

# **NETWORK ANALYSIS OF GUT MICROBIOME**

A dissertation submitted to The University of Manchester for the degree of Master of Science in the Faculty of Biology,  
Medicine and Health

**2025**

**11353031  
School of Biological Sciences**

## CONTENTS

<b>LIST OF FIGURES .....</b>	<b>3</b>
<b>ABSTRACT .....</b>	<b>4</b>
<b>DECLARATION .....</b>	<b>5</b>
<b>INTELLECTUAL PROPERTY STATEMENT .....</b>	<b>6</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>7</b>
<b>INTRODUCTION.....</b>	<b>8</b>
HUMAN GUT MICROBIOTA .....	8
GUT MICROBIOTA AND HEALTH .....	8
GUT DYSBIOSIS.....	11
NETWORK ANALYSIS IN GUT MICROBIOME STUDIES.....	13
<b>STUDY RATIONALE .....</b>	<b>14</b>
<b>AIMS AND OBJECTIVES.....</b>	<b>15</b>
<b>MATERIALS AND METHODS.....</b>	<b>16</b>
<b>GMREPO DATA.....</b>	<b>16</b>
DATA PREPROCESSING .....	16
STANDARDISATION OF DISEASE AND TAXONOMY NAMES .....	16
ALPHA DIVERSITY ANALYSIS .....	18
BETA DIVERSITY ANALYSIS.....	18
DIFFERENTIAL ABUNDANCE TAXA ANALYSIS.....	19
NETWORK CONSTRUCTION .....	20
NETWORK ANALYSIS .....	20
NETWORK METRICS.....	20
CODE AVAILABILITY .....	21
<b>RESULTS.....</b>	<b>22</b>
DATA PREPROCESSING SUMMARY .....	22
ALPHA DIVERSITY ANALYSIS.....	22
BETA DIVERSITY ANALYSIS.....	24
DIFFERENTIAL ABUNDANCE TAXA ANALYSIS.....	26
NETWORK CONSTRUCTION AND ANALYSIS.....	28
DISCUSSION .....	36
<b>REFERENCES.....</b>	<b>40</b>

WORD COUNT: approx. 7700 words

## LIST OF FIGURES

Figure 1: Overview of the data preprocessing, analysis, and network construction workflow..	17
Figure 2: Shannon Diversity Index Comparison Across Disease Categories at Genus and Species Levels.....	23
Figure 3: PERMANOVA analysis of the Bray-Curtis distance. ....	25
Figure 4: Differential taxa abundance analysis at both genus and species level. ....	27
Figure 5: Comprehensive Network Visualisation – Gut Microbiome Disease Network (Genus level). ....	29
Figure 6: Comprehensive Network Visualisation – Gut Microbiome Disease Network (Species level). .....	31
Figure 7: Comprehensive Network Visualisation – Gut Dysbiosis Co-occurrence Network (Genus level). ....	33
Figure 8: Comprehensive Network Visualisation – Gut Dysbiosis Co-occurrence Network (Species level).....	35

## ABSTRACT

In recent years, there has been an increasing recognition of the gut microbiome's role in various diseases, including gastric and non-gastric conditions, as more studies reveal underlying dysbiotic pathology in many diseases. This study proposes a novel network-based approach for gut microbiome analysis, which utilises an association network extending the Barabási Human Disease Network model. Here, the association between diseases and taxa, derived from differential abundance analysis of taxa, was used to construct the network.

This study utilised relative abundance data at the genus and species levels from the GMrepo database. To establish baseline values, conventional methods were used to compute the alpha and beta diversity metrics between the disease group and healthy controls. While Alpha diversity demonstrated a slight variation between the disease and healthy groups, beta diversity analysis revealed significant differences between the disease and healthy groups, confirming that the diseases in this study group had considerable variation in their gut microbiome composition when compared with the healthy counterparts.

To identify taxa associated with diseases, differential abundance taxa analysis was done using a permutation-based Monte Carlo sampling technique.

Two bipartite networks were constructed at the genus and species levels using the differential abundance taxa analysis-derived associations between disease and taxon, which were subsequently projected into unipartite networks for disease and taxon at both genus and species levels.

Network analysis of the disease unipartite networks revealed a shared microbial pattern linking these diseases, suggesting an underlying microbial signature shared across multiple diseases. The taxa projection networks revealed the dysbiotic microbial signature.

This novel network approach offered a biologically validated system-level understanding of the gut microbiome in dysbiosis.

## **DECLARATION**

I declare that this project report is my original work unless referenced clearly to the contrary and that no portion of the work referred to in the report has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

## **INTELLECTUAL PROPERTY STATEMENT**

- I. The author of this dissertation (including any appendices and/or schedules to this dissertation owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.
- II. Copies of this dissertation, either in full or in extracts and whether in hard or electronic copy may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has entered. This page must form part of any such copies made.
- III. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the dissertation, for example graphs and tables (“Reproductions”), which may be described in this dissertation, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.
- IV. Further information on the conditions under which disclosure, publication and commercialization of this dissertation, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy, in any relevant Dissertation restriction declarations deposited in the University Library, and The University Library’s regulations.

## **ACKNOWLEDGEMENTS**

I want to express my deepest gratitude to those who have supported me while completing this dissertation. I would like to thank my supervisor, **Dr Jean Marc Schwartz**, for his guidance. His expertise and insights have significantly shaped the direction of my research. I would also like to thank **Dr Jose Nacher**, Japan for his guidance. I thank the Research IT team for using the Computational Shared Facility at the University of Manchester.

I am incredibly grateful to my family for their unwavering support, love, and understanding. Their belief in me made this possible.

Thank you all.

## INTRODUCTION

Gut microbiota refers to all the microorganisms that colonise the gut and exist in a symbiotic relationship that mutually benefits both the host and themselves. These microbes form an ecosystem within the gut environment known as a niche in which they interact with one another and with the host. The gut microbiome, on the other hand, refers to the collection of all genetic material from the microorganisms present in the intestinal tract, including genes, genomes, and their products.

## HUMAN GUT MICROBIOTA

The human gut is estimated to contain approximately 100 trillion microorganisms, primarily bacteria, which account for more than 90% of the total microbial population (Thursby and Juge 2017, Hou et al. 2022). Although the microbes colonise the entire gastrointestinal tract from the mouth to the colon, the colon has by far the most abundant microbial population, ranging up to  $10^{12}$  organisms per millimetre of intestinal tissue (Ley et al. 2006). This high density in the colon is attributed to the slower colonic transit time, favourable conditions for survival, a constant supply of a nutrition-rich environment, low oxygen concentration and acidity of the intestinal lumen (Thursby and Juge 2017).

## GUT MICROBIOTA AND HEALTH

Gut microbiota, as popularly believed, are not harmful pathogenic microorganisms in the gut. In contrast, they are a group of microbes that colonise the host and perform several functions that are beneficial to the host organism. These functions can be classified as metabolic, structural, immunomodulatory and systemic functions.

### Metabolic functions of the Gut Microbiota

The microbial community performs most of the essential metabolic functions, which benefit both the microbes and the host. These microbes digest the undigested fibres and complex carbohydrates that reach the colon by a process known as fermentation and produce short chain fatty acids (SCFA) such as butyrate, propionate and acetate—bacterial species such as *Bacteroides*, *Bifidobacterium*, *Faecalibacterium* and *Enterobacteria* are the primary sources of SCFAs in the gut (Musso, Gambino and Cassader 2010). Butyrate is the preferred energy source for the colonocytes while propionate and acetate play important roles in glucose and lipid metabolisms (Morrison and Preston 2016, Corrêa-Oliveira et al. 2016, Martínez-Ruiz, Robeson and Piccolo 2025). In addition to synthesising SCFAs, the gut microbes also produce vitamins, supplying the host with vitamin K, various B vitamins and folate (LeBlanc et al. 2013, Pham et al. 2021, Tarracchini et al. 2024).

Commensal microbes such as *Bacteroides intestinalis*, *Bacteroides fragilis* and *E.coli* convert the primary bile acid from liver into secondary bile acid which were shown to have antimicrobial property that limits the overgrowth of pathogenic gut microbes (Kurdi et al. 2006). Further, bile acids also support intestinal barrier protection, immune regulation and also serve as signalling molecules (Hang et al. 2019, Song et al. 2025). The microbial conversion of secondary bile further augments

the reabsorption through the enterohepatic circulation (Collins et al. 2023). Beyond the role of synthesising useful metabolites beneficial for the host, the gut microbes also play an important role in metabolising toxic compounds including drugs, heavy metals and toxins and limit their absorption. In addition, they also supplement xenobiotic metabolism by upregulating cytochrome P450 expression (Collins and Patterson 2020).

### **Structural and protective functions**

Microbial metabolites, such as butyrate, not only support the growth of colonocytes but also protect them by regulating their proliferation, differentiation, and apoptosis (Roediger 1980). Furthermore, commensal bacteria, such as *Lactobacillus*, were also shown to secrete protein metabolites that inhibit cytokine-induced colonocyte apoptosis, thereby protecting the epithelial barrier (Yan and Polk 2002). Another mechanism by which the normal gut microbiota protects the epithelial layer is by enhancing the production of mucin by the colonocytes. Commensal bacteria such as *Lactobacillus*, *Bifidobacterium*, *Bacteroides*, and *Akkermansia* enhance mucin which act as a barrier between the colonic epithelial layer and the intestinal lumen, protecting them from invasion by pathogenic microbes or their toxic metabolites. The mucin layer, in turn, also protects and supports the growth of these commensal microbes (Jandhyala et al. 2015, Thursby and Juge 2017, Wang et al. 2024). Additionally, *Akkermansia muciniphila* was also shown to reduce luminal endotoxemia by increasing endocannabinoid levels, which further reinforces the epithelial tight junctions, making it impermeable to bacterial lipopolysaccharides (LPS) (Everard et al. 2013).

The commensals also prevent the invasion of pathogenic organisms by competing with them for nutrients, triggering the host defence system and also by secreting antimicrobial compounds such as bacteriocin and certain organic acids (Khan, Petersen and Shekhar 2019).

### **Immune modulation functions**

The normal gut microbiota plays an important role in immune modulation, as it regulates both the intestinal and systemic immune systems. In the gut, normal microbiota regulates both the innate and adaptive immune system by promoting the development of gut-associated lymphoid tissue (GALT). Studies have shown that an underdeveloped gut immune system in germ-free mice could be restored by microbial polysaccharides, further substantiating the role of gut microbes in GALT development (Mazmanian et al. 2005). Furthermore, the gut microbes promote immune tolerance in dendritic cells and intestinal macrophages by stimulating the production of IL-10, which is anti-inflammatory in dendritic cells and inhibits the release of pro-inflammatory cytokines from intestinal macrophages upon stimulation by microbial Toll-like receptor (TLR) signals (Choden and Cohen 2022). In addition, the SCFAs influence the adaptive immune system by regulating the differentiation of CD4<sup>+</sup> helper cells, CD8<sup>+</sup> T cells, and B cells, including the production of immunoglobulin A (IgA) and also regulate the balance between anti-inflammatory Treg cells and pro-inflammatory Th17 cells (Wu, Tian and Wang 2018).

At the systemic level, gut commensals promote Th17 differentiation, which is both pro- and anti-inflammatory. Additionally, through the production of Butyrate, gut microbiota also regulates the balance between pro-inflammatory Th17 and anti-inflammatory Treg cells (Ohkubo et al. 1990, Wu, Tian and Wang 2018, Choden and Cohen 2022). Additionally, the gut microbiota was also shown to regulate the development of neutrophils, natural killer cells and mast cells outside the gut (Sanos et al. 2009, Zhang and Frenette 2019).

### **Other systemic functions**

In addition to the effects on the local intestinal environment, the gut microbiota also affects distant organs such as the brain, liver, skin and pancreas through well-defined communication routes known as gut–organ axis (Ramezani and Raj 2014, Llorente and Schnabl 2015, Mahmud et al. 2022).

