0.019	0.122	-0.154	0.878
Streptococcus.thermophilus.JIM.8232.complete.genome	-0.015	0.103	-0.147	0.883
Eubacterium.sulci.ATCC.35585	0.020	0.139	0.140	0.888
Oscillibacter.valericigenes.Sjm18.20.DNA	-0.016	0.122	-0.135	0.893
Klebsiella.pneumoniae.subsp..pneumoniae.HS11286.plasmid.pKPHS2	-0.017	0.130	-0.127	0.899
Bacteroides.cellulosilyticus.strain.WH2	0.009	0.074	0.126	0.9
Bifidobacterium.dentium.JCM.1195...DSM.20436.DNA	-0.012	0.098	-0.125	0.9
Intestinimonas.butyriciproducens.strain.AF211	0.009	0.088	0.108	0.914
Clostridiales.genomosp..BVAB3.str..UPI9.5	-0.014	0.153	-0.089	0.929
Streptococcus.mutans.UA159.chromosome	0.010	0.123	0.082	0.935
Eggerthella.lenta.DSM.2243	0.006	0.082	0.071	0.943
Kluyvera.intermedia.strain.CAV1151	-0.008	0.122	-0.066	0.947
Escherichia.coli.UMN026.plasmid.p1ESCUM	-0.005	0.086	-0.062	0.951
Ruminococcus.bicirculans.chromosome.I	-0.005	0.075	-0.060	0.952
Lactobacillus.ruminis.ATCC.27782	0.007	0.129	0.054	0.957
Macroccoccus.caseolyticus.JCSC5402.plasmid.pMCCL2.DNA	-0.006	0.156	-0.039	0.969
Eubacterium.rectale.ATCC.33656	-0.003	0.074	-0.037	0.97
Pantoea.sp..PSNIH2.plasmid.pPSP.75c	-0.003	0.147	-0.020	0.984
Klebsiella.oxytoca.strain.CAV1374	0.002	0.104	0.018	0.986
Streptococcus.intermedius.B196	-0.002	0.131	-0.016	0.987
Bacteroides.salanitronis.DSM.18170.plasmid.pBACSA03	-0.001	0.095	-0.009	0.993
Bacteroides.vulgatus.ATCC.8482	0.001	0.067	0.008	0.994
Prevotella.melaninogenica.ATCC.25845.chromosome.II	0.001	0.168	0.004	0.997

dsdfdfs

Appendix III

	Estimate	Std. Error	t value	P_Value
<i>Lactobacillus.acidophilus</i> .30SC	3.119	0.786	3.969	0.000119
<i>Enterococcus.faecalis</i> .V583.chromosome	3.185	0.808	3.943	0.000131
<i>Acidaminococcus.intestini</i> .RyC.MR95	2.765	0.736	3.758	0.000258
<i>Streptococcus.galloyticus</i> .subsp..galloyticus.ATCC.43143.DNA	-2.226	0.786	-2.833	0.00535
<i>Enterococcus.faecium</i> .DO.chromosome	2.383	0.868	2.745	0.0069
<i>Lactobacillus.mucosae</i> .LM1	1.872	0.793	2.361	0.0197
<i>Shigella.dysenteriae</i> .Sd197.chromosome	-1.592	0.691	-2.302	0.0229
<i>Streptococcus.suis</i> .BM407.chromosome	1.951	0.874	2.233	0.0273
<i>Salmonella.enterica</i> .subsp..enterica.serovar.Typhi.str..CT18.plasmid.pHCM1	-1.986	0.968	-2.051	0.0423
<i>Streptococcus.gordonii</i> .str..Challis.substr..CH1	1.673	0.988	1.694	0.0926
<i>Enterobacteria.phage</i> .P1	-1.585	0.938	-1.690	0.0933
<i>Pseudoalteromonas.translucida</i> .KMM.520.chromosome.I	1.658	1.032	1.607	0.111
<i>Streptococcus.anginosus</i> .C238	1.331	0.962	1.384	0.169
<i>Adlercreutzia.equolifaciens</i> .DSM.19450.DNA	-0.930	0.736	-1.263	0.209
<i>Citrobacter.freundii</i> .CFNIH1.plasmid.pKEC.a3c	1.368	1.097	1.247	0.214
<i>Salmonella.enterica</i> .subsp..enterica.serovar.Typhi.str..CT18	-1.284	1.037	-1.238	0.218
