Data-driven exploration of cross-body site interactions between oral and gut microbiome

A Project Report

submitted by

Sakshi Mahajan

Under the supervision of
Dr. Tarini Shankar Ghosh

Submitted in partial fulfillment of the requirements for the degree of Master of Technology, Computational Biology

Center for Computational Biology,
Indraprastha Institute of Information Technology - Delhi
May, 2024

Certificate

This is to certify that the thesis titled *"**Data-driven exploration of cross-body site interactions between oral and gut microbiome**"* being submitted by **Sakshi Mahajan** to the Indraprastha Institute of Information Technology, Delhi, for the award of the **Master of Technology**, is an original research work carried out by her under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree/diploma.

Dr. Tarini Shankar Ghosh
Thesis Supervisor
Assistant Professor
Department of Computational Biology
Indraprastha Institute of Information Technology,
Delhi, New Delhi-110 020

Place :-New Delhi

Date:-20th May 2024

Acknowledgments

I would like to express my sincere gratitude to those who have played an instrumental role in the successful completion of my thesis. First and foremost, I am deeply grateful to Dr. Tarini Shankar Ghosh for his exceptional guidance, unwavering support, mentorship, and belief in me throughout the entire project. His invaluable insights, expert advice, support, and constructive criticism have been pivotal in shaping the direction of my research and enhancing the quality of my work. I am truly fortunate to have had the opportunity to work under his mentorship. From the beginning of my thesis, he continuously challenged me to reach higher goals and excel in my work. He generously offered his time to address any issues or clarify my doubts. Whenever I encountered challenges in my work, he provided valuable insights and refined our approach, significantly contributing to the timely completion of my tasks. I am truly grateful for his constant encouragement and support, which have been instrumental in my academic and personal growth.

I would like to express my gratitude to my friend and project partner, Omprakash, for his constant support, collaboration, and motivation throughout the duration of this project. His dedication, enthusiasm, and willingness to assist me during all the highs and lows have been truly remarkable. Working alongside him has not only made the project more enjoyable but also immensely rewarding.

I also thank my lab mates, Abhishek, Lavanya, Vindhya, and Shivangi, for supporting me throughout the work. Their presence and encouragement were invaluable.

I extend my heartfelt gratitude to all the faculty members and staff of the Department of Computational Biology at IIIT Delhi for their constant support and assistance throughout my journey as an M Tech student.

I would also like to thank my parents and batch mates for their constant support and cheer to make this journey smooth.

To all those mentioned above and many others who have been a part of my academic and personal growth, I offer my heartfelt thanks. Your contributions have been invaluable, and I am truly fortunate to have had such a supportive network of individuals around me.

Abstract

This study investigates the link between the oral and gut microbiome, two highly diverse microbial communities with established connections to human health. We performed meta- analysis on across those studies containing cross-body-site sample for oral and gut. We analyzed those paired oral-gut microbiome data from over 1600 individuals across 12 global cohorts. Our findings reveal a strong correlation between the composition of these microbiomes. Interestingly, we identified distinct groups of oral bacteria that appear to selectively influence either beneficial or detrimental members of the gut microbiota.In this study, we present the results and investigate the inter-relatedness, transmission pattern, connecting links and reproducibly predicting microbes.We identified extent of oral microbial species in the gut environment. These findings can potentially help to advance the understanding of interaction patterns between cross-body sites. This study suggests the potential for the oral microbiome to be utilized as part of targeted interventions aimed at reducing the risk and progression of gut-related diseases.

Keywords:- oral microbiome, gut microbiome, human microbiota, cross-body-site, microbiome translocation

Contents

List of Figures

Chapter 1

Introduction

Thousands of different microbiomes colonize at oral and gut environments, both are the most essential parts of our digestive tract. These active communities and their functional impact affect human health. The comprehensive analysis conducted by Segata et al.[1] not only reveals the complex microbial ecosystems within the human digestive tract but also highlights the frequent influence of oral bacteria colonies at different subsites in the oral and how they are associated with metabolic pathways across the gastrointestinal tract, illustrating the interconnection between local oral health and systemic well-being. The oral and gut microbiomes play crucial roles in maintaining human health, with each microbiome consisting of a complex community of microorganisms that inhabit different parts of the body. A variety of systemic diseases have been associated with the abundance of oral bacteria throughout the body[2]. Although various biotic and abiotic factors influence human microbial community composition [3].

Oral-to-gut microbial transmission is common, with saliva serving as a direct transmission vehicle, leading to higher strain-sharing rates in cohabiting individuals compared to non-cohabiting individuals[4]. Extensive cross-body site microbial transmission is common among healthy individuals, with increased levels in colorectal cancer and rheumatoid arthritis patients, shaping the gastrointestinal microbiome in health and disease [5].

Oral microbiota plays a crucial role in the onset and progression of various diseases, both localized and systemic. Dysbiosis in the oral cavity, characterized by a disturbance in the microbial composition, has been linked to conditions such as Alzheimer's disease (AD) [6], liver diseases through the oral-liver-gut axis, and even systemic infections like COVID-19 [7]. The human oral microbiome (HOM) is the second-largest microbial community in the body and can impact health significantly. Dysbiosis in the

oral microbiome can lead to metabolic imbalances, inflammation, and other pathological changes that contribute to disease states. Understanding the intricate interactions within the oral microbial community is essential for developing diagnostic, therapeutic, and preventive strategies to manage various health conditions associated with oral dysbiosis.The oral microbiome plays a crucial role in children's health, with implications for systemic diseases and developmental outcomes. Children's oral microbiota, rich in diverse microorganisms, can serve as an early indicator of systemic health issues and Understanding the oral microbiome in children is essential for promoting overall health and well-being.[8].

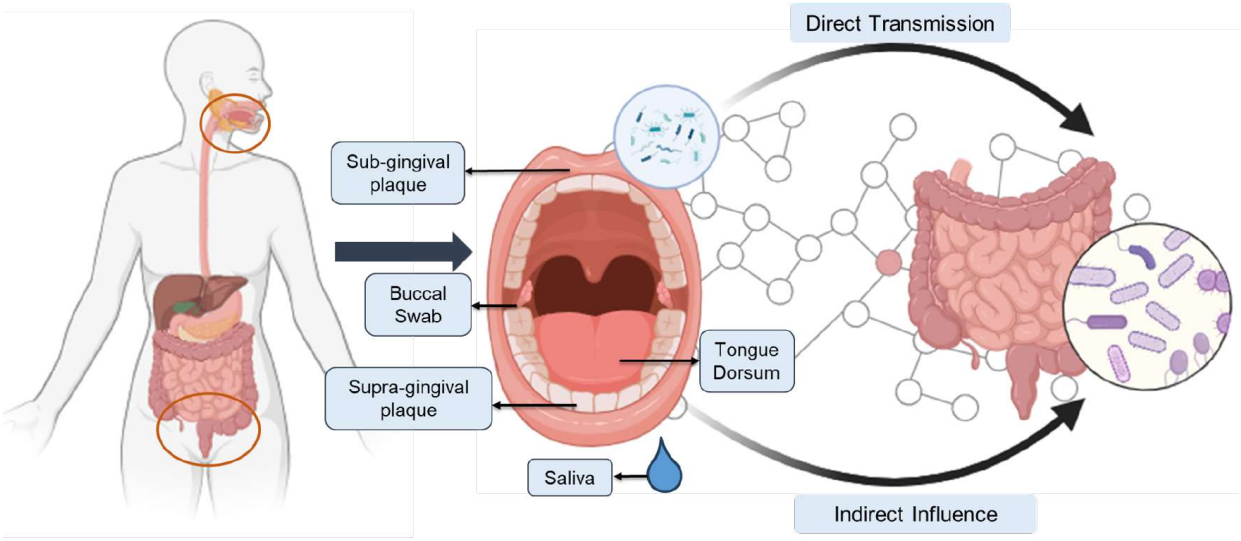


Figure 1.1: Two main ways by which the oral bacteria can affect the microbiome of the gut(oral to gut axis) & oral subsites representation

Further research could involve developing targeted interventions to prevent the transmission of disease-associated microbes and promote beneficial strains. This com-

prehensive approach will contribute to a deeper understanding of the oral-gut microbial axis and its implications for health and disease management. The paper delves into microbial translocation disorders induced by the host immune responses to extraintestinal gut microbes and their constituent parts, although it does not specifically address the oral-gut translocation of periodontal pathogens.

The concept of "cross-body site interactions" refers to the interplay between microbiomes located in different parts of the body, such as the mouth and the gut. Recent research has begun to uncover how these microbial communities influence each other and contribute to disease states or health conditions through complex inter-site interactions. The gut microbiota influences disease development through various mechanisms, including immune system modulation, metabolite production, and microbial interactions. Understanding these connections can lead to novel therapeutic strategies targeting the gut microbiome to manage and potentially prevent a wide range of diseases. Thus, based on prior research, we understand the importance of both communities and their relatedness.

In this study, we present the results and investigate the inter-relatedness, transmission pattern, connecting links and reproducibly predicting microbes in this two communities.

Chapter 2

MATERIALS AND METHODS

2.1 Data Collection and Preprocessing

To investigate the connections between the microbes in our oral and guts, we first needed to collect specific microbiome datasets from study cohorts that includes Whole Genome Sequencing (WGS) or 16S Amplicon Sequencing (16S) data from matched oral and gut (or fecal) samples taken from the same individuals. We identified 12 such study cohorts from various global locations. These cohorts provided a comprehensive collection of 3,864 microbiome samples, including 1,805 stool samples representing the gut microbiome and 1,825 oral samples representing the oral microbiome. Importantly, for our analysis, we focused on 1673 samples where both oral and gut samples came from the same person (matched samples).

Sequenced data and metadata of study cohorts collected from European Nucleotide Archive by accessing accession numbers. compressed reads location we get in fastq.gz format performed preprocessing and quality filtering using meta-analysis tool. To ensure consistent and reliable analysis across all the datasets in our study, we implemented a standardized taxonomic annotation approach. This means assigning labels to the different types of microbes present in the samples. We recognized that the data originated from two different sequencing methods: Whole Genome Shotgun Sequencing (WGS) and 16S ribosomal RNA gene sequencing (16S).For datasets derived from 16S rRNA sequencing, we used the SPINGO taxonomic profiler[9]. This approach allowed us to take advantage of the strengths of each method. SPINGO is particularly effective at providing high-resolution, species-level assignments from partial 16S sequences, enabling us to capture detailed insights into microbial communities. On the other hand, MetaPhlAn3 is designed for accurate species-level profiling of bacteria, archaea, and even eukaryotes from whole-genome shotgun sequencing data[10]. By using SPINGO