The gut–brain axis is a bidirectional communication pathway between the gut and the brain, established through nerves such as the vagus nerve and chemical messengers, including short-chain fatty acids, serotonin, gamma-aminobutyric acid, and bile. Through this axis, the gut influences mood, cognition, and appetite, while the brain can in turn affect gut physiology, including microbial composition, gastric motility, and glucose metabolism (Farzi, Fröhlich and Holzer 2018). Moreover, studies have shown that normal gut microbiota is essential for the development of the enteric and central nervous system (Carabotti et al. 2015).

The gut also communicates with other organs such as the lungs, kidneys, skin, pancreas, and liver through similar axes (Ahlawat, Asha and Sharma 2021).

## GUT DYSBIOSIS

Dysbiosis is the imbalance in the gut microbiota caused by disruption in the microbial balance that results in the loss of normal beneficial functions carried out by the gut microbiota supporting the host's health. It is characterised by a loss of microbial diversity and dysregulation in microbial interactions within themselves and with the host, leading to the overgrowth of pathogenic organisms and the elimination of commensal microbes. As the dysbiotic microbiome is a complicated collection of pathogenic strains and loss of commensals, the dysbiosis-induced pathogenic effects in the host occur through several independent and overlapping mechanisms. A few key mechanisms are discussed below.

### Gut barrier dysfunction

The intestinal barrier is the first line of defence that separates the intestinal microenvironment from the gut lumen. It prevents the movement of pathogens or their harmful metabolites from entering the systemic circulation while allowing nutrients to pass through. While the normal gut microenvironment is in a state of balance that suppresses the pathogen overgrowth, in dysbiosis, this balance is lost, allowing pathogens such as *Escherichia coli*, *Clostridium difficile*, and *Salmonella* to colonise the gut (Zeng, Inohara and Nuñez 2017, Seekatz and Young 2014). The toxins or the harmful metabolites produced by these pathogens damage the intestinal barrier, causing them to enter the gut microenvironment. In addition, the LPS produced by gram-negative bacteria also stimulates the release of proinflammatory cytokines that further weaken the intestinal barrier, increasing its permeability (Dmytriv, Storey and Lushchak 2024).

Loss of barrier function is associated with Inflammatory Bowel disease (IBD), an inflammatory gastrointestinal disease in which pathogens and their metabolites enter the gut microenvironment, triggering chronic inflammation. Notably, Barnich et al. (2025) demonstrated that adherent-invasive *E. coli* invade the intestinal mucosa in Crohn's disease, causing inflammation. A similar mechanism has also been described in Irritable Bowel syndrome with Shiga-like toxin-producing *E. coli* (Andresen et al. 2016). Moreover, in colorectal cancer, *Fusobacterium nucleatum*, another pathogenic strain, disrupts the epithelial barrier by degrading the tight junction proteins (Ou et al. 2022).

### Dysbiosis-induced immune dysregulation

Dysbiosis-induced immune dysregulation starts with a disrupted intestinal barrier, which allows microbial products like LPS to leak into the gut microenvironment. LPS-induced activation of Toll-like receptors (TLR4) triggers the release of proinflammatory cytokines like TNF- $\alpha$ , IL-1 $\beta$ , and IL-6, which in turn activate the adaptive immune system, causing gut inflammation (Rhee 2014). Moreover, dysbiosis-induced microbial imbalance also leads to loss of commensal bacterium such as *Faecalibacterium prausnitzii*, which is a major butyrate producer that leads to loss of inhibitory anti-inflammatory signals from Treg cells that depend on butyrate (Mousa, Chehadeh and Husband 2022). Additionally, there is an increased production of trimethylamine-N-oxide (TMAO) and bile

acids by pathogenic strains, which activates the TMAO–Bile Acid–Innate Immunity Axis, thereby amplifying the proinflammatory response (Wang et al. 2025).

### **Systemic effects of dysbiosis**

Healthy gut produces several metabolites that extend their beneficial effects beyond the intestines, affecting overall health. The key metabolites include SCFAs, bile acids, tryptophan metabolites and TMAO.

In dysbiosis, loss of microbial balance results in lower circulating levels of SCFAs such as butyrate due to reduction in commensal population especially *Faecalibacterium prausnitzii*. Reduced butyrate leads to impaired barrier integrity of the mucosal wall along with dysregulation in the anti-inflammatory Treg cells leading to a state of chronic systemic inflammation. This interferes with insulin signalling leading to insulin resistance, metabolic syndrome and type 2 diabetes. Moreover, impairment in lipid metabolism also interferes with lipid metabolism pathways causing obesity (He et al. 2020, Yu et al. 2025).

In the gut, the bile acid metabolism is a tightly controlled process aimed at maintaining the levels of primary and secondary bile. Dysbiosis-induced disruption leads to accumulation of toxic secondary bile in the gut which irritates the gut lining leading to impaired gut motility, damage to the intestinal barrier and dysregulated lipid metabolism. These process eventually results in the development of chronic systemic inflammation, insulin resistance, type 2 diabetes and even non-alcoholic fatty liver disease (Wang et al. 2019, Mancera-Hurtado et al. 2023). Moreover, high levels of secondary bile in the gut can also have carcinogenic effect within the gut (Bernstein and Bernstein 2023).

Tryptophan metabolism is another example on the role of gut dysbiosis on systemic health. Dysbiosis-induced impairment in tryptophan metabolism results in impaired indole production, a microbial metabolite useful in the maintenance of intestinal barrier integrity. Accumulation of tryptophan due to inadequate conversion into indole may result in the activation of kynureneine pathways which causes neuroinflammation leading to depression or neuro degenerative diseases (Gao et al. 2020).

Finally, dysbiosis-induced increased production of trimethylamine in the gut gets converted to TMAO in the liver. Elevated levels of TMAO has been shown to increase the levels of pro-inflammatory cytokines causing systemic inflammation (Wang et al. 2025). Moreover, studies have also shown that TMAO is atherogenic causing endothelial dysfunction and plaque formation thereby increasing cardiovascular risks (Zhu, Li and Jiang 2020).

## NETWORK ANALYSIS IN GUT MICROBIOME STUDIES

The gut microbiome is a complex and dynamic community of microbes, not a random gathering of different microorganisms, and it continuously interacts with other microbes and with the host microenvironment. These interactions are essential for maintaining a balanced and healthy gut, as well as the normal functioning of the gut microbiota (Coyte, Schluter and Foster 2015). Conventional methods in gut microbiome studies use statistical methods such as differential abundance taxa analysis and diversity analysis. While these methods provided information on the composition of the microbiome at the individual microbe level, they cannot comprehend the gut microbiome as a whole or explain the microbial interactions (Faust 2021).

Network-based approaches offer researchers a system-level understanding of how the microbes within the microbiota interact with one another and with their host environment. In network analysis, the gut microbiome is represented as a network, where microbes are mapped as nodes and their interactions are represented as edges. This allows a complete view of the gut microbiome, which reflects how microbes interact with one another and how changes in one component of the microbiome can affect the entire ecosystem. Here, methods such as statistical correlation, mutual information, association, and ecological inference are used to estimate microbial associations and interactions, depending on the type of network (Matchado et al. 2021).

In addition to representing the gut microbiome in a graphical format, network analysis also provides system-level insights. Some networks may show organised clusters, a property known as modularity or community in network terms. These clusters may represent an ecological niche or a specialised group of microbes within the microbiome (Kajihara and Hynson 2024). In addition to this, the network also provides other metrics such as connectivity which is the number of connections between the nodes in the network; key players in the network, hub nodes – those that are highly connected to the neighbours (degree centrality), bridging nodes – nodes with high betweenness centrality and has the shortest path between clusters and peripheral nodes – sparsely connected nodes (Kajihara and Hynson 2024).

Currently, four basic types of networks are mainly used in gut microbiome research and researchers often modify these approaches based on their study requirements.

1. Correlation Network – Edges based on statistical correlation analysis to identify significant correlation in abundance between microbes (e.g.: Pearson's, Spearman's or SparCC).
2. Co-occurrence Network – Edges based on microbial co-occurrences in the same sample.
3. Association Network – Edges based on statistical or machine learning models such as random forest to detect association between microbes.
4. Functional Network: Edges based on shared metabolic pathways, microbial metabolite or gene functions.

## **STUDY RATIONALE**

The current methods in network analysis approach used in gut microbiome studies rely heavily on data-driven methods such as correlation networks and machine learning-based models (Jayakrishnan et al. 2024, Rozera et al. 2025). These networks are built using statistical associations and correlation patterns derived from numerical data without including biological relevance. While these approaches have their benefits, they may not always provide meaningful insights about the gut microbiome (Friedman and Alm 2012, Papoutsoglou et al. 2023, Kajihara and Hynson 2024).

This study presents a novel method for analysing the gut microbiome, combining an association-based network analysis with differential abundance taxa analysis. Here, the disease-taxon relationship was derived from a differential abundance of taxa analysis, comparing them with healthy controls, which was used to construct a bipartite disease-taxon network. This was subsequently projected to a disease-disease and taxon-taxon unipartite network.

The network model used in this study was based on Barabási's Human Disease Network, which showed that mapping diseases through shared genes may reveal disease patterns not perceived by studying them in isolated conditions (Goh et al. 2007).

## **AIMS AND OBJECTIVES**

The aim of this project is to investigate the role of the gut microbiome in human diseases using an association-based network combined with differential abundance taxa analysis.

### **Objectives:**

1. To analyse the difference in gut microbial diversity between disease and health.
2. To analyse the difference in the microbial community composition between disease and health.
3. To identify taxa that are significantly associated with diseases using differential abundance taxa analysis.
4. Construct association-based bipartite disease-taxon network and project it into disease-disease and taxon-taxon unipartite networks.
5. To analyse the network by computing network metrics such as density, average clustering co-efficient, degree, closeness and betweenness centrality and to identify hub taxa, key nodes and shared microbial pattern or co-occurrence signatures.
6. To use higher-level disease or taxonomic classification on respective networks and assess for patterns or community formation.

## MATERIALS AND METHODS

All data preprocessing and analyses were performed in Python (v3.10.18) utilising standard libraries within a Jupyter Notebook environment in Microsoft Visual Studio Code. Network construction, visualisation and analysis were done using the NetworkX package (v3.4.2) and ipycytoscape (v1.3.3).

### GMREPO DATA

Data was obtained from the GMrepo repository ([www.gmrepo.hmangut.info/home](http://www.gmrepo.hmangut.info/home)), a database of curated and annotated human gut metagenomes (Wu et al. 2020). Downloaded files included the relative abundance table at both the genus and species levels, the MeSH table (Medical Subject Headings, National Library of Medicine), which contained standardised disease names, the NCBI taxonomy table, which contained taxon names, and the metadata file, which contained sample information and sequencing details.

