Predicting Schizophrenia from Human Gut Microbiome

Dong Wang[1,2], Will Russel[3], Ahmet Ay[2,3]

[1]Department of Computer Science, [2]Department of Mathematics, [3]Department of Biology, Colgate University, Hamilton, NY 13346

Background

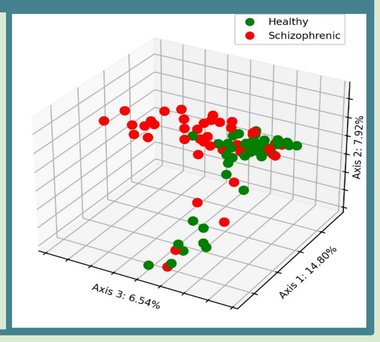
- Recent findings have supported an association between deviations in gut microbiome composition and mental disorders such as schizophrenia.
- This is related to the hypothesis that the microbiota affects brain activities through the gut-brain axis, the bidirectional interaction between the central nervous system and digestive system.
- We examined the dataset collected by Nguyen et al. 2021 [1] to investigate the differences in gut microbial composition between 48 schizophrenia (**SZ**) subjects and 48 matched healthy (**HC**) subjects.
- Nguyen et al. already established that beta diversity is different between healthy and schizophrenia groups
- First, we investigated clinical factors that also had an impact on gut microbiome, and how they correlate with differences observed due to schizophrenia diagnosis.
- Then, we focused on identifying specific taxa that may be critically associated with the difference in microbiome caused by schizophrenia.

Methods

- Using the microbiome analysis tool QIIME 2 [2], we compared beta diversity between HC and SZ groups, and also in subgroups divided by clinical factors
- We conducted mediation analysis [3] to investigate which factors create a mediation effect
- ANCOM [4] was used to identify differentially abundant taxa
- Combinations of feature selection methods and classifiers were used to develop a machine learning model to predict schizophrenia from gut microbiome
- Association networks [5] (bootstrap n=100) of microbial genera were constructed and analyzed

Clinical Factor Results

Figure 1. Beta diversity of microbial taxa was significantly different between SZ and HC groups. Principal coordinates analysis (PCoA) plot with Bray-Curtis distance was used to visualize the clear clustering of microbes. **Green**: HC; **Red**: SZ



Subgroup	Bray-Curtis	Unweight UniFrac
Race		
Caucasian: healthy (n=38) vs schizophrenia (n=31)	0.003	0.002
Other: healthy (n=10) vs. schizophrenia (n=17)	0.187	0.462
Antibiotic Use		
Non-users: healthy (n=36) vs. schizophrenia (n=35)	0.008	0.004
Users: healthy (n=12) vs. schizophrenia (n=13)	0.172	0.237
Smoking		
Not smoker: healthy (n=46) vs. schizophrenia (n=20)	0.027	0.003
Smoker: healthy (n=2) vs. schizophrenia (n=28)	0.727	0.99
Depression		
Non-depression: healthy (n=31) vs. schizophrenia (n=16)	0.019	0.031
Depression: healthy (n=10) vs. schizophrenia (n=28)	0.258	0.754

Table 1. Beta diversity Comparisons of subgroups divided by clinical features illustrate their influence on microbial diversity (results are p-values). Significance between HC and SZ groups was retained in non-smoker, non-antibiotic, non-depressed, non-Caucasian groups (color coded red). However, it was not seen in smoker, antibiotic, depressed, Caucasian groups (color coded blue).

Variable	Bray-Curtis	Jaccard	G-UniFrac	Permutation
Age (n=95)	0.911	0.940	0.395	0.667
Sex (n=96)	0.785	0.051†	0.194	0.131
Body-mass index (n=95)	0.097†	0.049*	0.066†	0.122
Race (n=96)	0.623	0.850	0.202	0.408
Anxiety score (n=85)	0.044*	0.022*	0.083†	0.060†
Depression score (n=85)	0.044*	0.042*	0.028*	0.077†
Antibiotic (n=96)	0.165	0.239	0.331	0.360
Smoking status (n=96)	0.012*	0.057†	0.061†	0.034*

Table 2. Mediation Analysis (results are p-values). Smoking status, depression score, anxiety score and BMI likely created mediation effects.