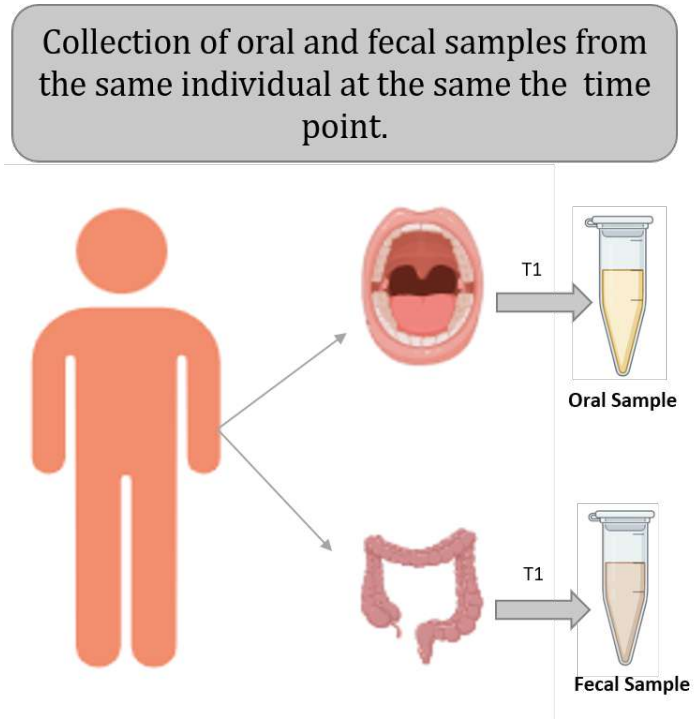
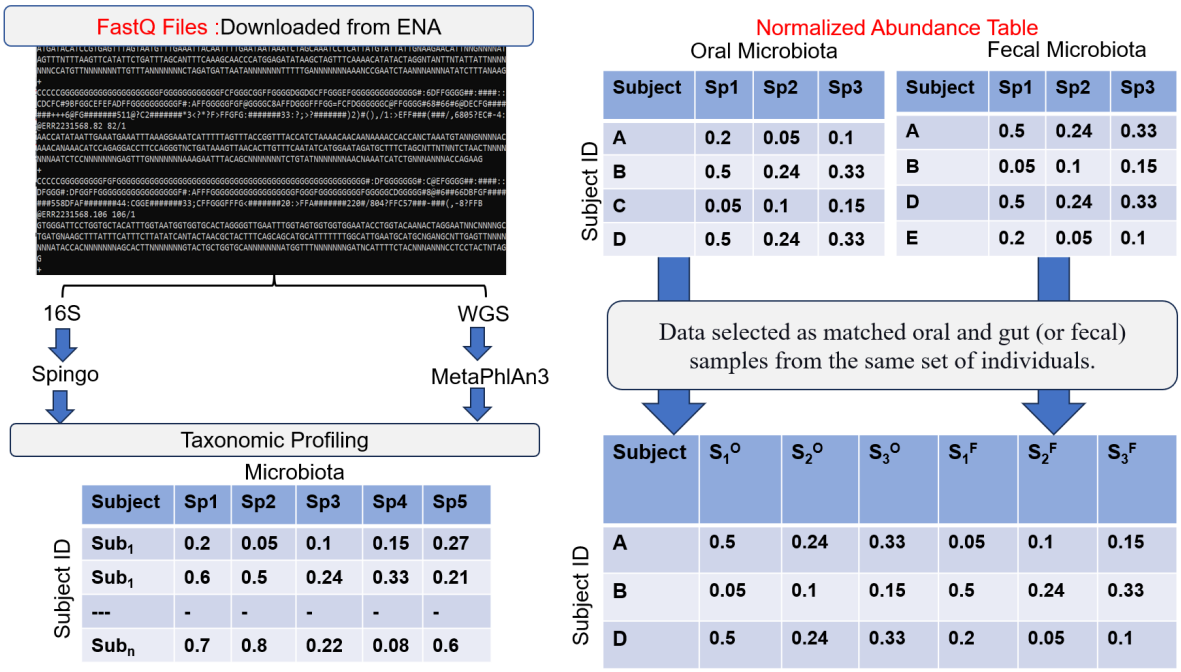


Figure 2.1: Schematic Representation of Simultaneous Oral and Gut Sample Collection at the Same Time Point from the Same set of Individual

for 16S data and MetaPhlAn3 for WGS data, we ensured that each type of sequencing data was analyzed using the most suitable method. After this step we get species level abundance profile for all studies separately which further we can use for analysis. We found a diverse range of oral sub-sites covered in the studies, but some sub-sites had very few samples. To ensure robustness in our analysis, we decided to focus on five primary oral sites: saliva, tongue dorsum, buccal swab, supragingival plaque, subgingival plaque, and stool. This selection provided a more substantial dataset to work with. To

further refine our dataset, we removed any duplicate samples from the lesser-studied oral sub-sites, as well as any stool or oral samples that did not have a matching counter pair. This curation process resulted in a final dataset consisting of 3,346 samples, which included 1,673 pairs of fecal and oral samples collected from the same individuals. This approach ensured that our analysis was based on well-matched and high-quality data, allowing for a more accurate investigation of the microbial communities in the oral and gut microbiomes. refer figure 2.2



Subject	Sp1	Sp2	Sp3	Sp4	Sp5
Sub_1	0.2	0.05	0.1	0.15	0.27
Sub_1	0.6	0.5	0.24	0.33	0.21
---	-	-	-	-	-
Sub_n	0.7	0.8	0.22	0.08	0.6

Subject ID

Subject	Sp1	Sp2	Sp3
A	0.2	0.05	0.1
B	0.5	0.24	0.33
C	0.05	0.1	0.15
D	0.5	0.24	0.33

Subject ID

Subject	Sp1	Sp2	Sp3
A	0.5	0.24	0.33
B	0.05	0.1	0.15
D	0.5	0.24	0.33
E	0.2	0.05	0.1

Subject	S_1^O	S_2^O	S_3^O	S_1^F	S_2^F	S_3^F
A	0.5	0.24	0.33	0.05	0.1	0.15
B	0.05	0.1	0.15	0.5	0.24	0.33
D	0.5	0.24	0.33	0.2	0.05	0.1

Subject ID

Figure 2.2: Data Preprocessing Pipeline for Matched Oral and Gut Samples

Sr. no	Study Name	Accession Number	Total Samples	Matched Samples (in pair)	Study condition		Diseased Type	Country	Body site	Sequence Type
					Control Samples	Diseased Samples				
1	BritoIL_2016	PRJNA217052	312	140	280	0	-	FJI	Saliva	WGS
2	FerrettiP_2018	PRJNA352475	215	54	108	0	-	ITA	Tongue dorsum	WGS
3	HMP_2012		749	352	489	0	-	USA	Buccal Swab, Tongue dorsum, Supragingival plaque	WGS
4	NagataN_2022	PRJNA832909	560	276	470	86	PDAC	JPN	Saliva	WGS/16s
5	Russo_2018	PRJNA356414	50	20	40	0	-	Italy	Saliva	16s
6	Stewart_2018	PRJNA413706	90	60	90	0	-	USA	Saliva, Buccal Swab	16s
7	Chaudhari_2020	PRJNA438728 PRJNA399246	100	37	74	0	-	India	Saliva	16s
8	Lindheim_2016	PRJNA326866	86	43	86	0	-	Austria, Europe	Saliva	16s
9	Pool_2019	PRJEB27308	527	206	412	0	-	USA	Saliva	16s
10	LokmerA_2020	PRJEB30836	250	105	212	0	-	Africa	Saliva	16s
11	KunathB_2022	PRJNA289586	218	70	0	140	T1DM	Europe	Saliva	WGS
12	Shoer_2023	PRJEB64861	706	310	620	0	-	Israel	Subgingival plaque	WGS

Figure 2.3: Detailed overview of Dataset Used in Analysis

2.2 Methods

2.2.1 Investigating the overall cross-body-site relatedness between the oral and fecal (gut) microbial communities within the same individuals across the different studies

We divided this objective into two parts, first was to profile the cross-community variations across the different individuals, within the same body sites (i.e. either oral or gut/fecal).To perform this, we used Principal Coordinate Analysis (PCoA)[11], a method that helps to convert higher dimension data into 2-Dimensional or 3-Dimensional to reduces the complexity of data by focusing on the most important feature. PCoA uses information about how different microbiome communities compare to each other

(beta-diversity) to create new variables called Principal Coordinates (PCs), which capture the most significant differences between communities. Beta diversity indicates the how divers are the two communities. We started by creating distance matrices for each body site and study cohort, showing how different the microbial communities are from each other using the Kendall distance.

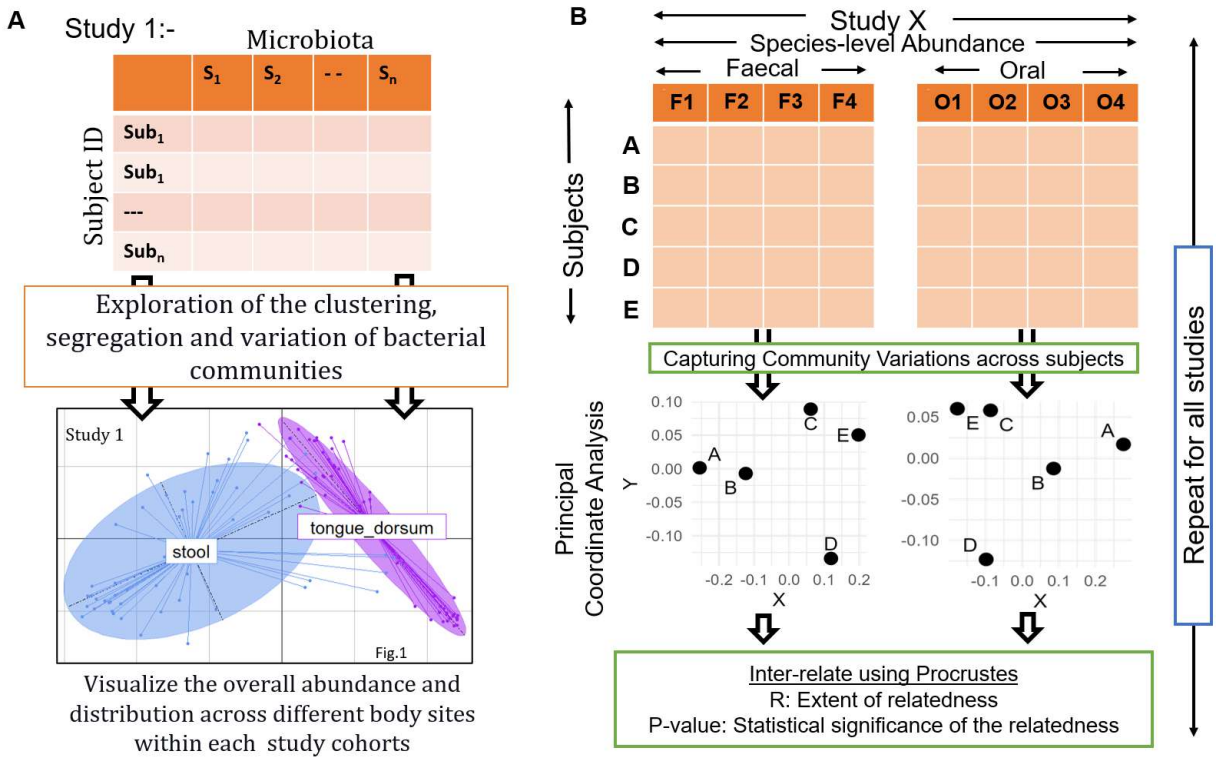


Figure 2.4: Methodology for Profiling Cross-Community Variations Across Different Individuals Within the Same Body Sites (Oral or Gut/Fecal) (A) and Correspondence Between Oral and Gut Microbiome Compositions Within the Same Individual Across Different Studies Using Procrustes Analysis (B)

We then applied PCoA to these distance matrices using the dudi.pco function from the vegan R package (version 2.6.4), which computes the PCs. These PCs provide a

simplified representation of the microbial communities, highlighting the main differences between them. We visualized the most informative PCs in two-dimensional or three-dimensional plots, making it easier to see how different microbial communities cluster together, separate from each other, and vary. This visualization allowed us to explore patterns such as similarities, differences, and groupings of bacterial communities. We performed separate PCoAs for each study cohort to visualize the overall abundance and distribution of microbial communities within each body site, examining the oral microbiome separately from the gut microbiome for each study group. PCoA helped us turn complex microbiome data into clear visual representations, enabling us to identify patterns and differences in bacterial communities across individuals for each study cohorts.