### DATA PREPROCESSING

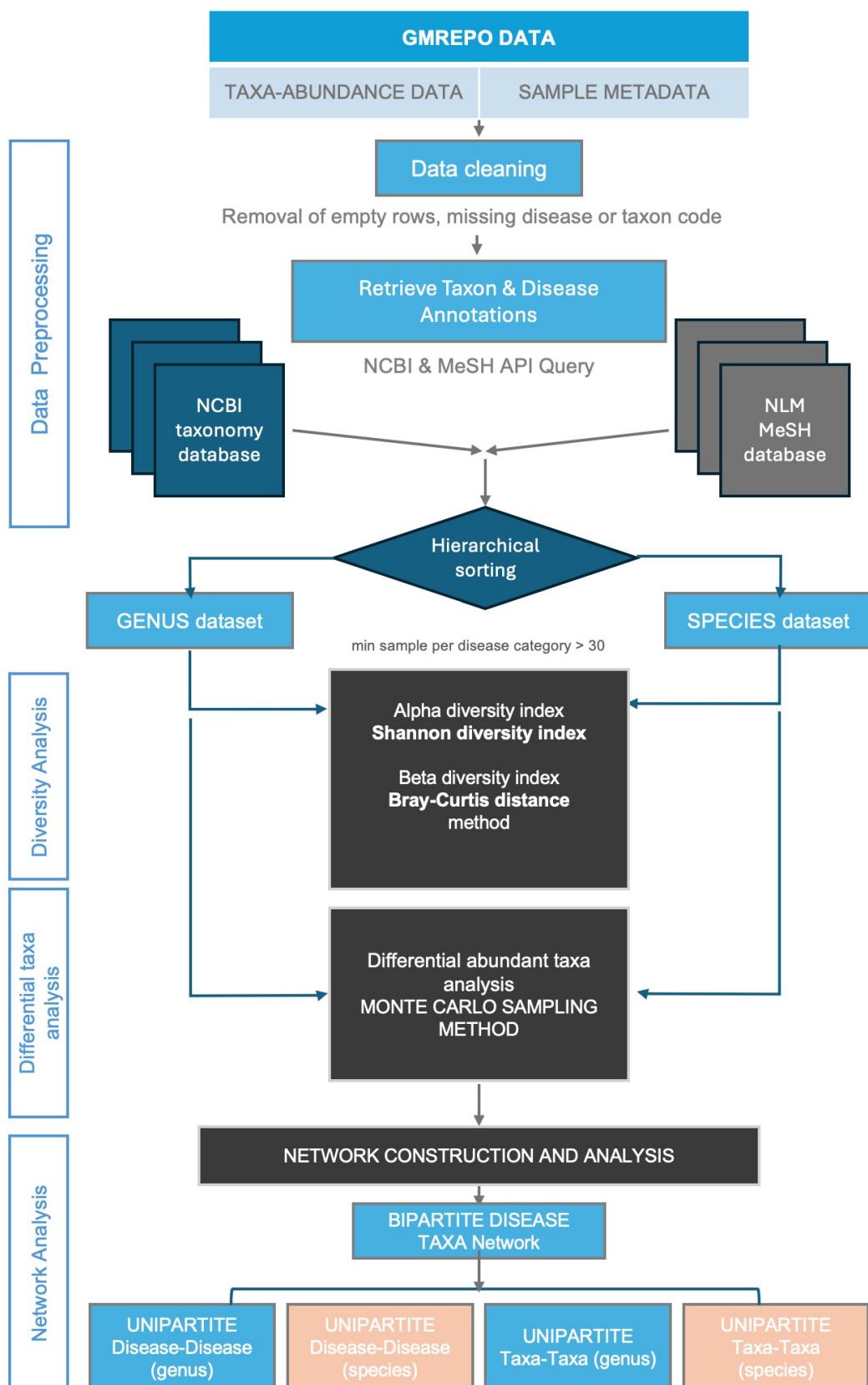
Data preprocessing was performed in stages using the Pandas and NumPy libraries. Rows were checked for duplicates, but none were found. Rows with missing entries and those that had taxa ID labelled as '-1' (unknown or unclassified taxa) were entirely removed. Post-processing, the disease codes were mapped to the sample IDs using the accession IDs in the sample metadata.

The dataset was then divided into genus and species-level abundance tables. Additionally, the dataset was filtered to exclude diseases with low sample count (<30 samples) as they limited statistical reliability.

### STANDARDISATION OF DISEASE AND TAXONOMY NAMES

The disease names in the MeSH table were found to be inconsistent. For example, the terms 'health', 'healthy', and 'normal' were used to represent healthy controls, while disease names such as 'Ulcerative colitis' were also represented as 'Colitis, Ulcerative'. To standardise disease names, MeSH terms, along with MeSH disease category terms, were directly downloaded from the NLM MeSH database (<https://www.ncbi.nlm.nih.gov/mesh>) using a custom Python script.

Similarly, taxon names along with phylum names were also directly downloaded from the NCBI taxonomy database (<https://www.ncbi.nlm.nih.gov/taxonomy>) using a custom Python script. For taxon IDs that have been deprecated in the NCBI database, the UniProt database was used. The downloaded name tables were stored as separate Python dictionaries for use in downstream analyses. A summary of the workflow is shown in Fig.1.



**Figure 1: Overview of the data preprocessing, analysis, and network construction workflow.**

GMRepo: Gut Microbiome Repository; NLM - national library of medicine, NCBI - National Centre for Biotechnology Information

## ALPHA DIVERSITY ANALYSIS

Alpha diversity (within-sample diversity) of the samples was computed using the Shannon diversity index, which measures the richness (number of taxa) and evenness (distribution of abundance) of the taxa. It was calculated using the formula:

$$\text{Shannon index} = - \sum_{i=1}^S p_i \cdot \ln(p_i)$$

where 'S' is the total number of taxa and 'pi' is the proportional relative abundance of the 'ith' taxon in the sample. Shannon index values range between 0 and  $\infty$  with higher Shannon index indicating a more diverse community, while lower values mean reduced diversity.

The Shannon index was calculated separately for each sample in the dataset at both the genus and species levels. The difference between the healthy and disease groups was tested by comparing the median Shannon index values using the Wilcoxon rank sum test, which was followed by Dunn's post hoc test for pairwise comparisons between the healthy and each disease group. Both the healthy and individual disease groups were randomly subsampled to match the size of the smallest group ( $n = 30$ ). Correction for multiple testing was done using the Benjamini–Hochberg false discovery rate (FDR) method, with corrected p-value  $\leq 0.05$  considered as statistically significant.

## BETA DIVERSITY ANALYSIS

Beta diversity (between-sample dissimilarity) was computed using Bray-Curtis (BC) distance, which is a measure of the dissimilarity between samples based on the presence and relative abundance of taxa. BC distance was calculated using the following formula:

$$\text{Bray – Curtis distance (BC)} = \frac{\sum_{i=1}^n |x_i - y_i|}{\sum_{i=1}^n |x_i + y_i|}$$

where  $x_i$  and  $y_i$  are the relative abundances of the ith taxon in sample X and sample Y, respectively, and 'n' is the total number of taxa in both samples. BC distance value ranges from 0, indicating identical composition, to 1, meaning entirely dissimilar.

Differences in the BC distance between the healthy and disease groups were tested statistically using permutation multivariate analysis of variance (PERMANOVA). Unequal sample sizes between the healthy and disease groups were handled by randomly subsampling without replacement to match the size of the smallest group ( $n=30$ ). The pseudo-F statistic (F) was then computed as the ratio of between-group to within-group variation; higher values indicating a greater strength of separation between groups (Anderson 2001). The pseudo-F was calculated using the formula:

$$F = \frac{\text{between group variation}}{\text{within group variation}}$$

Statistical significance between was calculated by permutation testing (no. of permutations 999), and the p-value was calculated using the formula:

$$p = \frac{\#\{F_{perm} \geq F_{obs}\} + 1}{no.\ of\ permutations + 1}$$

where  $F_{obs}$  is the observed pseudo-F statistic,  $F_{perm}$  are the pseudo-F values generated under the null hypothesis by randomly permuting the group labels and the numerator  $\#\{F_{perm} \geq F_{obs}\}$  denotes the number of times the condition was true. Correction for multiple testing was done using the Benjamini–Hochberg false discovery rate (FDR) method, with corrected p-value  $\leq 0.05$  considered as statistically significant.

## DIFFERENTIAL ABUNDANCE TAXA ANALYSIS

To identify the taxa that were significantly different in the disease group when compared with the healthy, a differential abundance taxa analysis was performed. Prior to analysis, an abundance filter (relative abundance  $>1\%$ ) and a prevalence filter (present in  $>10\%$  of the samples) was applied to remove rare and less prevalent taxa within the healthy and separately for each disease group.

Differentially abundant taxa were identified in each disease group by comparing the relative abundance values of each taxon in the disease group with that of healthy control samples using a non-parametric Monte Carlo permutation sampling method. Unequal sample sizes between the healthy and each disease group was handled by randomly subsampling without replacement to match the size of the smallest group. For taxa absent in the healthy group but present in disease samples, a relative abundance of zero was assigned to the healthy group. In contrast, if the taxa was present in the healthy but absent in the disease group, they were entirely excluded from analysis. A total of 999 permutations ( $N=999$ ) were performed for each taxon for every disease–healthy group combination. The permutation p-value was computed using the formula:

$$p = \frac{1 + \sum_{i=1}^N I(|\Delta_i| > |\Delta_{obs}|)}{N + 1}$$

where  $I()$  is the indicator function (equals to one if the condition is true, else 0),  $\Delta_{obs}$  is the observed difference in the mean relative abundance of the 'ith' taxon between the healthy and disease group, and  $\Delta_i$  is the difference in mean relative abundance of the 'ith' taxon obtained by permuting the group labels. Correction for multiple testing was done using the Benjamini–Hochberg false discovery rate (FDR) method, with corrected p-value  $\leq 0.05$  considered as statistically significant.

An upper ubiquity filter was applied on the taxa that were identified by the differential abundance taxa analysis to exclude taxa that were associated with more than 30 diseases.

## NETWORK CONSTRUCTION

Two bipartite networks (genus and species) were constructed using an association-based network model, in which diseases and taxa were linked based on biological associations derived from the differential abundance of taxa analysis. Here, diseases and taxa were represented as nodes and the edges represented the association between them. Each bipartite network (genus and species level) was then projected into two unipartite networks, where nodes of the same type (diseases or taxa) were connected based on the number of shared neighbours. The unipartite projections were:

1. Gut Microbiome Disease Network (GMDN): a disease–disease projection when two disease nodes were connected if they shared at least one taxon.
2. Gut Dysbiosis Co-occurrence Network (GDCN): a taxon–taxon projection when two taxon nodes were connected if they co-occurred in at least one disease.

## NETWORK ANALYSIS

Prior to network analysis, edges were thresholded based on their edge weights to remove weaker links. To determine the most appropriate edge-weight threshold, different edge weights were tested based on their percentile ranges (50<sup>th</sup> to 99<sup>th</sup> percentile). At each threshold ( $\theta$ ), basic network metrics, including node and edge counts and the network's overall connectivity was evaluated.

In addition, degree distribution analysis of the nodes was also performed to assess for scale-free distribution. This analysis was done in Python using the power-law package (v1.4.4), which fits the degree distribution of the nodes to the power-law model,  $P(k) \sim k^{-\alpha}$ , where  $k$  represented the degree of the node and  $\alpha$  is the scaling component ( $2 < \alpha \leq 3$ ; classic power-law fit) (Barabasi and Albert 1999). Goodness of fit between the empirical and the fitted power-law model was measured using the Kolmogorov–Smirnov (KS) statistic and Monte Carlo resampling, with a p-value greater than 0.1 as a theoretical power-law fit.

For visual confirmation of scale-free distribution, the complementary cumulative distribution function (CCDF) of the degree distribution was plotted on a log-log scale along with the fitted power-law for  $k \geq x_{\min}$ , where  $x_{\min}$  is the minimum degree which follows the power-law.

As higher threshold resulted in network fragmentation, lower threshold retained several edges and no threshold resulted in a power-law fit in all four networks, the networks were thresholded to retain the maximum number of nodes in the disease-disease projection, and to reduce overall network density and improve readability in the taxon-taxon projection.

## NETWORK METRICS

Using NetworkX, the following network metrics were computed: node and edge counts, network density, average clustering coefficient, network diameter (the shortest longest path between two nodes), node degrees, and centrality, including degree, closeness, and betweenness centrality.

## **CODE AVAILABILITY**

All python scripts used for data preprocessing, statistical analyses, network construction and analysis are available at [https://github.com/igurubalan/microbiome\\_project](https://github.com/igurubalan/microbiome_project).

OpenAI 2025 (ChatGPT4) was used for coding debugging.