Important Taxa Results

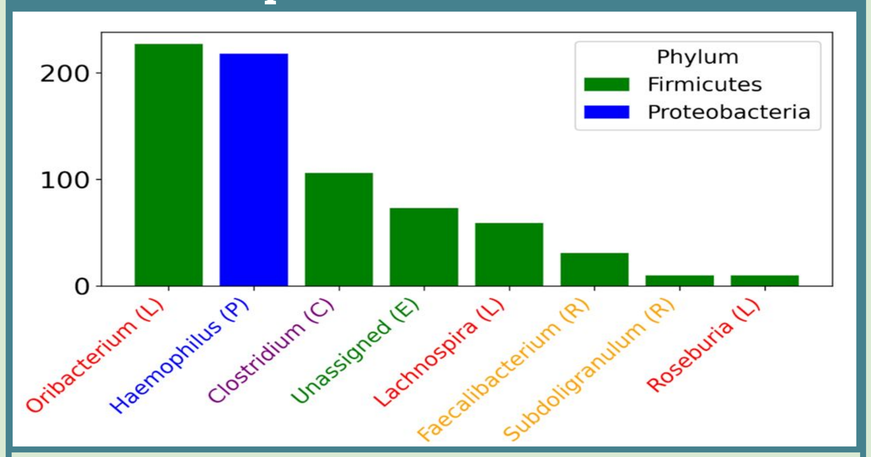


Figure 2. Top 8 differentially abundant genera between HC and SZ groups. Y-axis is measure of difference, bars are color coded by phylum, and genera names are color coded by family. We observed 3 genera from *Lachnospiraceae* family and 2 genera from *Ruminococcaceae* family.

Machine Learning Outcome

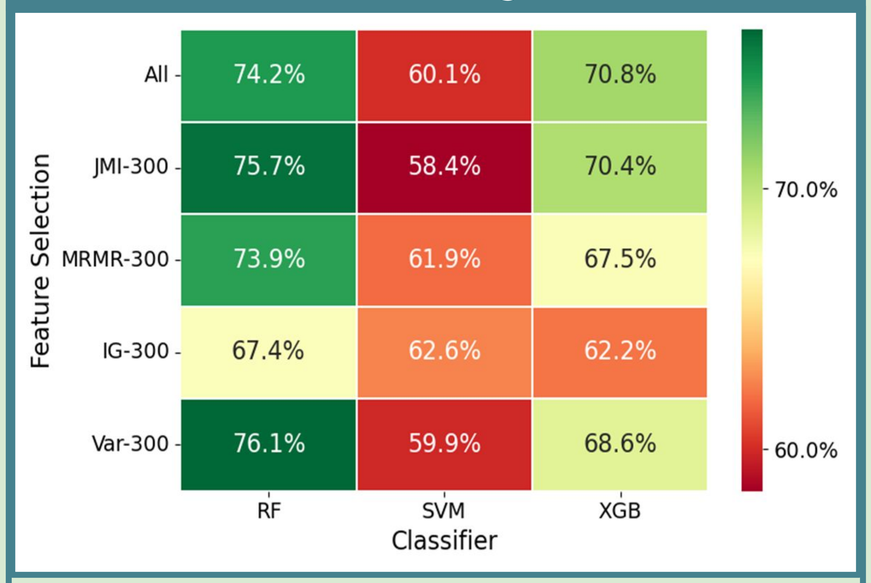


Figure 3. Heatmap of Machine Learning Prediction Accuracy. We chose 300 ASVs with 4 feature selection techniques and applied to 3 classifiers. Nguyen et al. (2021) reported accuracy of 75% from all ASVs with random forest. We achieved slightly higher 76.1% using random forest with a much smaller subset of 300 features selected based on variance in relative abundance.