The next step aimed to see if variations in the microbial communities between the oral and gut were related and how variations in microbial communities in one body site (oral) relate to variations in another body site (fecal/gut).If the composition of the microbial community in the oral cavity influences that in the gut, then the variations in these communities should be correlated. For instance, if two individuals have similar oral microbial communities (similar oral Principal Coordinates, or PCs), they are likely to have similar gut microbial communities (similar gut PCs), and vice versa. To assess this relationship, we used Procrustes analysis, which compares the similarity of two sets of data by aligning them to minimize differences[12]. Procrustes analysis allowed us to determine how well the microbial community structures from the oral and gut sites matched. We used the procuste.randtest function from the ade4 R package (version 1.7.22) for this purpose. This function takes the PCs of the oral and gut microbiomes as input and assesses the significance of their alignment. Specifically, we compared the bacterial community structures between the oral and gut microbiomes across the 12 studies. We performed these analyses separately for each body site and for each study to ensure a comprehensive assessment. We started by measuring the inter-community variations within the oral and gut microbiomes using three different distance measures: Kendall, Bray, and Aitchison. Additionally, we calculated a correlation matrix using the Spearman correlation coefficient to evaluate the relationship between the abundance data and the PCs. This analysis was performed separately for each study and each body site within those studies. In simpler terms, we used Procrustes analysis to see how similar the microbial communities in the oral are to those in the gut. By comparing these communities across different studies, we could understand how variations in one body site might influence the other. Here we aimed to uncover any significant correlations between the microbial communities of the oral cavity and the gut, providing insights into how these communities might influence each other across different individuals.

Schematic of this process is given in [Figure 2.4]

2.2.2 Identification of oral and fecal microbiome modules using meta-network analysis

To identify the modules we performed the meta-network analysis, it based on Random Effects Model (REM) plays a crucial role in analyzing the co-abundance relationships between microbial species across multiple studies. To analyze the relationships between the oral and gut microbiomes, we examined data from 12 different studies. Our goal was to identify consistent patterns of co-abundance between microbial species within the oral microbiome, within the gut microbiome, and between the oral and gut microbiomes. first, we combined the species-level abundance data from the fecal samples and each

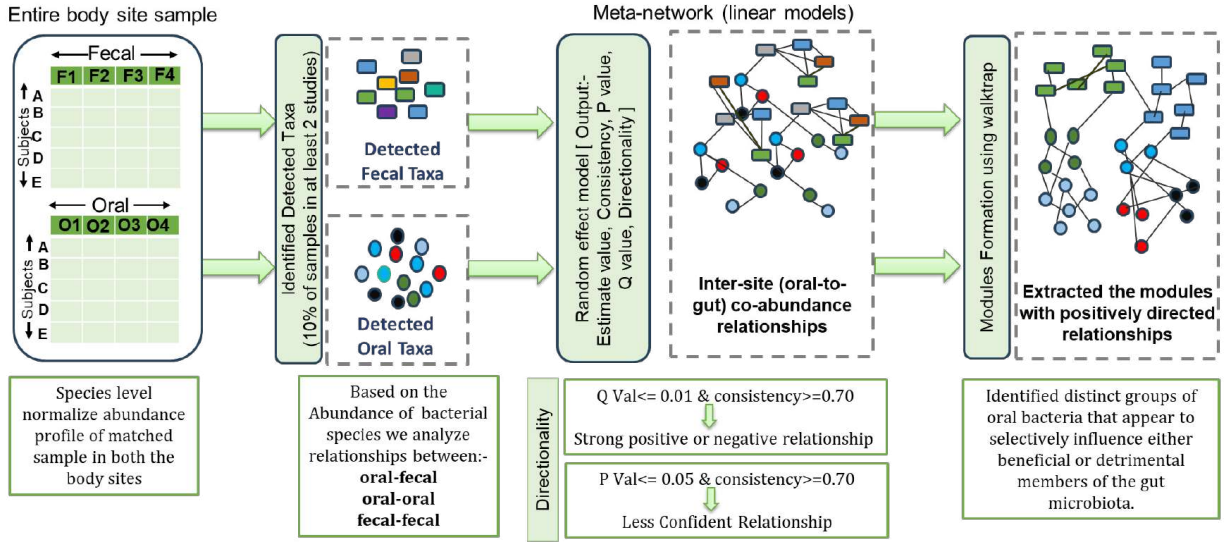


Figure 2.5: Representation of meta-network analysis for identification of oral and fecal microbiome modules.

oral subsite for every study. We labeled the species to distinguish between oral and gut

origins. We then created a merged dataset for each oral subsite, including all relevant studies. To differentiate between oral and gut microbes in the combine dataset we adds the separator '__O' for oral and '__F' for gut) to each species. From this combined dataset, we selected species that were present in at least 10% of samples in at least two studies for to use highly detected taxa for networking. These selected species were used for further analysis. Using this combine dataset of detected species on given criteria for that particular body site combination of oral to gut we employed a statistical technique called the Random Effects Model (REM) to perform a meta-analysis on the abundance data of any two species across the 12 studies. The Random Effects Model (REM) is a statistical approach used in meta-analysis to combine and analyze results from multiple studies. It is particularly useful when there is variability among the studies, which is common in biological and clinical research. REM accounts for both within-study variation and between-study variation. This allowed us to analyze relationships between species pairs within the oral microbiome, within the gut microbiome, and between the oral and gut microbiomes.

The results of this analysis included p-values, FDR-adjusted p-values (q-values), estimated relationship strengths, consistency across studies, and the direction of the relationships. For each significant co-abundance relationship, we obtained details explains the strength of the association (p-value), its reliability across studies (consistency) of observed relationships across different studies. A consistency threshold of $>= 0.70$ is applied. This means that for a relationship to be considered for further analysis, it must be observed in at least 70% of the studied cohorts, The directionality suggest direction of the effect (positive or negative). Here, q-value tells that any relationship with an adjusted p-value (q-value) less than or equal to 0.01 is considered statistically significant. The consistency matrix contains values representing the consistency of observed relationships across different studies. Each cell in this matrix represents the proportion of studies in which the relationship between a pair of species was observed.

Schematic of this process is given in [Figure 2.5]

Using the directionality values from the REM analysis, based on directionality we focused on the positively directed relationships and identified modules within the network. To do this, we applied the Walktrap algorithm, which uses random walks to find groups of nodes (species) that are closely connected to each other. The igraph package in R was used to perform this analysis. Once the modules were identified, we visualized the detailed network using Cytoscape software. Cytoscape is a platform that allows for the creation and visualization of complex networks. We provided the modules data to Cytoscape to observe clear groupings and connections between the oral and gut microbiomes. We performed these steps for different oral subsites, such as saliva, tongue dorsum & buccal swab, in relation to fecal microbiome abundance. This

approach allowed us to identify consistent and significant co-abundance relationships between microbial species in various parts of the oral and gut microbiomes.

2.2.3 Quantitatively assessing the extent of gut microbiome variations explained by different oral microbiome species

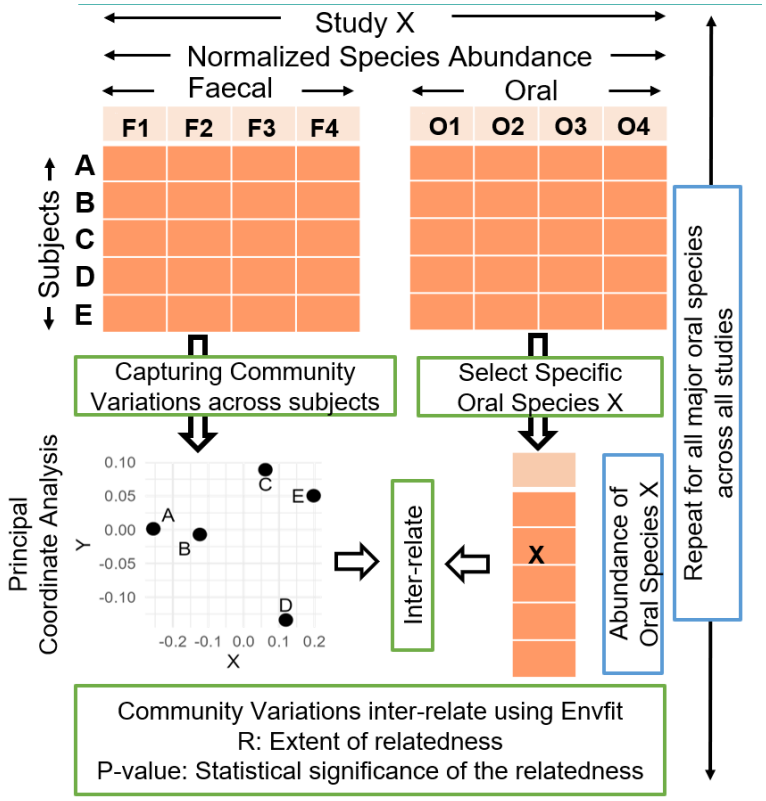


Figure 2.6: Schematic representation of assessing the extent of gut microbiome variations explained by different oral microbiome species using Envfit method

From this investigation we tried to find which specific species in the oral microbiome influenced the gut microbiome composition across different studies. We compare how different oral microbes appear in the gut across various diseases, different individuals, and their lifestyle habits. We look at what features of the gut microbiome might help these oral microbes move and survive in the gut. To explore this translocation of microbiomes we performed the meta-analysis technique known as Envfit it is regression

analysis. Envfit is a function of vegan package in R which helps to check community variation. This function fits environmental variables to ordination scores, helping us see which species are associated with which variables. We carried out this investigation on species level abundance profile. We reduce the dimension of fecal microbiome abundance using the PCoA (dimensionality reduction technique). The analysis involved performing PCoA with the dudi.pco function from the vegan package performed on distance matrix formed using Kendall tau correlation coefficient using the cor.fk function, Envfit then assessed whether the presence of specific oral microbial species, represented by the oral microbiome data, could statistically explain the variations observed in the gut microbiome composition based on oral. This approach allowed us to identify significant correlations between the presence of particular oral microbes and the overall composition of the gut microbiome across different studies. To analyze the relationships between the oral microbiome composition and environmental variables, we fitted the fecal PCoA ordination results with each species in oral microbiome abundance data using the envfit function. Repeat steps for each 12 studies for each bodysites.

This analysis gives correlation values, p-values, and other statistical outputs for each environmental variable(in our case oral species), where higher correlation values and lower p-values indicated stronger associations. It shows the strength and significance of these associations. We analyze using this, One group of communities varies due to any environmental factor that is influencing this community vise variation, how strongly that influence is, and whether that influence is significant or not. This analysis helped us gain insight into how environmental factors impact community variation in the gut microbiome, the strength of that impact, and whether it is significant. This investigation was performed on a each oral-species and each study level.

A subset of the oral and fecal microbial taxa were selected for further analysis based on their prevalence across the studies. This selection process ensured that only taxa consistently detected were included. Specifically, a threshold of being present in at least 10% of samples across at least two studies was employed. We initially selected 510 oral microbial taxa that were detected in at least 10 percent of samples across a minimum of two studies out of nine studies of saliva, and 524 fecal microbial taxa that were detected in at least 10 percent of samples across at least two studies from the same nine studies on saliva. Additionally, we selected 102 oral microbial taxa detected in at least 10 percent of samples across a minimum of two studies out of two studies on buccal swab, and similarly 96 fecal microbial taxa detected in at least 10 percent of samples across at least two studies from the same two studies of buccal swab. Furthermore, we selected 103 oral microbial taxa detected in at least 10 percent of samples across a minimum of two studies out of two studies on tongue dorsum, and 108 fecal microbial taxa detected in at least 10 percent of samples across at least two studies from the same two studies

on the tongue dorsum. And identified the oral species influence is significant or not on other community variation using this investigation. The schematic of this process is given in [Figure 2.6]

2.2.4 Identifying oral influencers of the gut microbiome community and vice-versa across different studies and identifying clusters of oral co-influencing the gut microbiome in specific studies (and the same for gut microbial taxa)

Figure 2.7: Representation of Identifying oral influencers using machine learning approach Random forest method

Machine learning technologies have been applied explicitly to microbiome studies from several years. In our study, we utilized Random Forest, a well-known bagging technique & its algorithm is the aggregation of the Decision trees. Random forest effectively helps to capture complex feature patterns in two different groups, which is why this method is used majorly for microbiome data analysis. This method is widely employed in microbiome data analysis due to its ability to discern complex patterns between distinct groups. Our research focused on leveraging Random Forest to identify fecal bacteria reproducibly predicted based on oral microbiome composition.