## **RESULTS**

### **DATA PREPROCESSING SUMMARY**

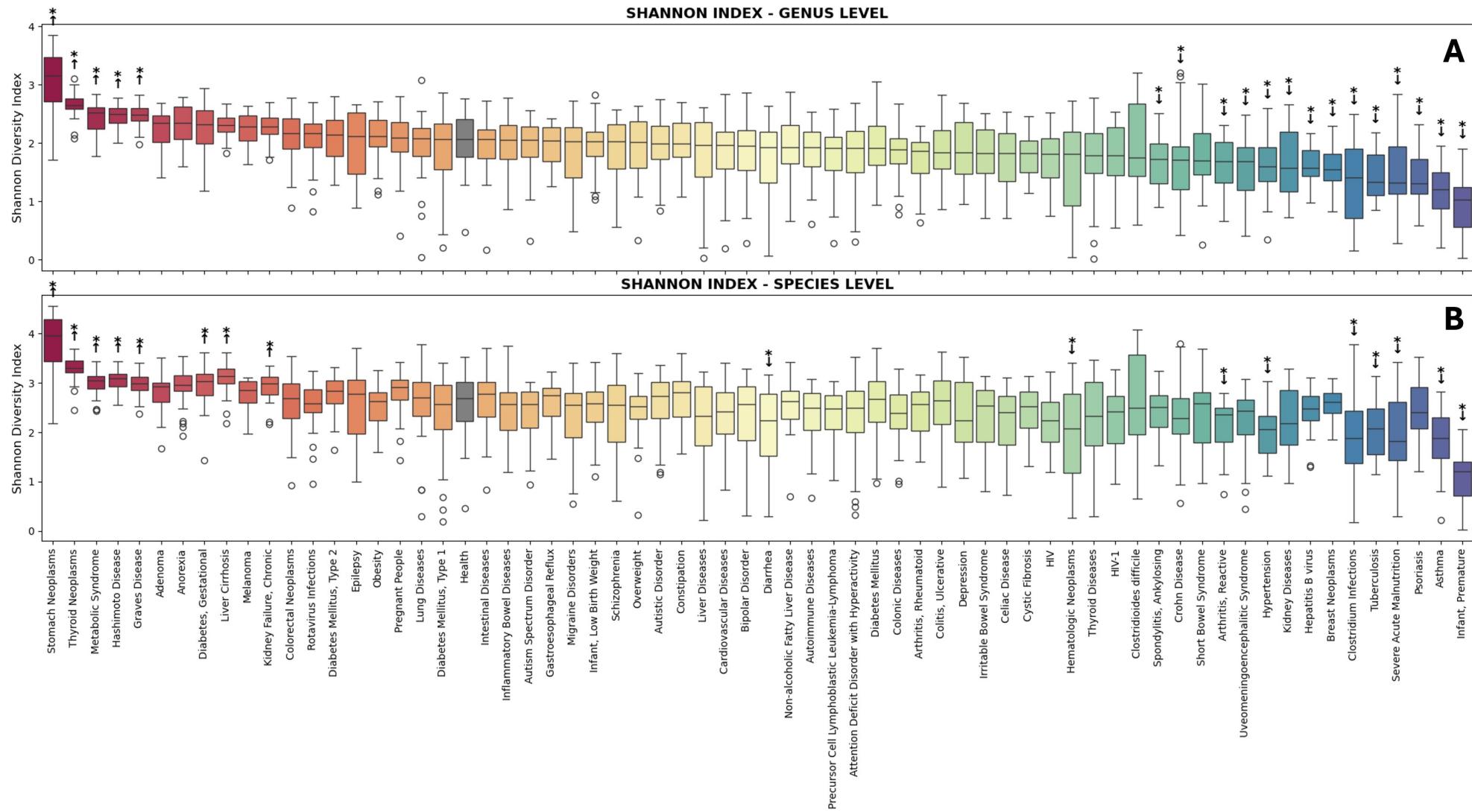
The original unprocessed dataset consisted of 32,244 samples, which included 93 diseases and contained 8,910 individual taxa at both genus and species ranks. Following data processing, which included de-duplication of rows, removal of rows with missing values, and removal of samples with missing disease or taxon names, a total of 26,834 samples were retained. Of these, 12,485 samples were healthy, and 14,389 were diseased, and contained 85 different disease groups. This filtered dataset consisted of 1,711 genera and 6,190 species. Furthermore, after applying a sample size threshold to remove disease groups with fewer than 30 samples, an additional 20 disease groups were withdrawn from the dataset, resulting in a total of 65 disease groups in the final dataset.

### **ALPHA DIVERSITY ANALYSIS**

Alpha diversity analysis revealed that most disease groups in this study showed no significant differences in their median Shannon index values when compared with the healthy control group at both the genus and species levels.

At the genus level, an increase in Shannon index value was observed in diseases such as stomach and thyroid neoplasms, metabolic syndrome, Hashimoto's thyroiditis, and Graves' disease, suggesting a gain in microbial diversity. In contrast, diseases such as premature birth, Clostridium infections, certain cancers, and a few other conditions were associated with lower values, suggesting a loss of microbial diversity (Fig. 2A).

Species-level results also showed a similar pattern to the genus-level. In addition, higher Shannon index levels were also observed in gestational diabetes, liver cirrhosis and chronic kidney failure at the species level. In contrast, diseases such as ankylosing spondylitis, Crohn's disease, uveo-meningo-encephalitis syndrome, kidney disease, hepatitis B and breast neoplasm, while exhibiting a significant reduction in Shannon index at the genus level, showed no significant difference at the species level (Fig. 2B).

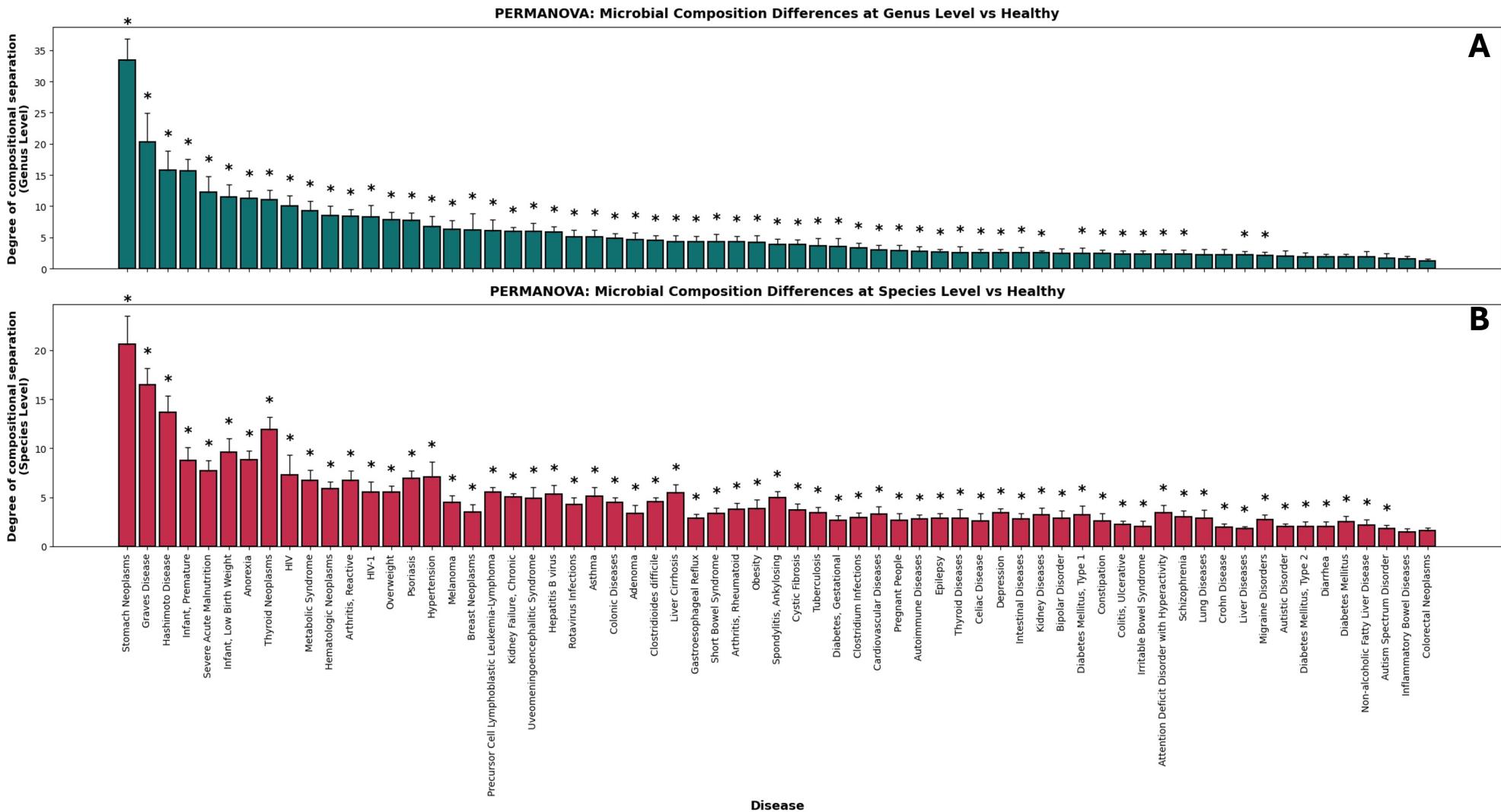


**Figure 2: Shannon Diversity Index Comparison Across Disease Categories at Genus and Species Levels.**

Graphs show the changes in the median Shannon index across disease groups, including the healthy (grey box). The top panel (A) shows the Shannon index at the genus level, while the bottom panel (B) shows it at the species level. Statistical significance, denoted by \*, represents pairwise comparisons between individual disease groups and healthy controls (FDR-adjusted p-value  $\leq 0.05$ ). Disease groups positioned to the left of "Healthy" generally show reduced diversity. In contrast, those to the right show increased diversity, with arrows indicating the direction of change relative to the healthy reference.

## BETA DIVERSITY ANALYSIS

At the genus level, PERMANOVA analysis revealed significant differences in pseudo-F statistic (the degree of compositional separation) for 55 out of 65 disease groups when compared with healthy controls. While diseases such as stomach neoplasm and Graves' disease exhibited the highest pseudo-F values, indicating a substantial shift in their microbial composition, diseases such as liver diseases and migraine showed minor but statistically significant difference (Fig. 3A). Species level analysis also revealed similar trends, in which nearly all disease groups, except autism spectrum disorder and colorectal neoplasm showed significant differences when compared with healthy control samples (Fig. 3B).

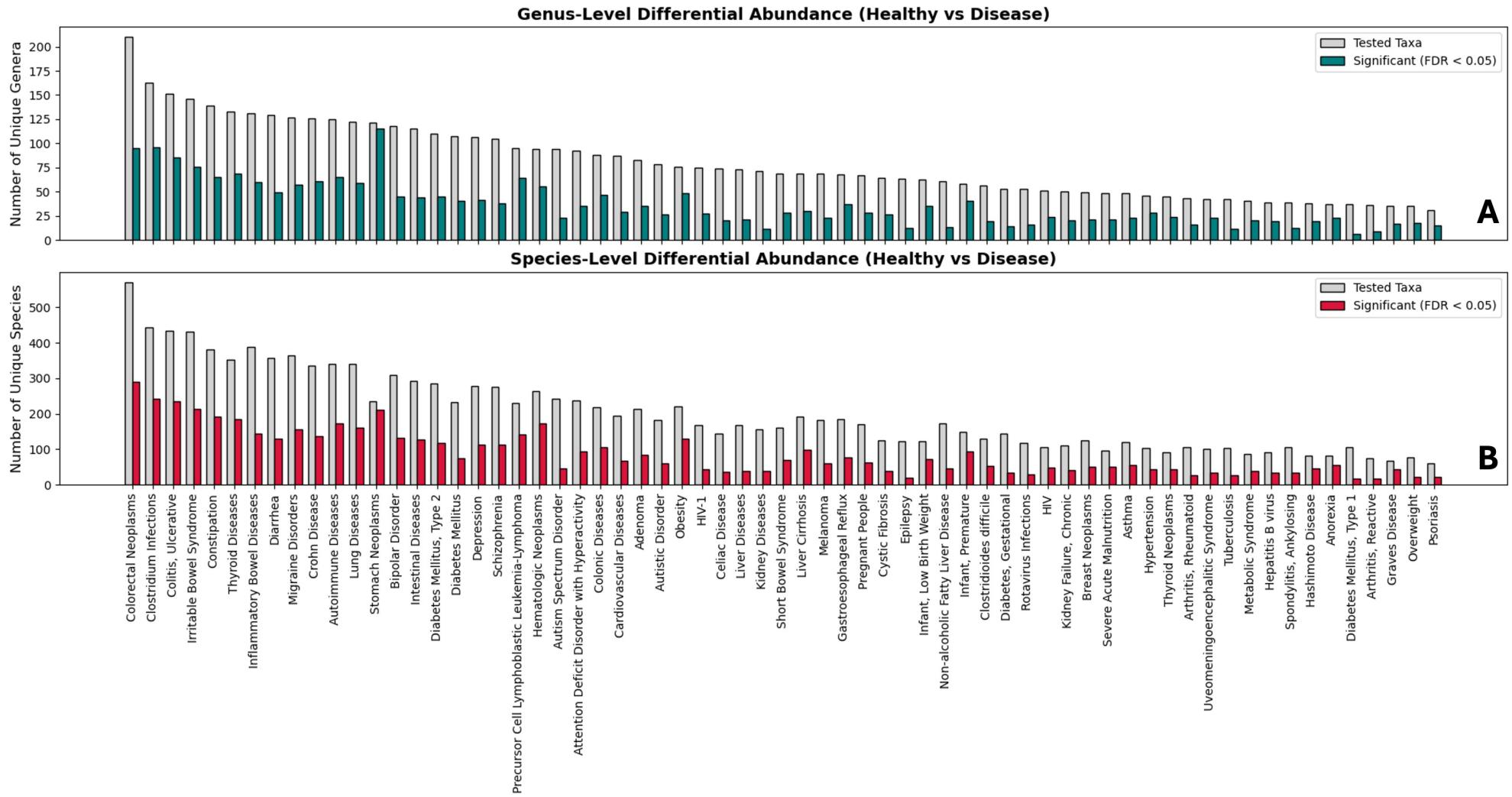


**Figure 3: PERMANOVA analysis of the Bray-Curtis distance.**

Bar plot represents the mean pseudo-F statistic ( $\pm$  standard deviation) for each disease category, calculated from PERMANOVA tests. Pseudo-F (degree of compositional separation) reflects compositional variation between a given disease group and the healthy control group. Since the healthy group served as the baseline for these comparisons, it is not represented as a separate bar in the plot. Individual bars are not compared against each other but instead represent the results of independent tests comparing each disease group with the healthy control. \* Statistical significance was defined as an adjusted  $p$ -value  $\leq 0.05$ .