<i>Porphyromonas.gingivalis</i> .ATCC.33277.DNA	1.716	1.408	1.219	0.225
<i>Klebsiella.pneumoniae</i> .subsp..pneumoniae.HS11286.plasmid.pKPHS4	1.185	1.017	1.166	0.246
<i>Clostridium.phytofermentans</i> .ISDg	-1.638	1.418	-1.155	0.25
<i>Clostridium.bolteae</i> .90A9.genomic.scaffold.acBQZ.supercont	0.800	0.722	1.109	0.27
<i>Bacteroides.ovatus</i> .strain.ATCC.8483	0.700	0.638	1.097	0.274
<i>Lactobacillus.salivarius</i> .UCC118.chromosome	-0.796	0.757	-1.051	0.295
<i>Enterobacteria.phage.lambda</i>	-1.204	1.208	-0.997	0.32
<i>Prevotella.enoea</i> .strain.F0113	1.202	1.216	0.988	0.325
<i>Streptococcus.intermedius</i> .B196	1.107	1.142	0.970	0.334
<i>Escherichia.coli</i> .O83.H1.str..NRG.857C.chromosome	-0.621	0.668	-0.929	0.355
<i>Escherichia.coli</i> .IAI39.chromosome	-0.613	0.661	-0.927	0.356
<i>Shigella.flexneri</i> .2a.str..301.plasmid.pCP301	-1.065	1.161	-0.917	0.361
<i>Klebsiella.oxytoca</i> .strain.CAV1374	0.829	0.908	0.913	0.363
<i>Enterobacteria.phage.mEp460</i>	-1.066	1.178	-0.905	0.367
<i>Pluralibacter.gergoviae</i> .strain.FB2	-1.220	1.356	-0.900	0.37

	Estimate	Std. Error	t value	P_Value
<i>Bacteroides. salanitronis</i> .DSM.18170.plasmid.pBACSA03	0.749	0.834	0.898	0.371
<i>Escherichia.coli</i> .O104.H4.str..2011C.3493.plasmid.pESBL.EA11	-0.685	0.784	-0.874	0.384
<i>Escherichia.coli</i> .O157.H7.str..Sakai.plasmid.pO157	-0.823	0.952	-0.865	0.389
<i>Bacteroides. cellulosilyticus</i> .strain.WH2	0.539	0.633	0.852	0.396
<i>Macrococcus. caseolyticus</i> .JCSC5402.plasmid.pMCCL2.DNA	1.145	1.362	0.841	0.402
<i>Shigella. flexneri</i> .2a.str..301.chromosome	-0.555	0.682	-0.814	0.417
<i>Enterococcus. faecium</i> .DO.plasmid.2	-1.072	1.371	-0.782	0.436
<i>Obesumbacterium. proteus</i> .strain.DSM.2777	-0.825	1.069	-0.772	0.441
<i>Bacteroides. fragilis</i> .YCH46.DNA	0.490	0.639	0.767	0.444
<i>Bifidobacterium. thermophilum</i> .RBL67	-0.874	1.142	-0.765	0.445
<i>Streptococcus. sanguinis</i> .SK36.chromosome	0.778	1.038	0.749	0.455
<i>Eubacterium. eligens</i> .ATCC.27750	0.478	0.642	0.745	0.458
<i>Eubacterium. rectale</i> .ATCC.33656	0.483	0.654	0.739	0.461
<i>Veillonella. parvula</i> .DSM.2008	-0.542	0.743	-0.730	0.467
<i>Bacteroides. vulgatus</i> .ATCC.8482	0.407	0.583	0.699	0.486
<i>Parabacteroides. distasonis</i> .ATCC.8503	0.429	0.638	0.673	0.502
<i>Campylobacter.coli</i> .RM4661.plasmid.pRM4661_48kbp	0.700	1.046	0.669	0.505
<i>Eubacterium. eligens</i> .ATCC.27750.plasmid	0.468	0.713	0.656	0.513
<i>Salmonella. enterica</i> .subsp..enterica.serovar.Typhimurium.str..LT2	-1.016	1.586	-0.641	0.523
<i>Enterococcus. faecalis</i> .V583.plasmid.pTEF1	0.612	0.966	0.634	0.527
<i>Oscillibacter. valericigenes</i> .Sjm18.20.DNA	-0.656	1.046	-0.627	0.532