The research was executed in three stages, focusing on five distinct body sites across twelve different studies. In the first stage, we utilized the randomForest library in R to predict the composition of fecal bacteria based on the microbiota found in the oral cavity. This involved using the abundance of various oral bacterial species as predictors to estimate the bacterial composition of fecal samples. The randomForest algorithm provided us with out-of-box correlation values, which measure the accuracy of our predictions(significance). Additionally, we obtained feature importance scores, indicating which oral bacterial species were most influential in predicting the composition of fecal

microbiota. This approach allowed us to understand the predictive relationship between oral and gut microbiomes and identify key oral bacteria that could influence gut bacterial composition.[Figure: 2.7]

In the second stage, we focused on predicting the only fecal specific taxa based on oral and oral species present in gut environment composition. However, this time, we repeated the prediction process using only those species that were specific to the fecal environment as the response variables. This allowed us to determine how well the oral microbiota could predict the presence and abundance of specific fecal bacteria, thus further elucidating the connections between the oral and gut microbiomes. By isolating fecal-specific species in our analysis, we aimed to refine our understanding of the interplay between these two microbial communities and identify potential markers of microbiome health that are consistent across different body sites.

In last stage we perform the feature selection, for this we use the feature importance score get in previous stages then rank the each response (fecal species) in between 0 to 1 across all the features. then we just use only those features whose ranked feature importance score is greater than 70% for that particular fecal species. Based on selected feature we predicted again to get more accurate prediction in two ways leave one out cross validation across all studies in that bodysite and out of box prediction within that specific study. This comprehensive approach allowed us to investigate the reproducible prediction of fecal bacteria based on oral microbiome composition across multiple studies and body sites. Using Random Forest provided robust insights into feature importance and predictive correlation, facilitating a deeper understanding of the intricate relationships between oral and fecal microbiomes.

Chapter 3

RESULTS

3.1 Results of profiling overall cross-body-site relatedness between the oral and fecal

3.1.1 Results of profiling the cross −community variations across the different individuals, within the same body sites using PCoA

In this study, we aimed to profile the cross-community variations across different individuals within the same body sites (oral or gut/fecal) using Principal Coordinate Analysis (PCoA). By converting higher-dimensional microbiome data into two-dimensional representations, we were able to simplify the complexity and focus on the most significant features. This section presents the results of our PCoA analysis, highlighting the discrete separation observed between the oral and gut microbiomes across different studies.

The results provide valuable insights into the complex relationships between oral and gut microbiomes. The PCoA plots consistently show that oral samples cluster tightly together, indicating consistent oral microbiome profiles among the individuals in each study. In contrast, the stool samples are positioned distinctly from the oral samples, reflecting the unique microbial environment of the gut. This clear segregation underscores the distinct microbial ecosystems present in the oral and gut environments.

The consistent patterns observed across all studies reinforce the robustness of our findings, demonstrating distinct microbial profiles for the oral and gut environments. Specific results for each study cohort, showing the cross-community variations, can be referred to Fig. 3.1 and Fig. 3.2 Additionally, an overall PCoA plot using Kendall

distance matrices, which shows discrete separation between oral and gut samples, can be referred to in Fig. 3.3

3.1.2 Correlation results of Inter-Individual Variations Between Oral and Gut Microbiomes

Our analysis revealed significant correlations between the microbial communities in the oral cavity and the gut, as evidenced by the Kendall distance metric. We observed significant results (p ¡ 0.05) across multiple studies 3.4, with p-values ranging from 0.001 to 0.205. Out of the 12 studies analyzed, 10 showed significant correlations, indicating a strong and consistent association between the microbial compositions at these two body sites.

Beta diversity, assessed using Bray-Curtis dissimilarity values, showed considerable variation across the studies, ranging from 0.001 to 0.947. This wide range highlights the substantial differences in community composition between the studies, reflecting the diverse factors that influence microbial communities at the oral and gut sites. A similar pattern was observed using Aitchison distance, further reinforcing the variability in microbial community composition across different studies 3.5.

The significant correlations observed through Kendall distance analysis suggest that variations in the microbial communities at one body site (oral) are likely to be mirrored at the other site (fecal/gut). Specifically, individuals with similar oral microbial communities, as indicated by similar oral Principal Coordinates (PCs), are likely to have similar gut microbial communities, reflected by similar gut PCs, and vice versa.

By incorporating both oral and gut microbiome data from the same individuals and applying diverse distance measures, we gained a more comprehensive understanding of the potential inter-site relationships. These correlations suggest that the microbial community composition in one body site can be correlated with the composition in another, providing valuable insights into the interconnected nature of microbial ecosystems within the human body.

3.2 Results of meta-network analysis showing oral and fecal microbiome modules

Our meta-network analysis, based on the Random Effects Model (REM), revealed significant co-abundance relationships between microbial species within and between the oral and gut microbiomes across different studies. By integrating species-level abundance data from fecal samples and various oral subsites, we identified consistent patterns

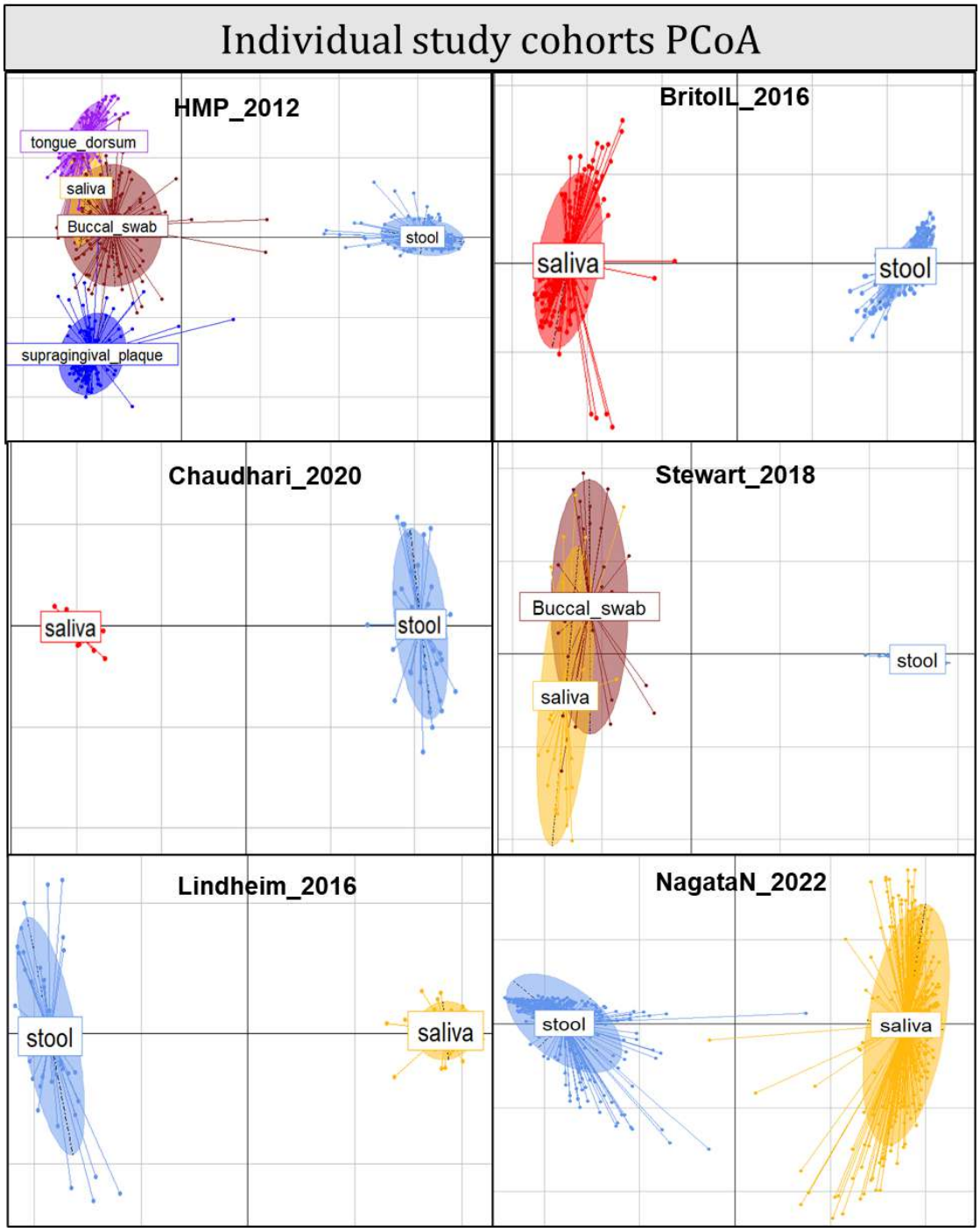


Figure 3.1: Principal Coordinate analysis Results for each study cohort showing the cross-community variations

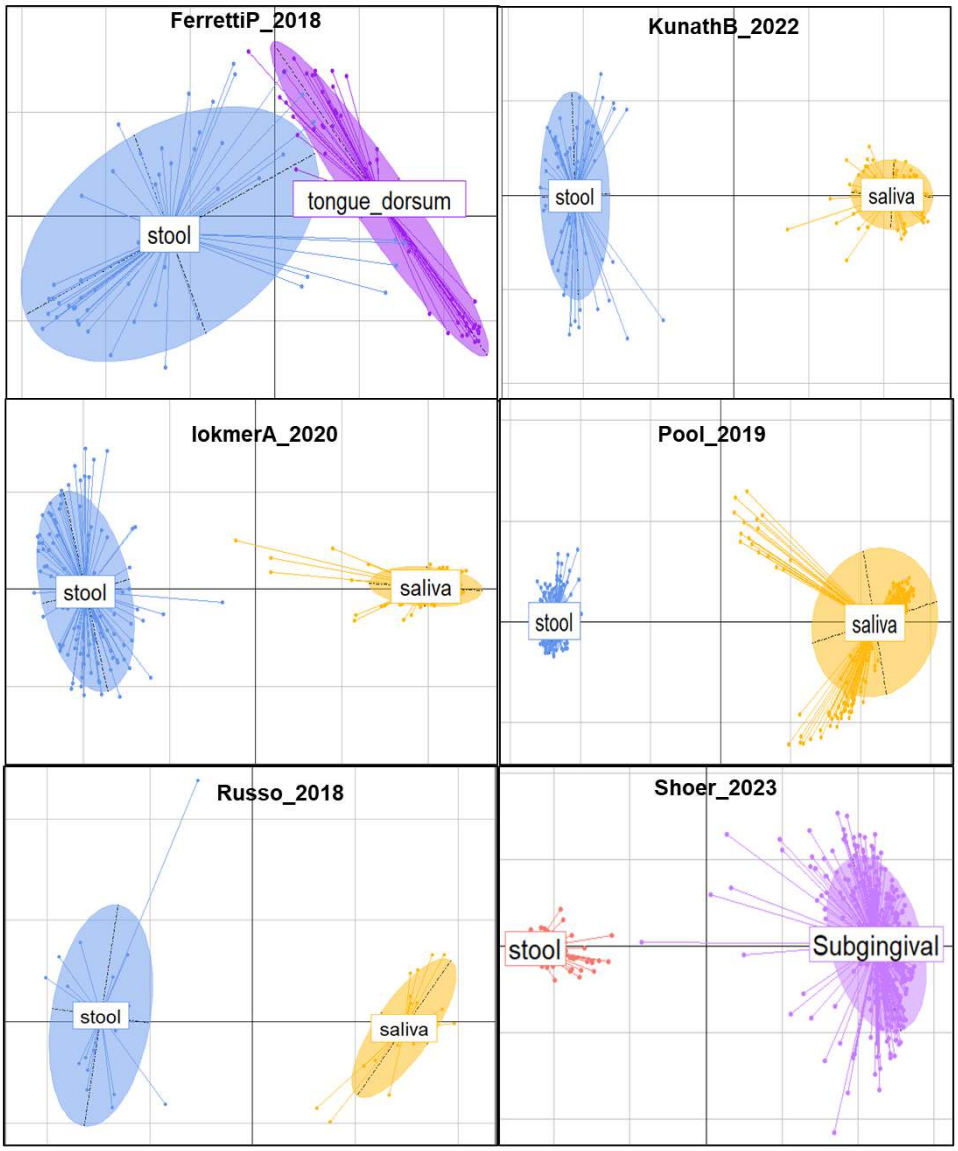


Figure 3.2: Principal Coordinate analysis Results for each study cohort showing the cross-community variations

of co-abundance, indicating potential inter-site interactions. The results observed the clearcut grouping and connecting link between oral and fecal microbiomes.