## DIFFERENTIAL ABUNDANCE TAXA ANALYSIS

The results of the differential abundance taxa analysis are summarised in Fig. 4. At the genus dataset level, a total of 520 genera passed the prevalence and relative abundance filter, of which 387 genera showed a significant difference in their relative abundance values when compared with healthy controls. Similarly, at the species dataset level, 1,547 species passed the filtering process, of which 1,150 showed a significant difference in their relative abundance values when compared with healthy controls. Figs. 4A and 4B represent individual disease-level changes at both the genus and species levels, respectively.



**Figure 4: Differential taxa abundance analysis at both genus and species level.**

Bar plots show the total number of unique taxa tested for each disease (grey) and the subset identified as statistically significant by Monte Carlo permutation-based differential abundance testing (coloured; adjusted p-value  $\leq 0.05$ ).

## NETWORK CONSTRUCTION AND ANALYSIS

### Gut Microbiome Disease Network – Genus level Network Properties

The genus-level GMDN comprised of 65 disease nodes and 1,930 edges. An edge-weight threshold at the 80<sup>th</sup> percentile ( $\theta = 9$ ) reduced the network from 65 disease nodes to 51 nodes and 413 edges beyond which the network fragmented into smaller disconnected components. This resulted in a moderately dense (density 0.32) and a compact structure (average clustering coefficient 0.77), which formed a single connected component with no modular community formation (Fig. 5B).

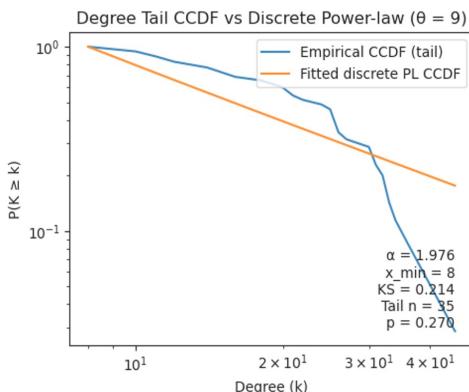
The degree distribution analysis of the GMDN, although, revealed a theoretical scale-free-like pattern with a scaling component ( $\alpha$ ) of 1.96 with a minimum node degree of 8 based on the Kolmogorov–Smirnov statistic (KS = 0.214; p-value=0.27), the complementary cumulative distribution function failed to align with the power-law model indicating that the network was composed of several high degree nodes rather than a single dominant superhub node (Fig. 5A).

The most central and well-connected node in the GMDN at the genus level was Irritable Bowel Syndrome, with a degree centrality of 0.9 and a closeness centrality of 0.9. However, the network also had other high-degree and well-connected nodes, such as those related to clostridial infection (degree centrality 0.72; closeness centrality 0.78), ulcerative colitis (degree centrality 0.72; closeness centrality 0.78), autoimmune diseases (degree centrality 0.68; closeness centrality 0.75), and lung disease (degree centrality 0.66; closeness centrality 0.73). The betweenness centrality in the GMDN was uniformly low throughout the network (mean 0.014; median 0.001), indicating that many alternative paths connected the nodes and no single disease acted as a bottleneck (Fig. 5B).

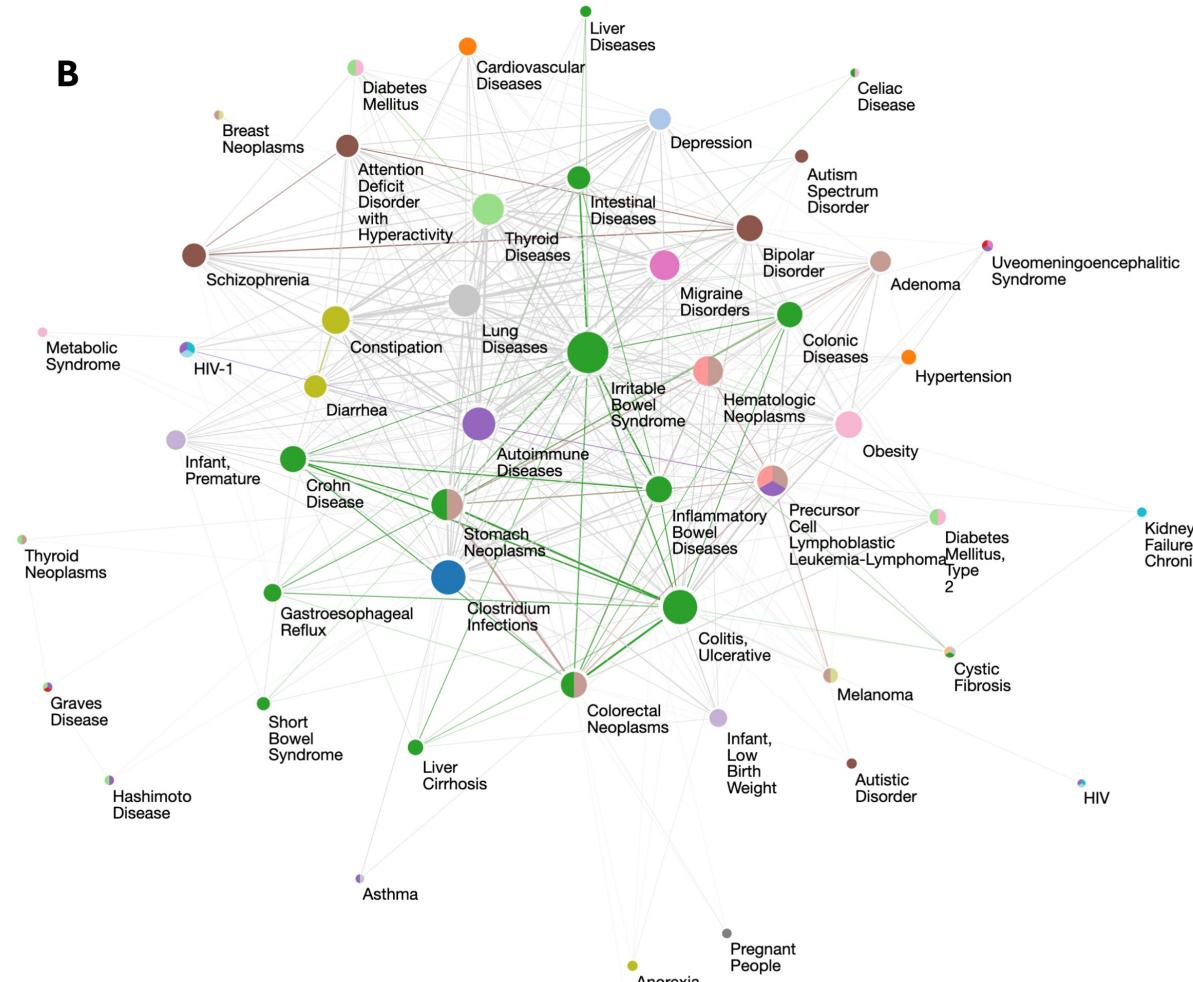
Colour-coding the nodes based on disease categories revealed several disease nodes within the central hub that predominantly belonged to digestive system diseases, such as Irritable bowel syndrome, Ulcerative colitis, Inflammatory Bowel Disease, Crohn's disease, and stomach neoplasm, along with hub nodes from other disease categories, including bacterial infections and immune system diseases.

# COMPREHENSIVE NETWORK VISUALISATION GUT MICROBIOME DISEASE NETWORK - GENUS LEVEL

**A**



**B**



**Figure 5: Comprehensive Network Visualisation – Gut Microbiome Disease Network (Genus level).**

**A.** GMDN degree distribution plot showing the empirical complementary cumulative distribution function (CCDF) (blue) representing the degree distribution of the nodes in a log-log plot and power-law CCDF (PL CCDF) (orange line). **B.** GMDN unipartite network. Nodes represent disease groups, and the edges represent the shared taxa. Edge thickness is proportional to the number of shared taxa (weighted edges). Disease nodes are colour-coded based on broad disease categories derived from the MeSH classification

### **Gut Microbiome Disease Network – Species level: Network Properties**

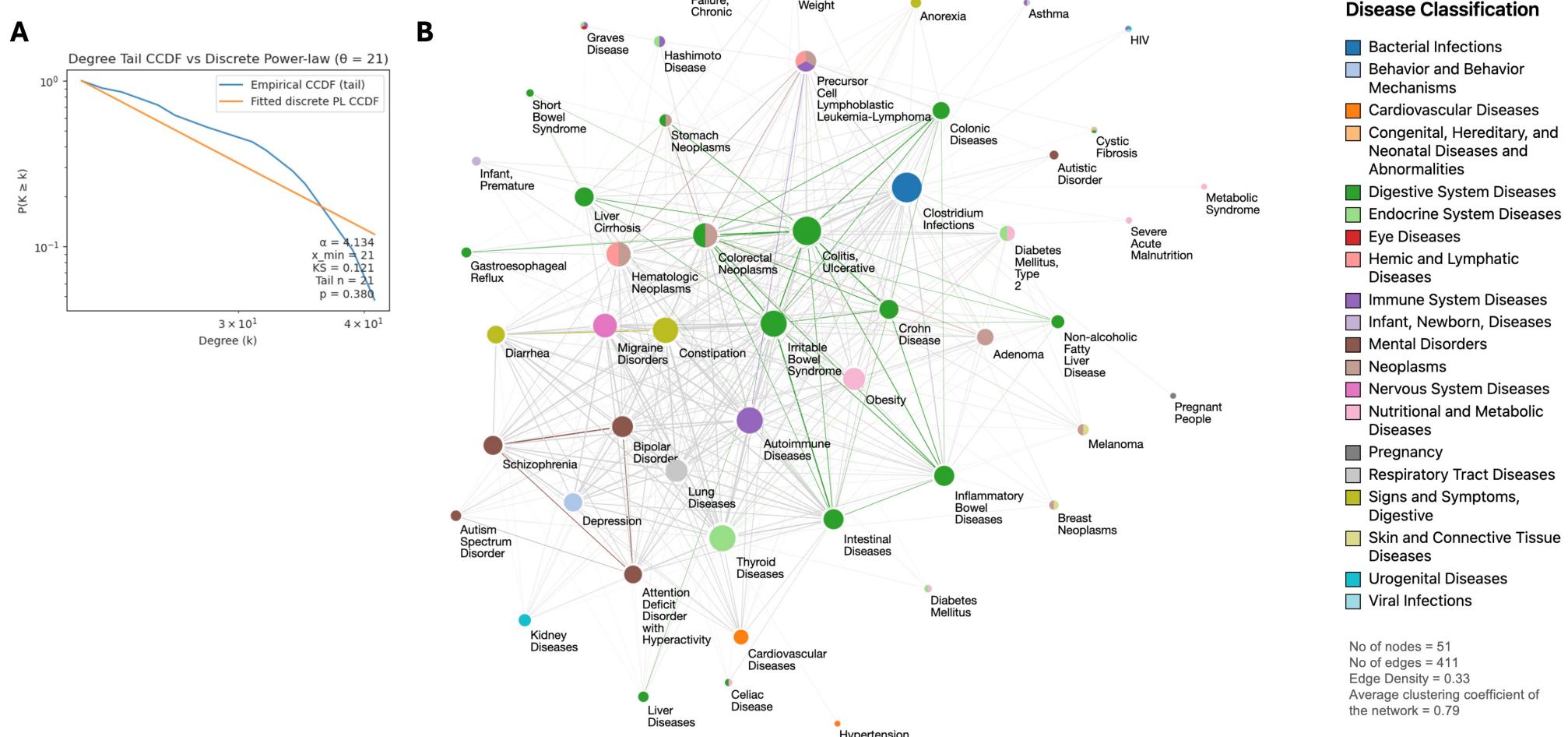
The GMDN at the species level exhibited a network architecture similar to that of the genus-level GMDN. The unthresholded species-level GMDN had 65 disease nodes and 2,041 edges. At the species level, the edge-weight threshold at the 80<sup>th</sup> percentile corresponded to an absolute edge-weight ( $\theta$ ) of 21 due to the higher number of unique species in the dataset, which resulted in 51 nodes and 411 edges. Like the genus-level GMDN, the species-level also resulted in a moderately dense (density 0.33) and compact structure (average clustering coefficient 0.79), forming a single connected component with no modular community formation (Fig. 6B).