Network Analysis

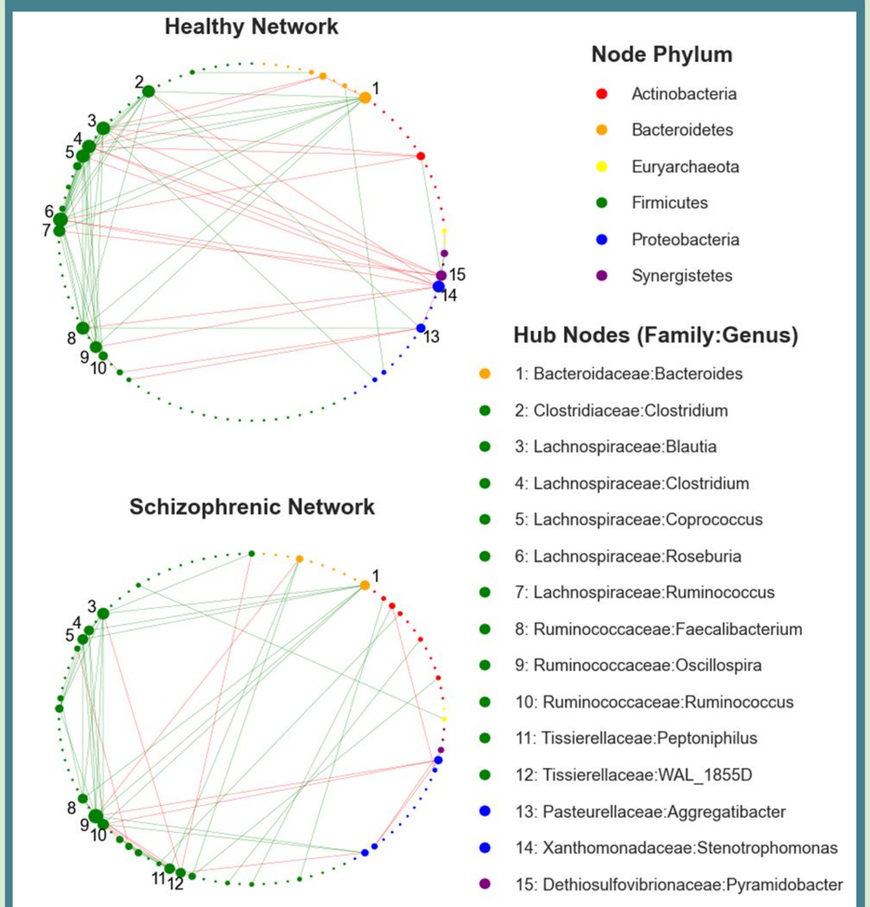


Figure 4. Network Visualization. Green edges represent positive associations. Red edges represent negative associations. Nodes with most edges are labelled as hub nodes.

Network Properties

Metric	Healthy	Schizophrenia	Healthy (mean)	Schizophrenia (mean)	P-Value
Nodes	100	100	98.370	101.590	<0.001
Edges	75	59	88.080	60.990	<0.001
Positive Edges	54	45	69.300	46.200	<0.001
Negative Edges	21	14	18.780	14.790	<0.001
Modularity	0.394	0.688	0.627	0.896	<0.001
Family Assortativity	0.002	0.135	0.048	0.113	<0.001
Shared Cliques	22	22	15.6	15.6	N/A
Shared Positive Cliques	22	22	15.2	15.2	N/A

Table 3. Network Properties. The HC network has higher density (more edges), while the SZ network has higher modularity and family assortativity. All shared cliques are positive. To prove network structures are statistically different, we conducted bootstrapping (n=100) and created 100 networks for each group. T-tests indicate that all properties are significantly different.

Discussion & Conclusions

- Subgroup beta diversity comparisons showed that smoking, antibiotic use, depression and cultural behavior (race), which are factors linked to microbial differences, will influence microbiome composition. The differences caused by these factors may mitigate the differences caused by schizophrenia, which will impact diversity results
- In particular, the links between anxiety, depression, smoking and schizophrenia are mediated through gut microbiome
- Taxa from families *Lachnospiraceae* and *Ruminococcaceae* were found to be differentially abundant between HC and SZ groups, supported by both differential abundance testing and feature selection
- Machine learning models can be developed to accurately predict schizophrenia from a small subset of microbial ASVs observed, making the computation faster and biological interpretation easier
- HC networks tend to have more edges, which shows that microbial interactions are more abundant in healthy subjects
- Interactions in SZ networks tend to be more modular and involve nodes close to each other taxonomically
- Negative associations involving nodes from *Lachnospiraceae* and *Ruminococcaceae* families to nodes in other phyla exist in the HC network, but disappear in the SZ network, which suggest that inter-phyla competitive interactions were reduced in SZ networks.

Limitations & Future Work

- Experiments in future microbiome studies should be careful to control for other clinical features
- The smaller sample size limits the generalizationality of these findings. More sizable datasets need to be analyzed to produce consistent results and conclusions
- The specificity of microbial association findings is relatively novel, and network analysis of other microbiome datasets is needed to provide more evidence of those particular negative associations being reduced in SZ networks

Literature Cited

[1]. Nguyen et al., Gut microbiome in schizophrenia: Altered functional pathways related to immune modulation and atherosclerotic risk, 2021.
[2]. Bolyen et al., Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2, 2019.
[3]. Jie Zhang, Zhi Wei, Jun Chen. A distance-based approach for testing the mediation effect of the human microbiome, 2018.
[4]. Mandal et al., Analysis of composition of microbiomes: a novel method for studying microbial composition, 2015.
[5]. Loftus et al., Bacterial associations in the healthy human gut microbiome across populations, 2021.

Acknowledgements

We would like to thank Yuntong Sun '22 for her contribution to this project during the past semester, and Professor Ken Belanger for his expertise and insightful advice.