In the saliva-fecal microbiome network, we identified four distinct oral modules that

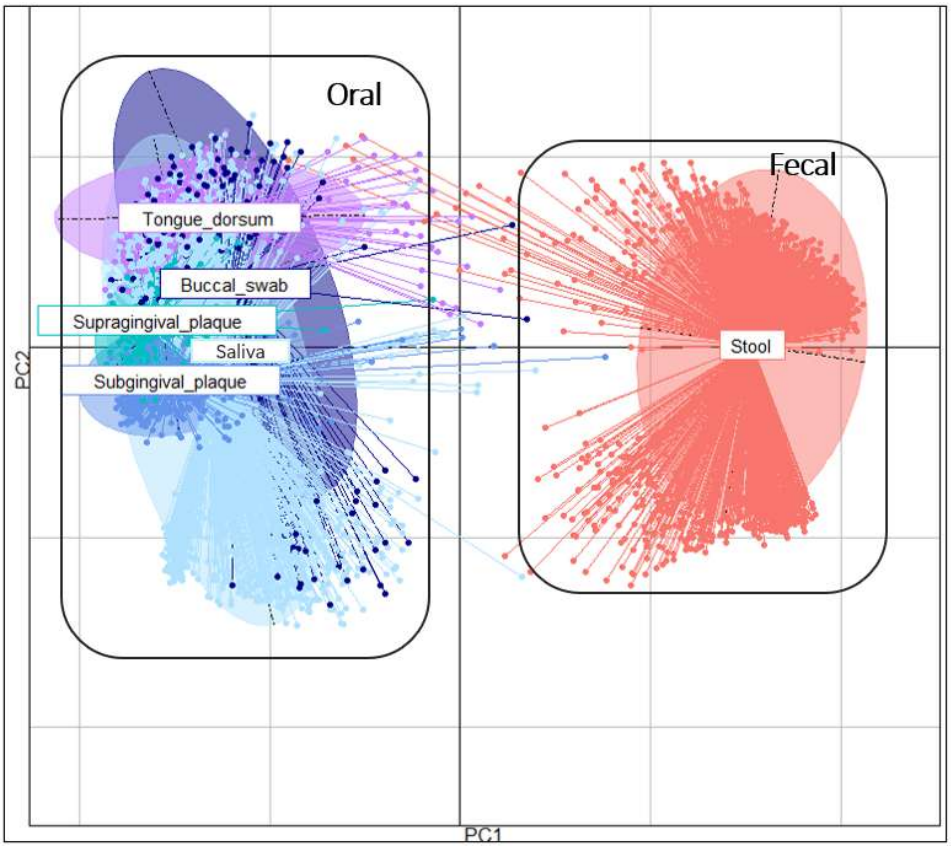


Figure 3.3: Principal Coordinate analysis plot using Kendall distance among all microbes at species level for all the sample in this investigation showing discrete seperation between oral and gut.

were interconnected with both beneficial and detrimental fecal bacteria. Salivary modules were divided into distinct hubs based on interaction patterns identified using the Walk-trap approach. This division revealed that certain oral microbes are associated with beneficial fecal microbes, while others are linked with potentially harmful fecal microbes. This indicates that the composition of the oral microbiome has the potential to modulate the gut microbiome, suggesting a bidirectional influence between these two sites.3.6 The buccal swab-fecal microbiome network revealed eight distinct oral modules. Similar to the saliva network, these modules showed connections to both beneficial and detrimental fecal bacteria. This pattern underscores the significant role of the oral microbiome in influencing gut microbiome composition and highlights specific oral microbial groups that may be key modulators of gut health.3.7 In the tongue dorsum-fecal

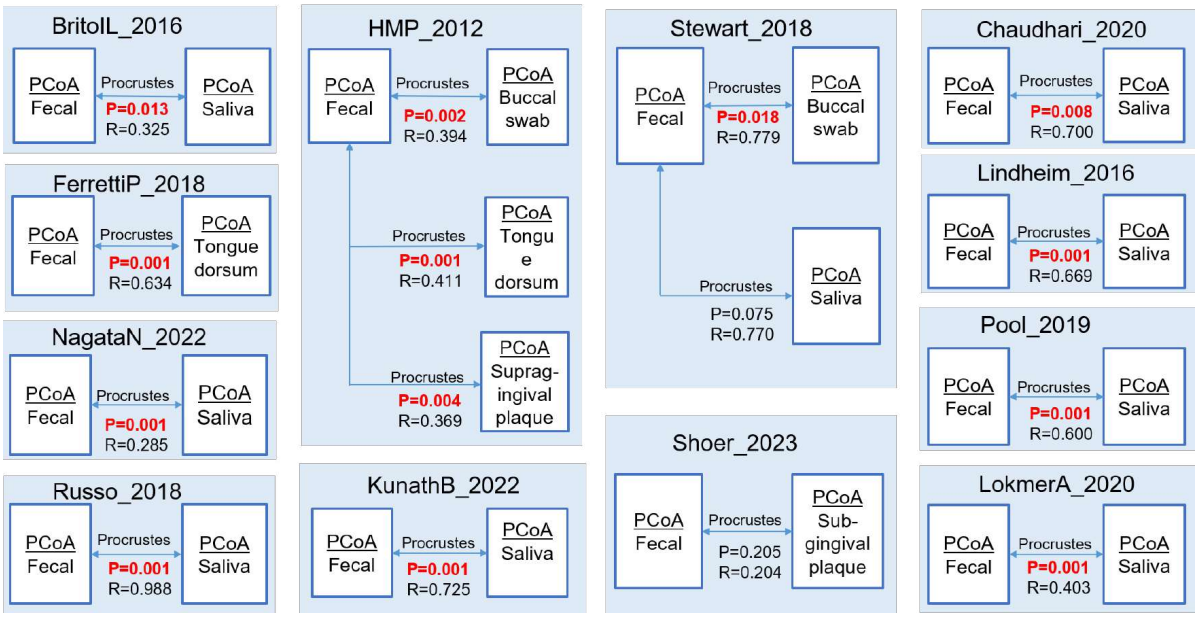


Figure 3.4: Significant results observed for each study using Kendall distance where $p < 0.05$ (across studies p-value ranges from 0.001 to 0.205) where 10 studies show significant results out of 12.

microbiome network, we identified five distinct oral modules. These modules, like those identified in the saliva and buccal swab networks, demonstrated connections to both beneficial and detrimental fecal bacteria. The consistent pattern observed across different oral subsites suggests a generalized mechanism through which the oral microbiome can affect gut microbiome composition.3.8 Overall, our meta-network analysis uncovered significant and consistent co-abundance relationships between the oral and gut microbiomes. The distinct modules identified within the oral microbiomes and their connections to beneficial and detrimental fecal bacteria highlight the potential of the oral microbiome to modulate gut microbiome composition. These findings provide valuable insights into the interconnected nature of microbial ecosystems within the human body and underscore the importance of considering oral-gut microbiome interactions in future research and therapeutic strategies.

Study name	Body_site [vs stool]	P value			R value		
		Bray	Kendall	Aitchison	Bray	Kendall	Aitchison
BritoIL_2016	Saliva	0.493	0.013	0.043	0.274	0.325	0.319
FerrettiP_2018	Tongue dorsum	0.001	0.001	0.001	0.562	0.634	0.621
HMP_2012	Buccal swab	0.149	0.002	0.083	0.331	0.394	0.367
	Tongue dorsum	0.084	0.001	0.001	0.302	0.411	0.375
	Supragingival plaque	0.136	0.004	0.013	0.321	0.369	0.362
NagataN_2022	Saliva	0.008	0.001	0.001	0.224	0.285	0.254
Russo_2018	Saliva	0.152	0.001	0.111	0.765	0.989	0.930
Stewart_2018	Buccal swab	0.239	0.018	0.132	0.618	0.779	0.746
	Saliva	0.134	0.075	0.346	0.601	0.770	0.737
Chaudhari_2020	Saliva	0.001	0.009	0.293	0.601	0.700	0.644
Lindheim_2016	Saliva	0.235	0.001	0.21	0.516	0.669	0.616
Pool_2019	Saliva	0.001	0.001	0.001	0.433	0.600	0.560
IokmerA_2020	Saliva	0.501	0.001	0.001	0.304	0.403	0.409
KunathB_2022	Saliva	0.001	0.001	0.001	0.527	0.725	0.597
Shoer_2023	Subgingival plaque	0.947	0.205	0.035	0.168	0.204	0.215

Figure 3.5: Procrustes results using three distance measures, i.e., Kendall distance, Bray-Curtis distance, and Aitchison distance.

3.3 Identified result of significant oral microbial species influencing gut microbiome variation.

Our analysis identified significant correlations between specific oral microbial species and variations in gut microbiome composition across multiple studies, highlighting the potential translocation and influence of oral microbes on the gut microbiome.

As shown in Figure 3.9, the bar plot illustrates the number of studies in which each oral species demonstrated significant results in the Envfit analysis. The top three species—Mycoplasma salivarium, Kingella denitrificans, and Atopobium rimae—showed significant associations with gut microbiome variations in more than 50% of the studies analyzed, along with several other species.