The degree distribution analysis of the GMDN at the species level, although, revealed a theoretical scale-free-like pattern with a scaling component ( $\alpha$ ) of 4.13 with a minimum node degree of 21 based on the KS statistic (KS = 0.124; p-value=0.38), the CCDF failed to align with the power-law model indicating that the network was composed of several high degree nodes (Fig. 6A).

The most central and well-connected node at the species level in the GMDN was Clostridial infection with a degree centrality of 0.83 and closeness centrality of 0.85. The network also included other high-degree and well-connected nodes, such as those related to ulcerative colitis (degree centrality 0.72; closeness centrality 0.78), thyroid diseases (degree centrality 0.71; closeness centrality 0.77), Irritable bowel syndrome, (degree centrality 0.71; closeness centrality 0.77), and autoimmune diseases (degree centrality 0.71; closeness centrality 0.77), which formed the central core of the GMDN. The betweenness centrality in the GMDN was uniformly low throughout the network (mean 0.014; median 0.0009), indicating that many alternative paths connected the nodes and no single disease acted as a bottleneck (Fig. 6B).

Like its genus-level counterpart, colour coding at the species level revealed a similar network structure with central hub nodes predominated by digestive system diseases, such as Irritable bowel syndrome, Ulcerative colitis, Inflammatory Bowel Disease, Crohn's disease, and stomach neoplasm, along with hub nodes from other disease categories, including bacterial infections and immune system diseases.

# COMPREHENSIVE NETWORK VISUALISATION GUT MICROBIOME DISEASE NETWORK - SPECIES LEVEL



**Figure 6: Comprehensive Network Visualisation – Gut Microbiome Disease Network (Species level).**

**A.** GMDN degree distribution plot showing the empirical CCDF (blue) representing the degree distribution of the nodes in a log-log plot and power-law CCDF (PL CCDF) (orange line). **B.** GMDN unipartite network. Nodes represent disease groups, and the edges represent the shared taxa. Edge thickness is proportional to the number of shared taxa (weighted edges). Disease nodes are colour-coded based on broad disease categories derived from the MeSH classification.

### Gut Dysbiosis Co-occurrence Network - Genus Level: Network Properties

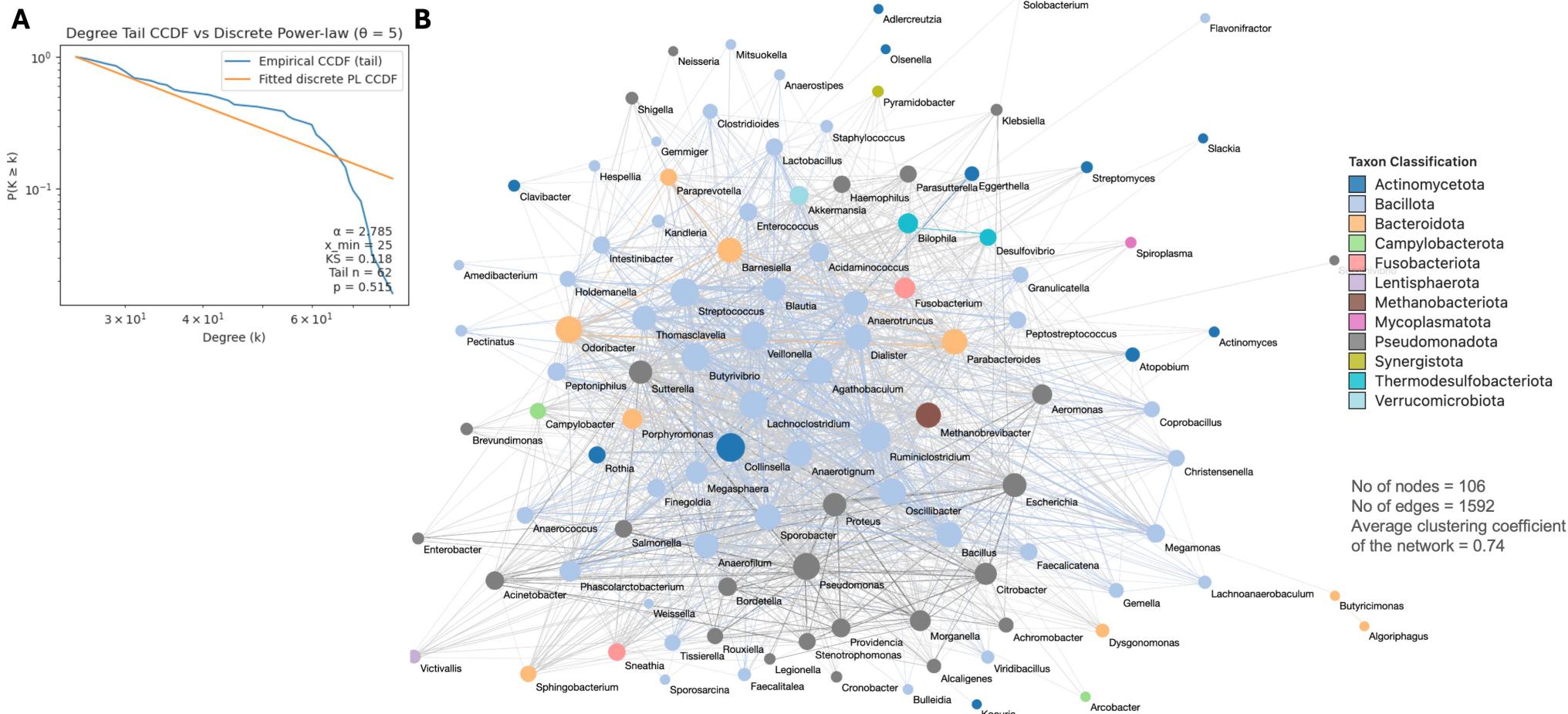
The genus-level GDCN comprised of 375 taxa nodes and 23,415 edges. An edge-weight threshold at the 95<sup>th</sup> percentile ( $\theta = 5$ ) reduced the network to 106 taxa nodes and 1,592 edges beyond which the network fragmented into smaller disconnected components. This resulted in a moderately dense (density 0.28) and a compact structure (average clustering coefficient 0.74), which formed a single connected component with no modular community formation (Fig. 5B).

The degree distribution analysis of the GDCN, although, revealed a theoretical scale-free-like pattern with a scaling component ( $\alpha$ ) of 2.78 with a minimum node degree of 8 based on the KS statistic of 0.118 (p-value=0.38), the CCDF failed to align with the power-law model indicating that the network was composed of several high degree nodes rather than a single dominant superhub node (Fig. 7A).

The most central taxon node is the genus *Ruminiclostridium*, with a degree centrality of 0.77 and a closeness centrality of 0.8. Several other high-degree nodes also formed the central core of the network; some of the top genera include *Collinsella* (degree centrality 0.71; closeness centrality 0.77), *Streptococcus* (degree centrality 0.7; closeness centrality 0.76), *Lachnoclostridium* (degree centrality 0.68; closeness centrality 0.75), and *Butyrivibrio* (degree centrality 0.68; closeness centrality 0.75). Betweenness centrality was low throughout the network (mean 0.008; median 0.0005), indicating that many alternative paths connected the nodes, and no single genus acted as a bottleneck (Fig. 7B).

The GDCN at the genus level showed that the network was dominated by members of the phyla *Bacillota* (47%), followed by *Pseudomonadota* (24%), *Actinomycetota* (11%), and *Bacteroidota* (9%). In addition, most of the central core nodes also belong to the phylum *Bacillota*, suggesting that the phylum *Bacillota* not only dominates the network but also the central core.

# COMPREHENSIVE NETWORK VISUALISATION GUT DYSBIOSIS CO-OCCURRENCE NETWORK - GENUS LEVEL



**Figure 7: Comprehensive Network Visualisation – Gut Dysbiosis Co-occurrence Network (Genus level).**

**A.** GDCN degree distribution plot showing the empirical CCDF (blue) representing the degree distribution of the nodes in a log-log plot and power-law CCDF (PL CCDF) (orange line). **B.** GDCN unipartite network. Nodes represent taxa (genera), and the edges represent the shared diseases. Edge thickness is proportional to the number of shared diseases (weighted edges). Taxa nodes are colour-coded based on phylum derived from the NCBI taxonomy classification.