<i>Pantoea. sp.</i> .PSNIH2.plasmid.pPSP.75c	-0.772	1.263	-0.611	0.542
<i>Escherichia.coli</i> .O104.H4.str..2011C.3493.plasmid.pAA.EA11	-0.579	0.957	-0.605	0.546
<i>Staphylococcus. haemolyticus</i> .JCSC1435.DNA	0.586	0.997	0.587	0.558
<i>Bacteroides. thetaiotaomicron</i> .VPI.5482.chromosome	0.369	0.646	0.571	0.569
<i>Bacteroides. thetaiotaomicron</i> .strain.7330	0.404	0.712	0.567	0.571
<i>Klebsiella. oxytoca</i> .strain.CAV1374.plasmid.pCAV1374.228	0.604	1.069	0.565	0.573
<i>Lactobacillus. ruminis</i> .ATCC.27782	-0.617	1.098	-0.562	0.575
<i>Escherichia.coli</i> .UMN026	0.364	0.652	0.558	0.577
<i>Shigella. phage</i> .StfV	-0.629	1.142	-0.550	0.583
<i>Streptococcus. thermophilus</i> .JIM.8232.complete.genome	0.471	0.891	0.529	0.598
<i>Clostridiales. genomosp.</i> .BVAB3.str..UPII9.5	-0.685	1.323	-0.518	0.606

	Estimate	Std. Error	t value	P_Value
Haemophilus.influenzae.Rd.KW20.chromosome	-0.586	1.136	-0.516	0.607
Parvimonas.micra.strain.KCOM.1535	0.587	1.143	0.514	0.608
Escherichia.coli.O104.H4.str..2011C.3493.chromosome	-0.316	0.636	-0.498	0.62
Citrobacter.freundii.CFNIH1	-0.454	0.929	-0.489	0.626
Aggregatibacter.aphrophilus.strain.W10433	-0.547	1.148	-0.477	0.634
Clostridium.perfringens.ATCC.13124	0.405	0.858	0.471	0.638
Peptoclostridium.difficile.630	0.417	0.893	0.468	0.641
Streptococcus.oligofermentans.AS.1.3089	0.487	1.067	0.457	0.649
Fusobacterium.nucleatum.subsp..animalis.7_1	0.662	1.448	0.457	0.649
Intestinimonas.butyriciproducens.strain.AF211	0.348	0.763	0.456	0.649
Bifidobacterium.breve.DSM.20213....JCM.1192.DNA	0.293	0.655	0.448	0.655
Streptococcus.mitis.B6	0.463	1.047	0.443	0.659
Escherichia.coli.O157.H7.str..Sakai	-0.306	0.697	-0.439	0.662
Citrobacter.amalonaticus.strain.FDAARGOS_165	0.435	1.005	0.433	0.666
Eubacterium.limosum.KIST612	0.384	0.896	0.429	0.669
Bacteroides.salanitronis.DSM.18170	0.307	0.816	0.376	0.708
Desulfovibrio.fairfieldensis.strain.CCUG.45958	-0.335	0.909	-0.369	0.713
Eubacterium.sulci.ATCC.35585	0.429	1.196	0.359	0.72
Klebsiella.pneumoniae.subsp..pneumoniae.HS11286.plasmid.pKPHS2	0.399	1.116	0.358	0.721
Odoribacter.splanchnicus.DSM.20712	0.233	0.670	0.348	0.728
Klebsiella.pneumoniae.subsp..pneumoniae.HS11286.plasmid.pKPHS3	0.363	1.047	0.347	0.729
Klebsiella.oxytoca.strain.CAV1374.plasmid.pCAV1374.150	0.365	1.119	0.326	0.745
Streptococcus.salivarius.strain.NCTC.8618	0.227	0.703	0.324	0.747
Enterobacter.aerogenes.KCTC.2190.chromosome	0.274	0.888	0.309	0.758
Klebsiella.pneumoniae.subsp..pneumoniae.HS11286.chromosome	0.184	0.640	0.287	0.774
Salmonella.phage.SJ46	-0.258	0.905	-0.286	0.776
Escherichia.coli.str..K.12.substr..MG1655	-0.157	0.596	-0.264	0.793