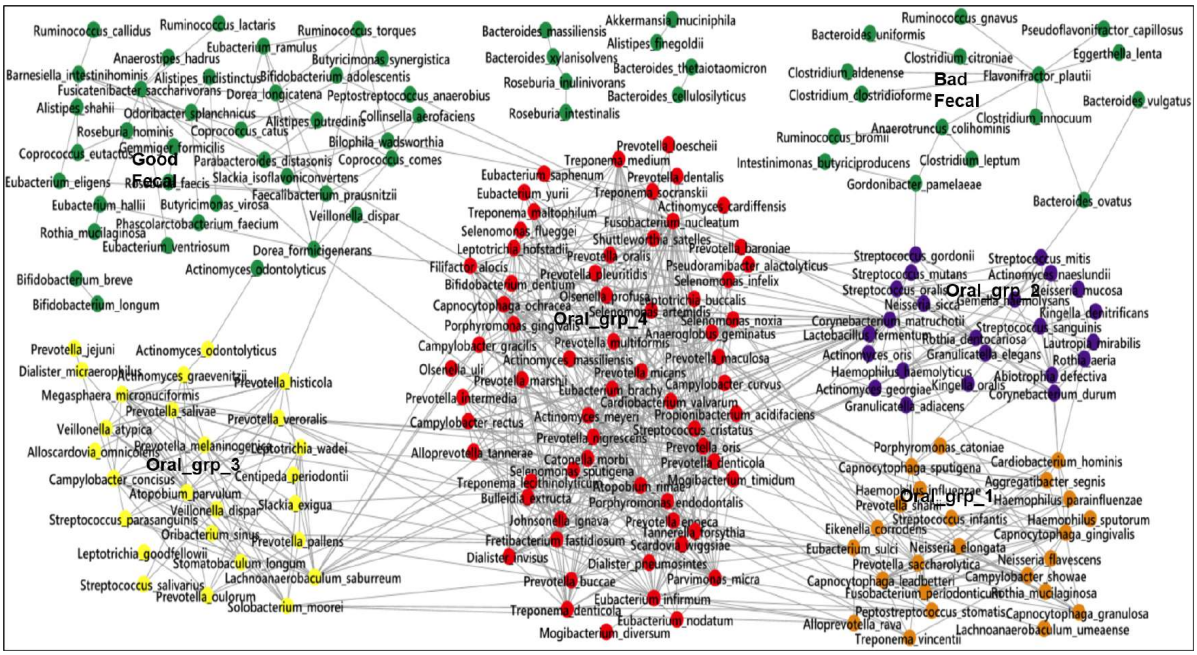


Figure 3.6: Visualization of modules between saliva and fecal bodysite microbiome composition, formed using meta-network analysis based on random effects model and plotted using Cytoscape tool.We can see four distinct oral modules that were interconnected with both beneficial and detrimental fecal bacteria.

These findings suggest that certain oral microbes consistently influence gut microbiome composition, despite differences in disease states, individual characteristics, and lifestyle factors across the studies. The presence and significance of these oral microbes in the gut indicate potential pathways for microbial translocation and colonization that might be influenced by environmental or host-specific factors.

The high significance levels observed for these top oral species across multiple studies underscore the robustness of their impact on the gut microbiome. This consistency provides strong evidence for the role of specific oral microbes in shaping gut microbial communities, which could have important implications for understanding the intercon-

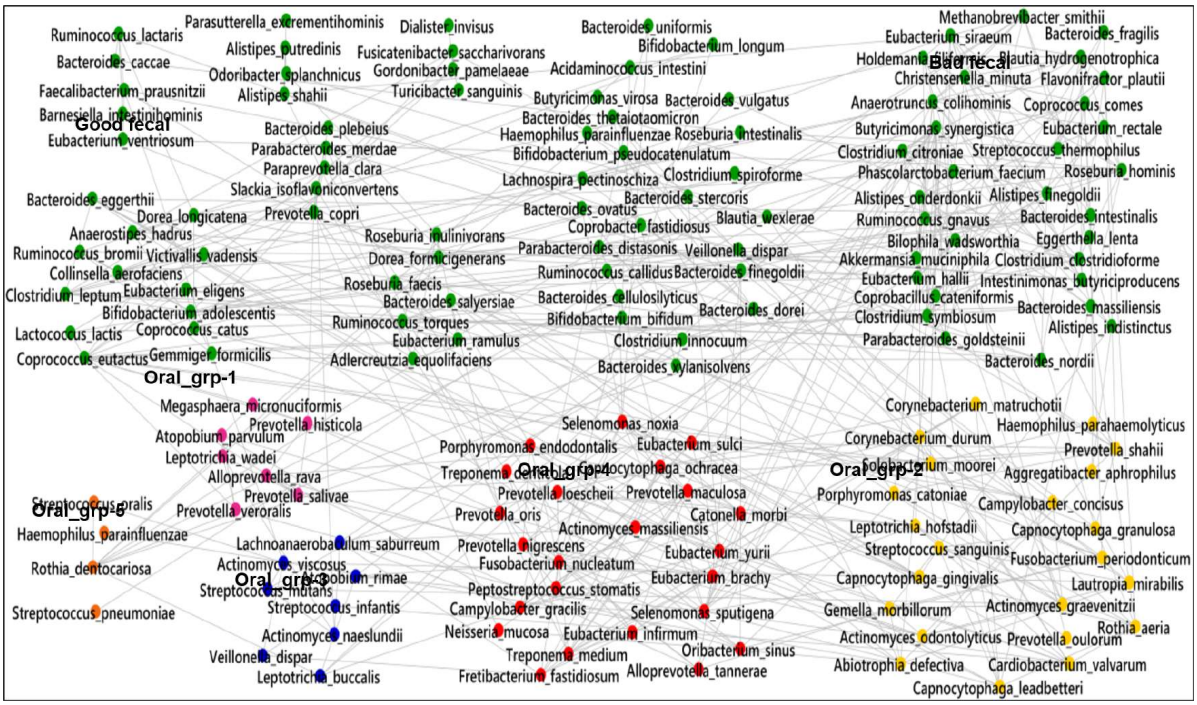


Figure 3.7: Visualization of modules between Buccal swab and fecal bodysite microbiome composition, formed using meta-network analysis based on random effects model and plotted using Cytoscape tool.We can see five distinct oral modules that were interconnected with both beneficial and detrimental fecal bacteria

nected nature of the oral and gut microbiomes. However, these results need further validation with larger sample sizes to ensure their reliability and to cross-check the findings.

Our results highlight the importance of considering oral microbial species when studying gut microbiome variations. This approach can provide valuable insights into the mechanisms through which oral microbes might influence gut health and disease, potentially guiding future therapeutic strategies aimed at modulating the microbiome to improve health outcomes.

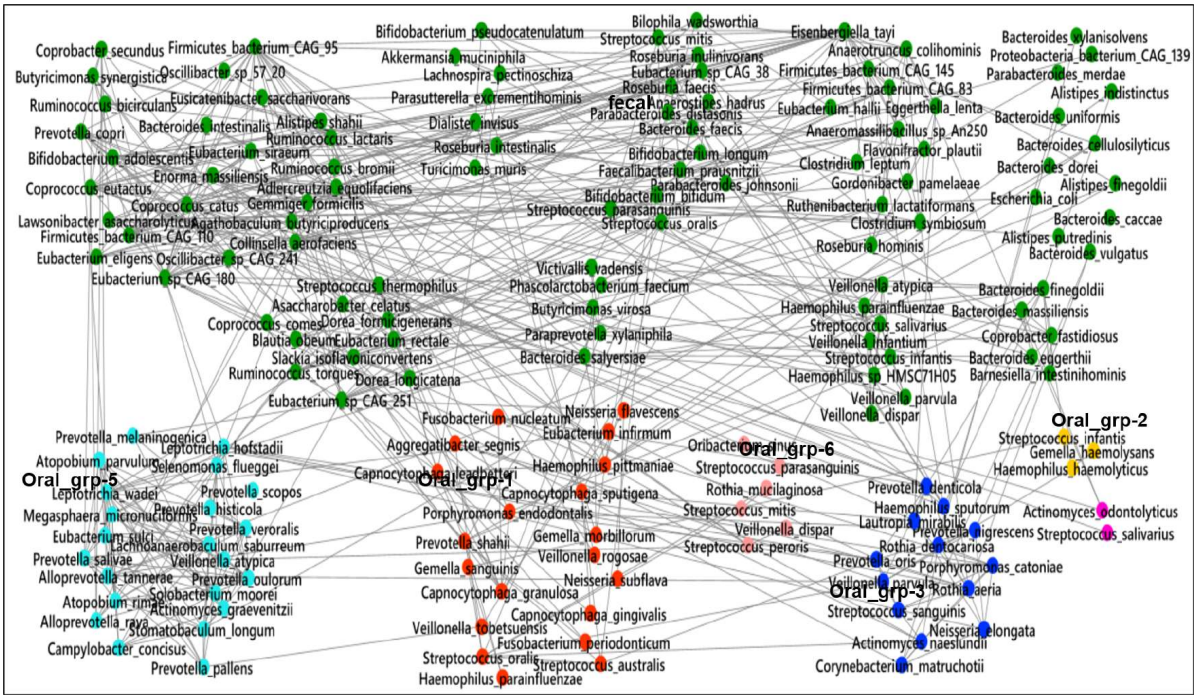


Figure 3.8: Visualization of modules between tongue dorsum and fecal bodysite microbiome composition, formed using meta-network analysis based on random effects model and plotted using Cytoscape tool.We can see six distinct oral modules that were interconnected with both beneficial and detrimental fecal bacteria.

3.4 Reults of Machine Learning Insights into Oral-Gut Microbiome Interactions.

This results focuses on Predicting Fecal Microbiome based on Oral Composition. This approach enabled us to identify key oral bacterial species influencing gut microbiota across various studies and body sites.

Figure 3.10shows the out-of-box (OOB) correlation results for predicting fecal bacterial composition based on saliva microbiota across nine studies. The heatmap illustrates clusters of studies with high predictive accuracy, indicating that certain oral bacte-

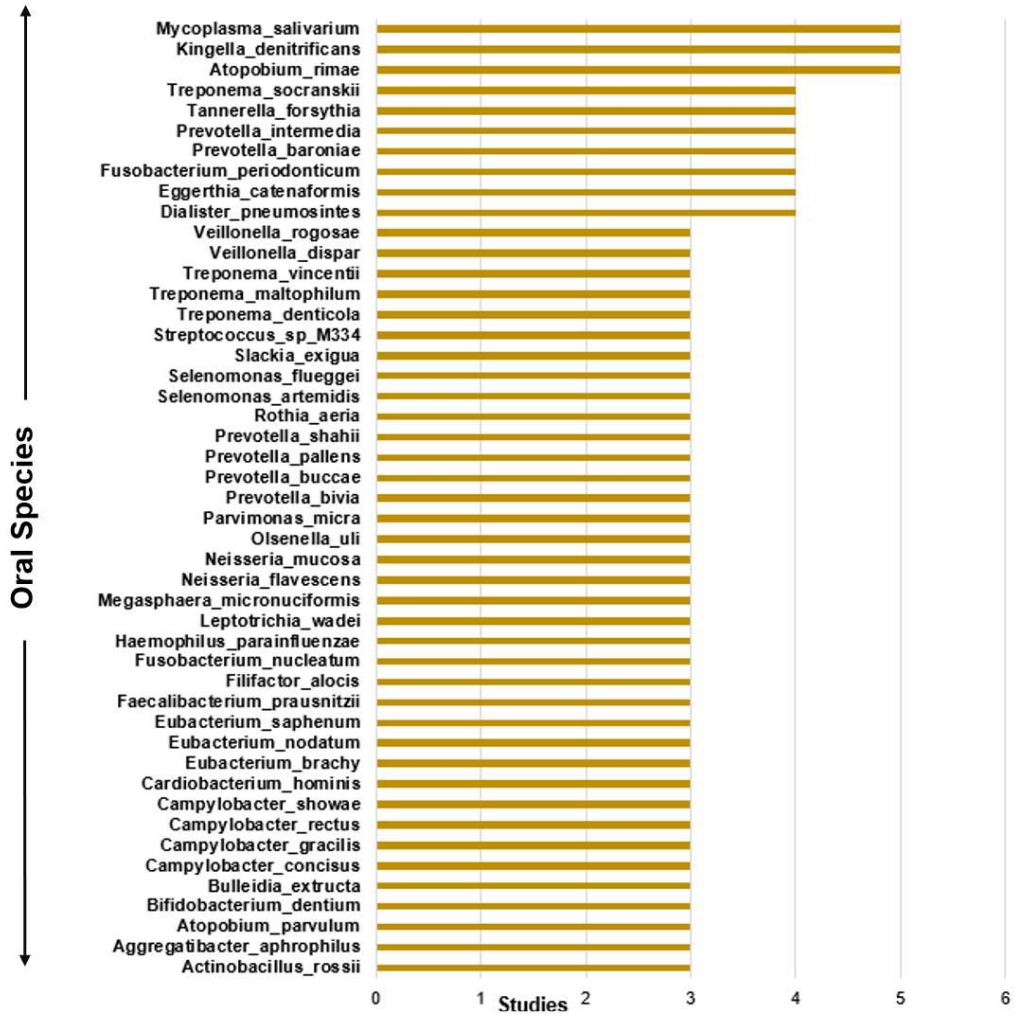


Figure 3.9: Significantly Observed oral species in more than 50% of studies after EnvFit analysis.The top three species: Mycoplasma salivarium, Kingella denitrificans, and Atopobium rimae showed significant associations with gut microbiome variations in more than 50% of the studies analyzed, along with several other species.

ria consistently predict gut microbiota composition. Notably, three studies clustered together, exhibiting high correlation values, while others displayed distinct patterns, suggesting variability in predictability among different cohorts.