### Gut Dysbiosis Co-occurrence Network - Species Level: Network Properties

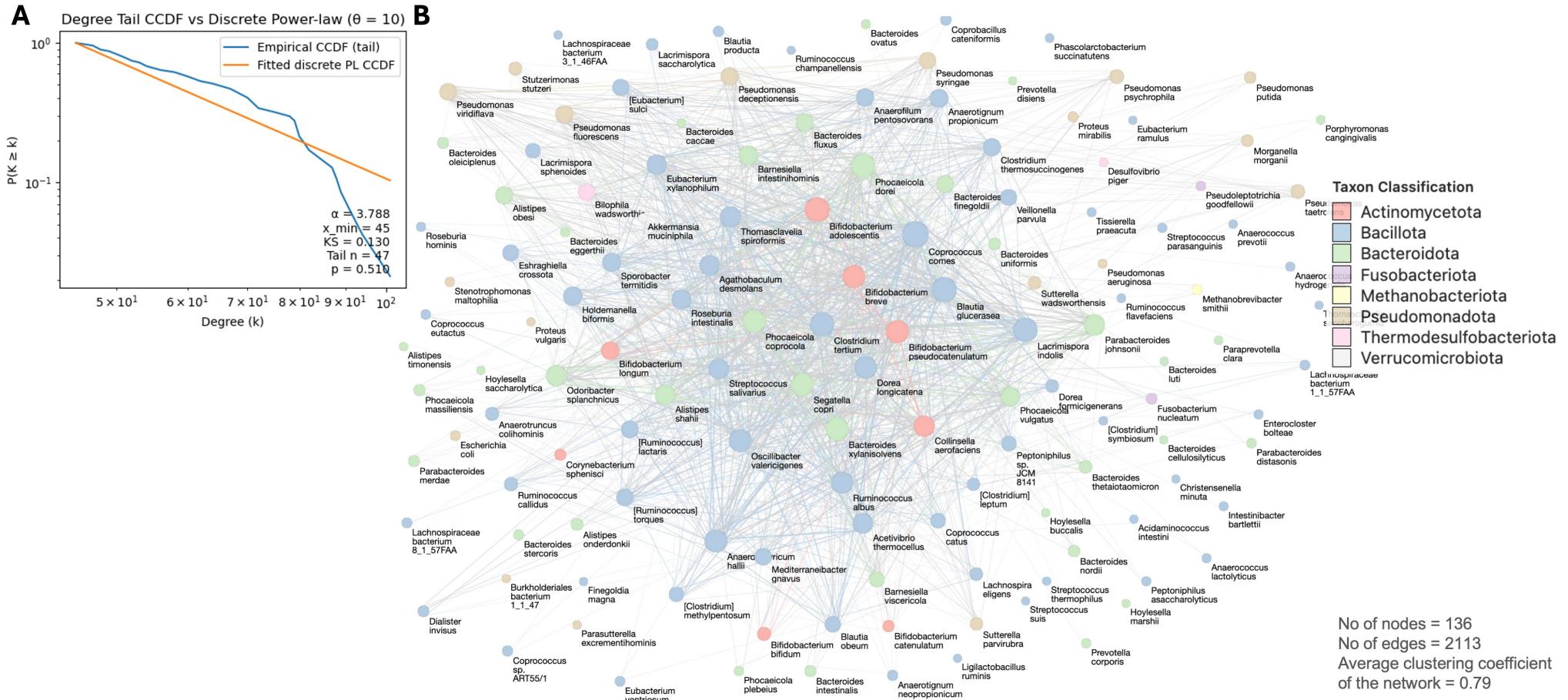
The GDCN at the species level exhibited a network architecture similar to that of the genus-level GDCN. The unthresholded species-level GDCN had 1,144 taxa nodes and 199,073 edges. At the species level, the edge-weight threshold at the 99th percentile corresponded to an absolute edge-weight ( $\theta$ ) of 10 due to the higher number of unique species in the dataset, which resulted in 136 nodes and 2113 edges. Like the genus-level GDCN, the species-level also resulted in a moderately dense (density 0.28) and compact structure (average clustering coefficient 0.74), forming a single connected component with no modular community formation (Fig. 8B).

The degree distribution analysis of the GDCN at the species level, although, revealed a theoretical scale-free-like pattern with a scaling component ( $\alpha$ ) of 3.78 with a minimum node degree of 47 based on the KS statistic (KS = 0.13; p-value=0.51), the CCDF failed to align with the power-law model indicating that the network was composed of several high degree nodes (Fig. 8A).

The most central taxon node is the species *Coprococcus comes*, with a degree centrality of 0.74 and a closeness centrality of 0.79. Several other high-degree nodes also formed the central core of the network; some of the top species include *Blautia glucerasea* (degree centrality 0.69; closeness centrality 0.76), *Bifidobacterium adolescentis* (degree centrality 0.67; closeness centrality 0.75), *Phocaeicola dorei* (degree centrality 0.65; closeness centrality 0.74), and *Lacrimispora indolis* (degree centrality 0.65; closeness centrality 0.74). Betweenness centrality was low throughout the network (mean 0.008; median 0.0005), indicating that many alternative paths connected the nodes and no single disease acted as a bottleneck (Fig. 8B).

The GDCN at the species level showed that the network was dominated by members of the phyla *Bacillota* (49%), followed by *Bacteroidota* (27%) and *Pseudomonadota* (13%). In addition, most of the central core nodes also belong to the phylum *Bacillota*, suggesting that the phylum *Bacillota* not only dominates the network but also the central core.

# COMPREHENSIVE NETWORK VISUALISATION GUT DYSBIOSIS CO-OCCURRENCE NETWORK - SPECIES LEVEL



**Figure 8: Comprehensive Network Visualisation – Gut Dysbiosis Co-occurrence Network (Species level).**

**A.** GDCN degree distribution plot showing the empirical CCDF (blue) representing the degree distribution of the nodes in a log-log plot and power-law CCDF (PL CCDF) (orange line). **B.** GDCN unipartite network. Nodes represent taxa (species), and the edges represent the shared diseases. Edge thickness is proportional to the number of shared diseases (weighted edges). Taxa nodes are colour-coded based on phylum derived from the NCBI taxonomy classification.

## DISCUSSION

The present study provides a comprehensive analysis of the gut microbiome in dysbiosis using a novel approach by combining conventional methods such as alpha and beta diversity and differential abundance taxa analysis with network analysis. Here, alpha and beta diversity analyses were used to evaluate the changes in the gut microbiota at disease level, differential abundance taxa analysis was used to identify taxa that were significantly associated with the diseases and network analysis offered a system-level understanding of the overall microbial pattern across multiple disease states.

In this study, results from the beta diversity analysis showed that nearly all disease groups demonstrated significant variations in their microbial composition when compared with healthy controls (Fig 3. A and B). However, these changes were not reflected in their microbial diversity as a substantial number of diseases showed no differences in their alpha diversity index when compared with healthy controls (Fig. 2A and B). This discrepancy suggested that changes in microbial composition were not always followed by alterations in microbial diversity.

Furthermore, the dataset used in this study was from the GMrepo database, a collection of curated metagenome samples and there was no evidence that the samples in the database were in a dysbiotic state at the time of collection, despite being collected from diseased patients. Moreover, the repository contained samples from different geographical locations, covering a wide range of age groups and dietary habits, with various host genotypes (Wu et al. 2020). Therefore, the observed differences in beta diversity analysis could also be attributed to the demographic differences between samples rather than an actual dysbiotic state. This suggests that not all changes observed in the beta diversity analysis necessarily represent dysbiotic state (Williams, Hammer and Williams 2024, Amato and Carmody 2023).

In addition, the alpha diversity analysis from the present study also included disease groups such as stomach neoplasm, Hashimoto's thyroiditis, and Graves' disease, which showed a significant increase in microbial diversity when compared with healthy controls (Fig 2 A and B). Studies have shown that, the loss of microbial diversity and changes in microbial composition are the hallmark signs of dysbiosis (Mosca, Leclerc and Hugot 2016). While this may be partially true, the present study showed that dysbiosis can also manifest as an increase in microbial diversity, at least in certain diseases. Similar increase in microbial diversity in dysbiosis have also been reported in diseases such as metabolic disorders and obesity (Sze and Schloss 2016, Que et al. 2021, Wang et al. 2022, Li et al. 2024). In conclusion, alpha or beta diversity on their own will not be an effective indicator of dysbiosis and must always be integrated with population-level data, such as age group, geographic location, dietary habits, along with host genetics if available.

### **The disease-centred view:**

GMDN at both genus and species levels showed similar network structure and components. This observation was in agreement with the work of Renard et al (2020), which demonstrated that the overall network structure was preserved across genus and species levels, with variations noted only at the level of individual network metrics.

Network analysis of the GMDN at both the genus and species levels revealed that both networks were moderately dense and highly interconnected, with no modular community separation (Figs. 5 and 6). In addition, although both GMDN showed a plausible power-law-like fit theoretically, upon visual inspection, neither conformed to a scale-free distribution, suggesting that both networks were composed of several high-degree nodes, rather than a few superhub nodes (Barabási, Gulbahce and Loscalzo 2011, Barabási and Oltvai 2004). Taken together, this suggests that the GMDN at both levels were composed of multiple high-degree nodes forming the central hub surrounded by peripheral low-degree nodes. These highly interconnected nodes, that formed the central hub, imply the presence of an underlying core set of microbes that represent a shared dysbiotic microbial signature across many clinically distinct diseases. Sun et al. (2024) in a population-scale meta-analysis showed that an universal dysbiotic microbial signature was shared across 28 clinically distinct diseases. Similarly Duvallet et al. (2017) demonstrated similar shared dysbiotic signature in 10 different diseases.

In addition, grouping of disease nodes using broader categories showed that, although the central hub nodes included diseases from different categories, they were dominated by the diseases of the digestive system, such as Irritable bowel syndrome, ulcerative colitis, and stomach neoplasms. This could be either due to the overrepresentation of digestive system diseases in the dataset or may be a more direct effect of the gut dysbiosis in the digestive system diseases, as the gut is the primary site of microbial interaction (Duvallet et al. 2017, Opazo et al. 2018). This suggests that, in gut dysbiosis, the diseases of the digestive system play an important role in forming the dysbiotic microbial pattern and may act as a key link to other diseases. This may lead to the hypothesis that non-intestinal diseases may have a gut-related cause due to shared microbial dysbiosis.

### **The taxon-centred view:**

While the GMDN showed how diseases share microbial patterns, the GDCN showed how the microbes co-occurred across shared diseases.

Like the GMDN, both genus and species level GDCN showed similar network architecture (Fig 7 and 8). They were both moderately dense and highly interconnected, except that the species-level GDCN had more nodes (species) and edges, even after the edge weights were thresholded at the 99th percentile. As expected, at both the taxonomic levels, these networks contained several highly interconnected nodes and were not dominated by superhub nodes. This was in agreement with the report by Hall et al (2019), who showed a similar node arrangement in their taxa co-occurrence network in the healthy gut microbiome.

In the GDCN, the central high-degree hub nodes showed microbes that frequently co-occurred in most diseases in the dataset, representing a shared dysbiotic signature, while the peripheral nodes represented disease-specific microbes. In summary, the GDCN demonstrated that gut dysbiosis is a coordinated community-level transition of microbes, rather than changes at the individual taxon level. This result aligned with experimentally validated studies by Loftus et al. (2021) and Bajaj et al. (2024), who demonstrated a common dysbiotic core with a disease-specific periphery in Crohn's disease, intestinal tuberculosis and colorectal cancer.

In the GDCN, when the taxa nodes were coloured by phylum, no specific community patterns emerged at both the genus and species levels. Instead, it was observed that many of the central hub nodes belong to the phyla Bacillota, Bacteroidota and Actinomycetota at both genus and species levels. Several studies have demonstrated that these phyla dominate the gut microbiota of both diseased and healthy individuals (Magne et al. 2020, Shin et al. 2024, Hul et al. 2024). In the healthy gut microbiome, members of the phyla Bacteroidota and Bacillota live in a mutualistic relationship in which Bacteroidota digests the dietary fibres and Bacillota utilises the byproducts and produces SCFAs, which are essential in maintaining the gut ecosystem and several other essential functions like maintaining the epithelial integrity and immune modulation (Ríos-Covián et al. 2016, Louis and Flint 2017, Magne et al. 2020).

In the present study, several genera were identified as key hubs in the GDCN at the genus level, and only a few are discussed below. Studies have shown that *Ruminiclostridium* and *Collinsella* were found to be elevated in metabolic disorders like obesity, autoimmune diseases (Plaza-Díaz et al. 2022, Han et al. 2025) and inflammatory conditions such as atherosclerosis and rheumatoid arthritis (Ruiz-Limón et al. 2022, Han et al. 2025) respectively. Additionally, it was reported that *Blautia* and *Butyrivibrio*, both butyrate-producing organisms, were found in low levels in colorectal cancer (Pozuelo et al. 2015, Chen et al. 2012, Wang et al. 2012). *Pseudomonas* and *Streptococcus* are other examples of opportunistic bacteria that frequently proliferate in dysbiosis, especially when mucosal injury and immunological suppression are present. *Pseudomonas* was commonly found in immunocompromised states, including antibiotic-induced dysbiosis, HIV, critical illnesses, and colorectal cancer (Vujkovic-Cvijin et al. 2013, Evans et al. 2023, Ubeda et al. 2010), while *Streptococci spp.* were linked to colorectal cancer as reported in Pimentel et al (2016) and Yu et al (2017).

Similarly, some of the species identified as key hubs in the species-level GDCN were reported to be associated with diseases. Especially, *Coprococcus comes*, *Blautia glucerasea*, and *Bifidobacterium adolescentis* are butyrate-producing commensal bacteria reported to be reduced in diseases such as inflammatory bowel disease, metabolic disorders such as obesity and type 2 diabetes and autoimmune diseases (Louis and Flint 2009, Liu et al. 2021, AL-Ishaq, Samuel and Büselberg 2023, Yang et al. 2023).

In conclusion, the findings from both the disease-centric and taxon-centric networks suggest that the pattern of gut dysbiosis is not specific to a particular disease or disease group. Instead, they were highly interconnected and overlapped across several diseases.