Cronobacter.sakazakii.strain.ATCC.29544	-0.236	0.927	-0.254	0.8
Bifidobacterium.longum.NCC2705.chromosome	-0.138	0.564	-0.244	0.807
Bifidobacterium.dentium.JCM.1195...DSM.20436.DNA	-0.189	0.842	-0.225	0.822
Roseburia.hominis.A2.183	0.140	0.693	0.203	0.84
Bifidobacterium.bifidum.PRL2010.chromosome	0.137	0.716	0.191	0.849

	Estimate	Std. Error	t value	P_Value
<i>Bibersteinia trehalosi</i> .USDA.ARS.USMARC.192	0.224	1.171	0.191	0.849
<i>Bacteroides thetaiotaomicron</i> .VPL5482.plasmid.p5482	0.117	0.656	0.178	0.859
<i>Enterobacter cloacae</i> .subsp..cloacae.ATCC.13047.chromosome	0.161	0.903	0.178	0.859
<i>Escherichia coli</i> .UMN026.plasmid.p1ESCUM	0.138	0.779	0.176	0.86
<i>Rothia mucilaginosa</i> .DY.18.DNA	-0.191	1.145	-0.167	0.868
<i>Eggerthella lenta</i> .DSM.2243	0.119	0.718	0.166	0.868
<i>Lactococcus lactis</i> .subsp..lactis.II1403.chromosome	-0.201	1.247	-0.161	0.872
<i>Streptococcus mutans</i> .UA159.chromosome	0.158	1.054	0.149	0.881
Lachnospiraceae.bacterium.2_1_46FAA.genomic.scaffold.acIXT.supercont	-0.118	0.823	-0.143	0.887
<i>Enterobacter cloacae</i> .subsp..cloacae.ATCC.13047.plasmid.pECL_A	-0.179	1.254	-0.142	0.887
<i>Escherichia coli</i> .O83.H1.str..NRG.857C.plasmid.pO83_CORR	0.115	0.807	0.142	0.887
<i>Yersinia pestis</i> .CO92.chromosome	-0.191	1.351	-0.141	0.888
<i>Streptococcus parasanguinis</i> .ATCC.15912	-0.109	0.815	-0.134	0.893
<i>Clostridium saccharolyticum</i> .WM1	0.134	1.020	0.132	0.895
<i>Streptococcus thermophilus</i> .strain.S9	0.138	1.056	0.131	0.896
<i>Streptococcus pneumoniae</i> .R6.chromosome	-0.161	1.232	-0.131	0.896
<i>Barnesiella viscericola</i> .DSM.18177	-0.106	0.811	-0.130	0.896
<i>Tannerella forsythia</i> .92A2	0.096	0.738	0.129	0.897
<i>Flavonifractor plautii</i> .strain.YL31	-0.084	0.678	-0.124	0.901
<i>Filifactor alocis</i> .ATCC.35896	-0.150	1.296	-0.116	0.908
<i>Alistipes finegoldii</i> .DSM.17242	-0.078	0.703	-0.112	0.911
<i>Kluyvera intermedia</i> .strain.CAV1151	-0.107	1.079	-0.099	0.921
<i>Kluyvera intermedia</i> .strain.CAV1151.plasmid.pCAV1151.215	0.095	1.117	0.085	0.932
<i>Bacteroides fragilis</i> .YCH46.plasmid.pBFY46	-0.087	1.267	-0.069	0.945
<i>Acidaminococcus fermentans</i> .DSM.20731	-0.065	1.128	-0.058	0.954
<i>Kluyvera intermedia</i> .strain.CAV1151.plasmid.pCAV1151.296	0.037	1.053	0.035	0.972
<i>Bifidobacterium adolescentis</i> .ATCC.15703.DNA	-0.021	0.604	-0.034	0.973
<i>Prevotella melaninogenica</i> .ATCC.25845.chromosome.II	-0.039	1.366	-0.029	0.977
<i>Clostridium</i> .sp..SY8519.DNA	0.021	0.910	0.023	0.982
<i>Ruminococcus bicirculans</i> .chromosome.I	-0.012	0.657	-0.018	0.985
<i>Akkermansia muciniphila</i> .ATCC.BAA.835	-0.005	0.601	-0.008	0.994
<i>Ruminococcus bicirculans</i> .chromosome.II	0.004	0.595	0.006	0.995