Figure 3.11 presents a heatmap of mean ranked feature importance scores for predicting fecal bacteria using saliva microbiota. The heatmap shows four distinct bands, each representing groups of fecal species that are significantly predicted by specific oral bacterial species. This indicates that certain oral bacteria have a strong influence on predicting multiple fecal species, highlighting their potential role in modulating gut microbiota.

Figure 3.12 shows a similar heatmap for buccal swab and fecal microbiota. Here, we observe four clear bands of feature importance scores, similar to the saliva-fecal results. This consistent pattern across different oral sites underscores the robustness of our findings, suggesting that specific oral bacteria are crucial predictors of gut microbiome composition across different body sites.

Figure 3.13 and Figure 3.14 illustrate the results of feature selection and subsequent Random Forest predictions for saliva-fecal pairs. In Figure 4, we used only features with a ranked feature importance score greater than 0.7 for each specific fecal species, resulting in improved OOB correlation values for the nine studies. The heatmap shows enhanced predictive accuracy, indicating that selecting the most important features significantly boosts model performance.

Figure 3.14 compares the OOB results with leave-one-out cross-validation (LOOCV) for saliva-fecal predictions across the nine studies. The OOB results demonstrate superior performance, suggesting the presence of cohort-specific signatures that are captured more effectively with this method. This highlights the variability among different cohorts and the importance of using robust validation techniques to ensure accurate predictions.

Our findings indicate that it is possible to predict both beneficial and harmful gut bacteria based on their oral composition. The high correlation and feature importance scores across multiple studies and body sites underscore the significant role of specific oral microbes in shaping gut microbiome composition. The superior performance of OOB results compared to LOOCV suggests that cohort-specific signatures are prevalent, emphasizing the need for incorporating additional datasets to enhance generalizability and validate these patterns further.

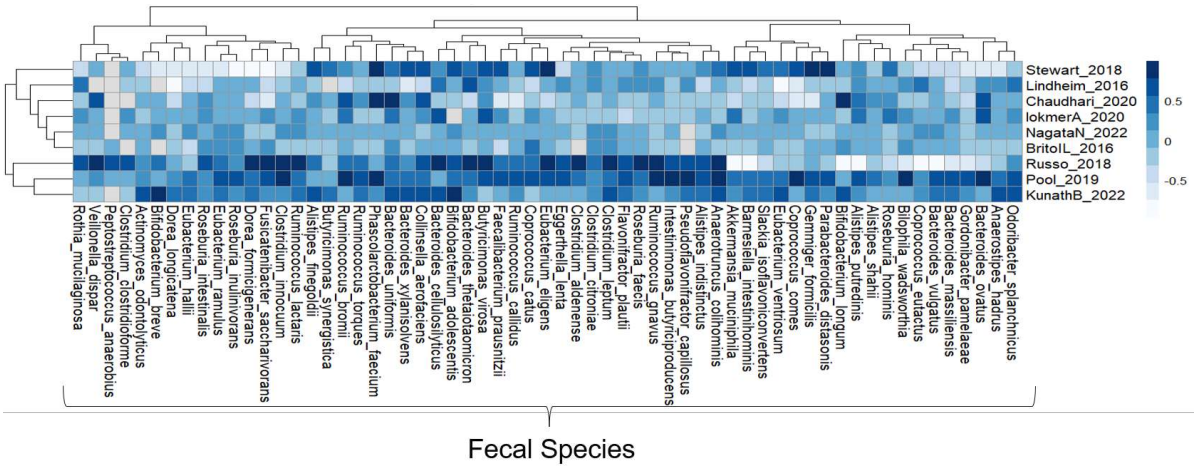


Figure 3.10: Heatmap represents Out of box correlation for predicted fecal species based on fecal using random forest for saliva bodysite.

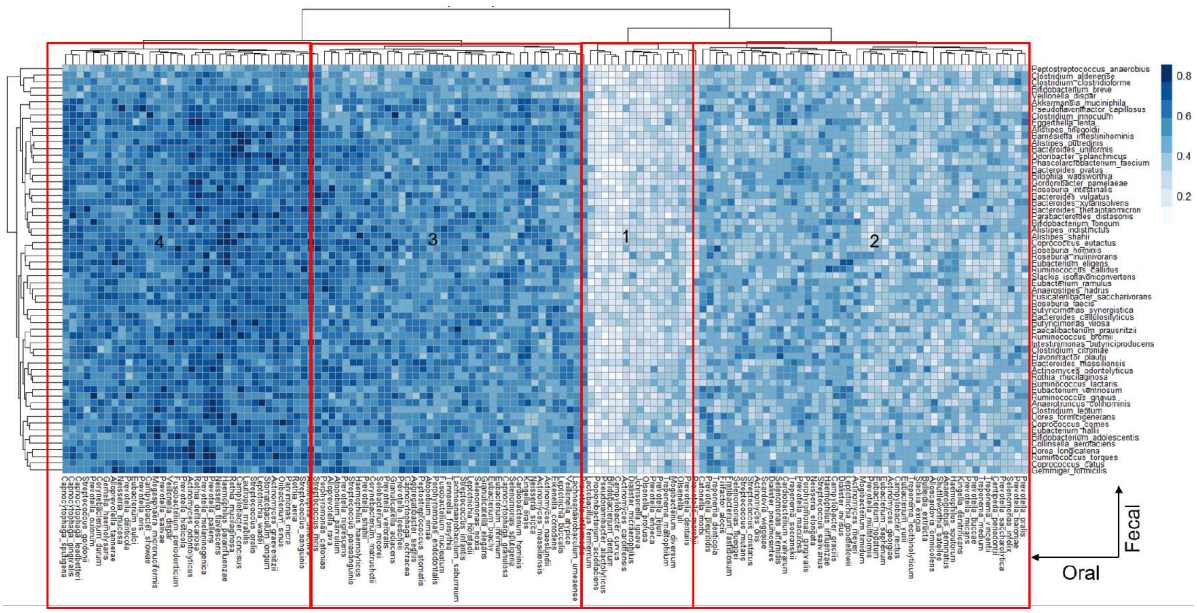


Figure 3.11: Heatmap represents Mean Ranked Feature Importance Score for saliva body site obtained using random forest.

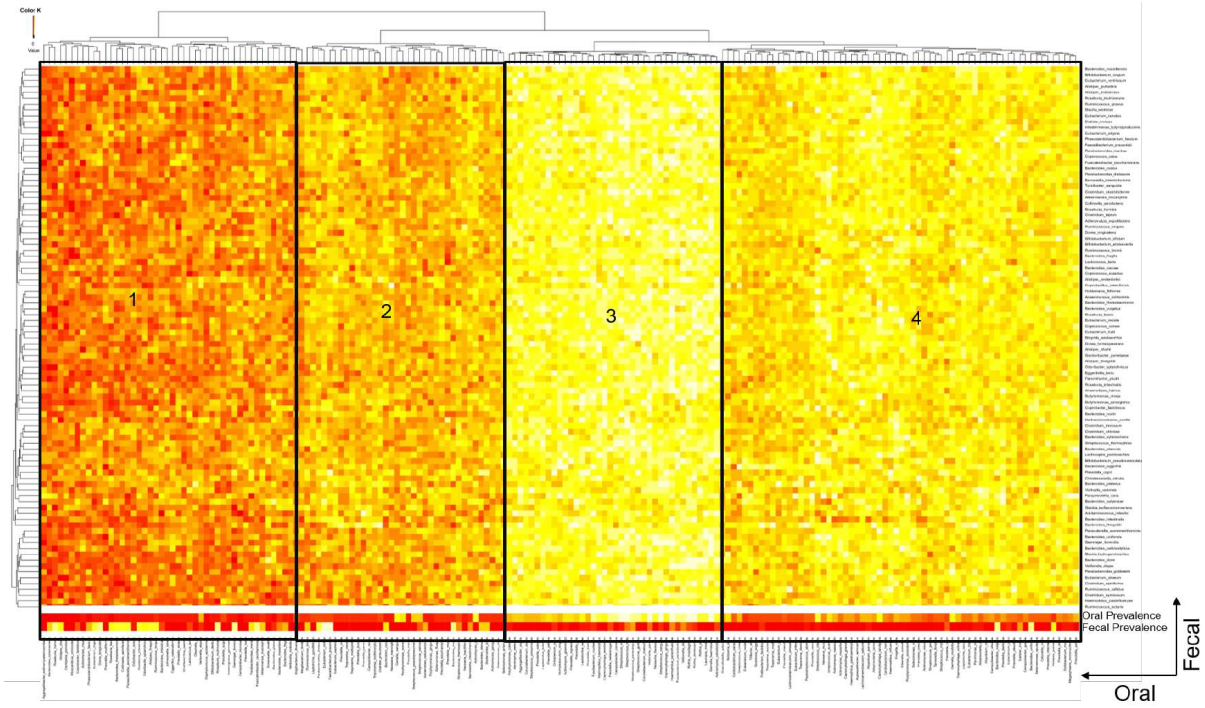


Figure 3.12: Heatmap represents Mean Ranked Feature Importance Score for Buccal swab body site obtained using random forest with showing feature(oral species) prevalence in oral and fecal.

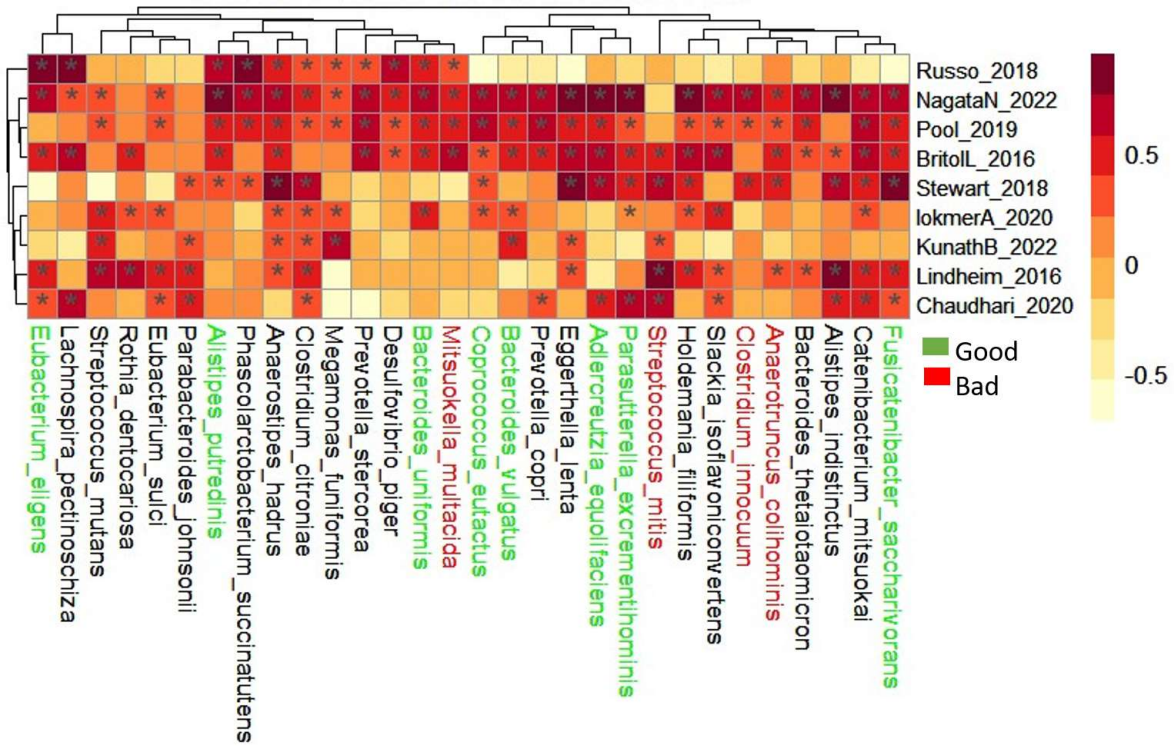


Figure 3.13: The heatmap illustrates Random Forest out-of-box (OOB) estimates for salivary datasets, showcasing enhanced prediction accuracy through feature selection. Cells highlighted with a star mark indicate significant positive correlations, demonstrating our ability to predict both beneficial and harmful bacteria using this approach.

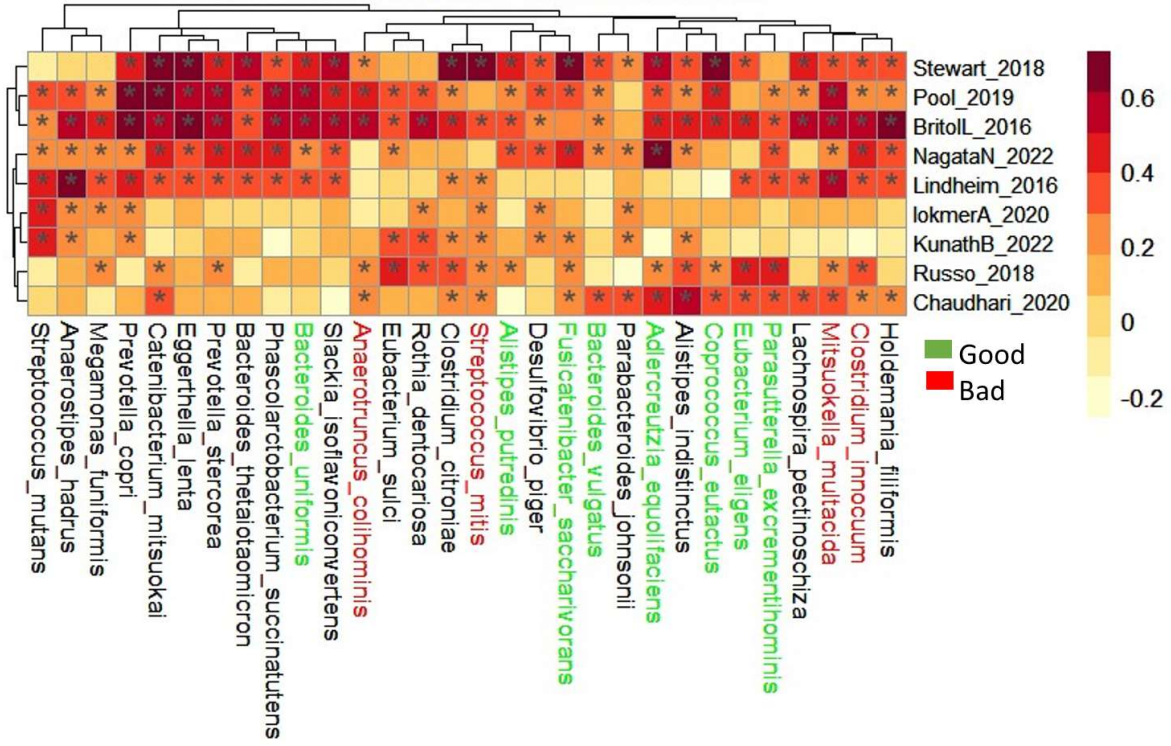


Figure 3.14: The heatmap illustrates Random Forest Leave one out cross validation(LOOCV) estimates for salivary datasets, showcasing enhanced prediction accuracy through feature selection. Cells highlighted with a star mark indicate significant positive correlations, demonstrating our ability to predict both beneficial and harmful bacteria using this approach.

Chapter 4

Conclusion and Future Scope

4.1 Conclusion

This data-driven exploration of cross-body site interactions between oral and gut microbiome communities revealed significant correlations across multiple studies, indicating a robust link between these two distinct body sites. Our meta-analysis, encompassing over 1,500 human subjects from 12 global cohorts, identified distinct modules within the microbiome that demonstrate significant co-abundance relationships between oral and gut species. This co-abundance suggests potential functional interactions that could be crucial for understanding the dynamics of microbiome-associated health and disease. Identifying specific oral microbes that correlate with gut health opens new avenues for therapeutic and diagnostic development. Future research should focus on isolating these microbes and understanding their roles in gut health. This knowledge can be leveraged to develop novel probiotics, prebiotics, and microbiome-modulating therapies. Observing that certain oral taxa can influence the gut microbiome without physically translocating suggests the presence of remote control mechanisms which adds good insight in this are. The out-of-bag (OOB) performance, superior to leave-one-out cross-validation (LOOCV), indicates the presence of cohort-specific factors influencing the predictions. These cohort-specific signatures suggest that while there are common patterns, individual variations must be considered. In conclusion, this comprehensive analysis highlights the significant correlation between oral and fecal microbiome communities and their potential functional interactions. The findings emphasize the importance of considering cross-body site microbiome interactions in health and disease studies. By expanding datasets, investigating mechanistic pathways, and developing predictive models, we can advance our understanding of microbiome dynamics and their implications for human health.

4.2 Future Perspectives

Further research is needed to explore the underlying mechanisms of this relationship and the influence of other factors on the gut microbiome. By elucidating the specific pathways and interactions through which oral microbes influence gut health.A comprehensive analysis should correlate microbiome changes with various environmental factors such as diet, antibiotics usage, stress, and pollutants. Understanding how these external factors modulate the oral and gut microbiomes will provide a more holistic view of their impact on human health and disease progression. Identifying specific taxa that migrate across body sites and observing cohort-specific migration patterns can offer insights into the dynamics of microbiome interactions. By understanding how and why certain microbes translocate, we can better comprehend their role in disease and health.Observing that certain oral taxa can influence the gut microbiome without physically translocating suggests the presence of remote control mechanisms. This could involve microbial metabolites, signaling molecules, or other indirect interactions. Investigating these remote influences will be crucial for understanding the full spectrum of oral-gut microbiome interactions. The identification of specific oral microbes that can serve as diagnostic markers for gut health and potential therapeutic targets opens new avenues for medical interventions. By targeting these microbes, it may be possible to develop probiotics, prebiotics, or other microbiome-modulating therapies to maintain or restore gut health. By providing global insights into the mechanistic details of the oral-to-gut axis, this data-driven study for the first time offers potentially novel insights into the utilization of the oral microbiome as a diagnostic marker and a promising therapeutic target of gut health. These findings can potentially help to advance the understanding of interaction patterns between cross-body sites, leading to more effective health interventions and preventive strategies.

Bibliography

[1] N. Segata, S. Haake, P. Mannon, K. P. Lemon, L. Waldron, D. Gevers, C. Huttenhower, and J. Izard, "Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples," *Genome Biology*, vol. 13, no. 6, p. R42, 2012. [Online]. Available: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2012-13-6-r42

[2] I. Olsen, "From the acta prize lecture 2014: the periodontal-systemic connection seen from a microbiological standpoint: Summary of the Acta Odontologica Scandinavia Price Lecture 2014 presented at the meeting of the IADR/Pan European Region in Dubrovnik, September 10–13. 2014," *Acta Odontologica Scandinavica*, vol. 73, no. 8, pp. 563–568, Nov. 2015. [Online]. Available: http://www.tandfonline.com/doi/full/10.3109/00016357.2015.1007480

[3] The Human Microbiome Project Consortium, "Structure, function and diversity of the healthy human microbiome," *Nature*, vol. 486, no. 7402, pp. 207–214, Jun. 2012. [Online]. Available: https://www.nature.com/articles/nature11234

[4] S. Andreu-Sánchez, J. Wu, and J. Fu, "Beyond personal space: Unveiling the transmission pattern of the human gut and oral microbiome," *iMeta*, vol. 2, no. 2, p. e98, May 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/imt2.98

[5] T. S. Schmidt, M. R. Hayward, L. P. Coelho, S. S. Li, P. I. Costea, A. Y. Voigt, J. Wirbel, O. M. Maistrenko, R. J. Alves, E. Bergsten, C. De Beaufort, I. Sobhani, A. Heintz-Buschart, S. Sunagawa, G. Zeller, P. Wilmes, and P. Bork, "Extensive transmission of microbes along the gastrointestinal tract," *eLife*, vol. 8, p. e42693, Feb. 2019. [Online]. Available: https://elifesciences.org/articles/42693

[6] P. Sudhakara, A. Gupta, A. Bhardwaj, and A. Wilson, "Oral Dysbiotic Communities and Their Implications in Systemic Diseases," *Dentistry Journal*,

vol. 6, no. 2, p. 10, Apr. 2018. [Online]. Available: http://www.mdpi.com/2304-6767/6/2/10

[7] I. Soffritti, M. D'Accolti, C. Fabbri, A. Passaro, R. Manfredini, G. Zuliani, M. Libanore, M. Franchi, C. Contini, and E. Caselli, "Oral Microbiome Dysbiosis Is Associated With Symptoms Severity and Local Immune/Inflammatory Response in COVID-19 Patients: A Cross-Sectional Study," *Frontiers in Microbiology*, vol. 12, p. 687513, Jun. 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmicb.2021.687513/full

[8] J. Xiao, K. A. Fiscella, and S. R. Gill, "Oral microbiome: possible harbinger for children's health," *International Journal of Oral Science*, vol. 12, no. 1, p. 12, Dec. 2020. [Online]. Available: https://www.nature.com/articles/s41368-020-0082-x

[9] G. Allard, F. J. Ryan, I. B. Jeffery, and M. J. Claesson, "SPINGO: a rapid species-classifier for microbial amplicon sequences," *BMC Bioinformatics*, vol. 16, no. 1, p. 324, Dec. 2015. [Online]. Available: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-015-0747-1

[10] L. J. M. Francesco Beghini, "Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3," May 2021. [Online]. Available: https://elifesciences.org/articles/65088

[11] G. Armstrong, G. Rahman, C. Martino, D. McDonald, A. Gonzalez, G. Mishne, and R. Knight, "Applications and Comparison of Dimensionality Reduction Methods for Microbiome Data," *Frontiers in Bioinformatics*, vol. 2, p. 821861, Feb. 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fbinf.2022.821861/full

[12] Z. Sun, S. Huang, M. Zhang, Q. Zhu, N. Haiminen, A. P. Carrieri, Y. Vázquez-Baeza, L. Parida, H.-C. Kim, R. Knight, and Y.-Y. Liu, "Challenges in benchmarking metagenomic profilers," *Nature Methods*, vol. 18, no. 6, pp. 618–626, Jun. 2021. [Online]. Available: https://www.nature.com/articles/s41592-021-01141-3