### **Limitations and future works**

One of the significant limitation in the present study was the construction of the bipartite disease-taxon network using unweighted edges. This led to the assumption that every taxon identified by differential abundance taxa analysis to be equally associated with the disease irrespective of its relative abundance value. Another potential limitation was that the taxa that were present in the healthy but absent in the diseased samples were excluded from the differential abundance taxa and the network analysis.

Future work should include the construction of an annotated network that incorporates demographic details, such as age, sex, diet, and population information, as the gut microbiome is sensitive to these factors. Secondly, using genetic information for disease from databases such as OMIM, a gene-disease-taxa tripartite network can be constructed, linking disease genes with taxa.

## REFERENCES

- Ahlawat, S., Asha, and Sharma, K.K., 'Gut–Organ Axis: A Microbial Outreach and Networking', *Letters in Applied Microbiology*, 72/6 (2021), 636–68
- AL-Ishaq, Raghad Khalid, Samuel, Samson Mathews, and Büsselberg, Dietrich, 'The Influence of Gut Microbial Species on Diabetes Mellitus', *International Journal of Molecular Sciences*, 24/9 (2023), 8118
- Amato, Katherine R., and Carmody, Rachel N., 'Gut Microbial Intersections with Human Ecology and Evolution', *Annual Review of Anthropology*, 52/Volume 52, 2023 (2023), 295–311
- Anderson, Marti J., 'A New Method for Non-Parametric Multivariate Analysis of Variance', *Austral Ecology*, 26/1 (2001), 32–46
- Andresen, Viola, Löwe, Bernd, Broicher, Wiebke, Riegel, Björn, Fraedrich, Katharina, von Wulffen, Moritz, et al., 'Post-Infectious Irritable Bowel Syndrome (PI-IBS) after Infection with Shiga-like Toxin-Producing Escherichia Coli (STEC) O104:H4: A Cohort Study with Prospective Follow-Up', *United European Gastroenterology Journal*, 4/1 (2016), 121–31
- Bajaj, Aditya, Markandey, Manasvini, Samal, Amit, Goswami, Sourav, Vuyyuru, Sudheer K., Mohta, Srikant, et al., 'Depletion of Core Microbiome Forms the Shared Background against Diverging Dysbiosis Patterns in Crohn's Disease and Intestinal Tuberculosis: Insights from an Integrated Multi-Cohort Analysis', *Gut Pathogens*, 16 (2024), 65
- Barabasi, A. L., and Albert, R., 'Emergence of Scaling in Random Networks', *Science (New York, N.Y.)*, 286/5439 (1999), 509–12
- Barabási, Albert-László, Gulbahce, Natali, and Loscalzo, Joseph, 'Network Medicine: A Network-Based Approach to Human Disease', *Nature Reviews. Genetics*, 12/1 (2011), 56–68
- Barabási, Albert-László, and Oltvai, Zoltán N., 'Network Biology: Understanding the Cell's Functional Organization', *Nature Reviews. Genetics*, 5/2 (2004), 101–13
- Barnich, Nicolas, Arthur, Janelle C., Buisson, Anthony, Campbell, Barry J., Carbonnel, Franck, Chassaing, Benoit, et al., 'Adherent-Invasive Escherichia Coli in Crohn's Disease: The 25th Anniversary', *Gut*, 2025 <<https://gut.bmjjournals.org/content/early/2025/06/04/gutjnl-2025-335331>> [accessed 28 August 2025]
- Bernstein, Harris, and Bernstein, Carol, 'Bile Acids as Carcinogens in the Colon and at Other Sites in the Gastrointestinal System', *Experimental Biology and Medicine*, 248/1 (2023), 79–89
- Carabotti, Marilia, Scirocco, Annunziata, Maselli, Maria Antonietta, and Severi, Carola, 'The Gut-Brain Axis: Interactions between Enteric Microbiota, Central and Enteric Nervous Systems', *Annals of Gastroenterology: Quarterly Publication of the Hellenic Society of Gastroenterology*, 28/2 (2015), 203–9
- Chen, Weiguang, Liu, Fanlong, Ling, Zongxin, Tong, Xiaojuan, and Xiang, Charlie, 'Human Intestinal Lumen and Mucosa-Associated Microbiota in Patients with Colorectal Cancer', *PLOS ONE*, 7/6 (2012), e39743
- Choden, Tenzin, and Cohen, Nathaniel Aviv, 'The Gut Microbiome and the Immune System', *Exploration of Medicine*, 3/3 (2022), 219–33
- Collins, Stephanie L., and Patterson, Andrew D., 'The Gut Microbiome: An Orchestrator of Xenobiotic Metabolism', *Acta Pharmaceutica Sinica. B*, 10/1 (2020), 19–32

- Collins, Stephanie L., Stine, Jonathan G., Bisanz, Jordan E., Okafor, C. Denise, and Patterson, Andrew D., 'Bile Acids and the Gut Microbiota: Metabolic Interactions and Impacts on Disease', *Nature Reviews Microbiology*, 21/4 (2023), 236–47
- Corrêa-Oliveira, Renan, Fachi, José Luís, Vieira, Aline, Sato, Fabio Takeo, and Vinolo, Marco Aurélio R., 'Regulation of Immune Cell Function by Short-Chain Fatty Acids', *Clinical & Translational Immunology*, 5/4 (2016), e73
- Coyte, Katharine Z., Schluter, Jonas, and Foster, Kevin R., 'The Ecology of the Microbiome: Networks, Competition, and Stability', *Science (New York, N.Y.)*, 350/6261 (2015), 663–66
- Dmytriv, Tetiana R., Storey, Kenneth B., and Lushchak, Volodymyr I., 'Intestinal Barrier Permeability: The Influence of Gut Microbiota, Nutrition, and Exercise', *Frontiers in Physiology*, 15 (2024), 1380713
- Duvallet, Claire, Gibbons, Sean M., Gurry, Thomas, Irizarry, Rafael A., and Alm, Eric J., 'Meta-Analysis of Gut Microbiome Studies Identifies Disease-Specific and Shared Responses', *Nature Communications*, 8/1 (2017), 1784
- Evans, Tess, Ali, Umar, Anderton, Ryan, Raby, Edward, Manning, Laurens, and Litton, Edward, 'Lower Gut Dysbiosis and Mortality in Acute Critical Illness: A Systematic Review and Meta-Analysis', *Intensive Care Medicine Experimental*, 11 (2023), 6
- Everard, Amandine, Belzer, Clara, Geurts, Lucie, Ouwerkerk, Janneke P., Druart, Céline, Bindels, Laure B., et al., 'Cross-Talk between Akkermansia Muciniphila and Intestinal Epithelium Controls Diet-Induced Obesity', *Proceedings of the National Academy of Sciences of the United States of America*, 110/22 (2013), 9066–71
- Farzi, Aitak, Fröhlich, Esther E., and Holzer, Peter, 'Gut Microbiota and the Neuroendocrine System', *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics*, 15/1 (2018), 5–22
- Faust, Karoline, 'Open Challenges for Microbial Network Construction and Analysis', *The ISME Journal*, 15/11 (2021), 3111–18
- Friedman, Jonathan, and Alm, Eric J., 'Inferring Correlation Networks from Genomic Survey Data', *PLOS Computational Biology*, 8/9 (2012), e1002687
- Gao, Kan, Mu, Chun-long, Farzi, Aitak, and Zhu, Wei-yun, 'Tryptophan Metabolism: A Link Between the Gut Microbiota and Brain', *Advances in Nutrition*, 11/3 (2020), 709–23
- Goh, Kwang-II, Cusick, Michael E., Valle, David, Childs, Barton, Vidal, Marc, and Barabási, Albert-László, 'The Human Disease Network', *Proceedings of the National Academy of Sciences of the United States of America*, 104/21 (2007), 8685–90
- Hall, Caitlin V., Lord, Anton, Betzel, Richard, Zakrzewski, Martha, Simms, Lisa A., Zalesky, Andrew, et al., 'Co-Existence of Network Architectures Supporting the Human Gut Microbiome', *IScience*, 22 (2019), 380–91
- Han, Eui-Jeong, Ahn, Ji-Seon, Choi, Yu-Jin, Kim, Da-Hye, and Chung, Hea-Jong, 'Changes in Gut Microbiota According to Disease Severity in a Lupus Mouse Model', *International Journal of Molecular Sciences*, 26/3 (2025), 1006
- Hang, Saiyu, Paik, Donggi, Yao, Lina, Kim, Eunha, Trinath, Jamma, Lu, Jingping, et al., 'Bile Acid Metabolites Control TH17 and Treg Cell Differentiation', *Nature*, 576/7785 (2019), 143–48

- He, Jin, Zhang, Peiwen, Shen, Linyuan, Niu, Lili, Tan, Ya, Chen, Lei, et al., 'Short-Chain Fatty Acids and Their Association with Signalling Pathways in Inflammation, Glucose and Lipid Metabolism', *International Journal of Molecular Sciences*, 21/17 (2020), 6356
- Hou, Kaijian, Wu, Zhuo-Xun, Chen, Xuan-Yu, Wang, Jing-Quan, Zhang, Dongya, Xiao, Chuanxing, et al., 'Microbiota in Health and Diseases', *Signal Transduction and Targeted Therapy*, 7/1 (2022), 135
- Hul, Matthias Van, Cani, Patrice D., Petitfrère, Camille, Vos, Willem M. De, Tilg, Herbert, and El-Omar, Emad M., 'What Defines a Healthy Gut Microbiome?', *Gut*, 73/11 (2024), 1893–1908
- Jandhyala, Sai Manasa, Talukdar, Rupjyoti, Subramanyam, Chivkula, Vuyyuru, Harish, Sasikala, Mitnala, and Reddy, D Nageshwar, 'Role of the Normal Gut Microbiota', *World Journal of Gastroenterology : WJG*, 21/29 (2015), 8787–8803
- Jayakrishnan, Thejus T., Sangwan, Naseer, Barot, Shimoli V., Farha, Nicole, Mariam, Arshiya, Xiang, Shao, et al., 'Multi-Omics Machine Learning to Study Host-Microbiome Interactions in Early-Onset Colorectal Cancer', *Npj Precision Oncology*, 8/1 (2024), 146
- Kajihara, Kacie T., and Hynson, Nicole A., 'Networks as Tools for Defining Emergent Properties of Microbiomes and Their Stability', *Microbiome*, 12/1 (2024), 184
- Khan, Rabia, Petersen, Fernanda Cristina, and Shekhar, Sudhanshu, 'Commensal Bacteria: An Emerging Player in Defense Against Respiratory Pathogens', *Frontiers in Immunology*, 10 (2019) <<https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2019.01203/full>> [accessed 27 August 2025]
- Kurdi, Peter, Kawanishi, Koji, Mizutani, Kanako, and Yokota, Atsushi, 'Mechanism of Growth Inhibition by Free Bile Acids in Lactobacilli and Bifidobacteria', *Journal of Bacteriology*, 188/5 (2006), 1979–86
- LeBlanc, Jean Guy, Milani, Christian, de Giori, Graciela Savoy, Sesma, Fernando, van Sinderen, Douwe, and Ventura, Marco, 'Bacteria as Vitamin Suppliers to Their Host: A Gut Microbiota Perspective', *Current Opinion in Biotechnology*, 24/2 (2013), 160–68
- Ley, Ruth E., Turnbaugh, Peter J., Klein, Samuel, and Gordon, Jeffrey I., 'Microbial Ecology: Human Gut Microbes Associated with Obesity', *Nature*, 444/7122 (2006), 1022–23
- Li, Xiaxi, Li, Xiaoling, Xiao, Haowei, Xu, Jiaying, He, Jianquan, Xiao, Chuanxing, et al., 'Meta-Analysis of Gut Microbiota Alterations in Patients with Irritable Bowel Syndrome', *Frontiers in Microbiology*, 15 (2024), 1492349
- Liu, Xuemei, Mao, Bingyong, Gu, Jiayu, Wu, Jiaying, Cui, Shumao, Wang, Gang, et al., 'Blautia-a New Functional Genus with Potential Probiotic Properties?', *Gut Microbes*, 13/1 (2021), 1–21
- Llorente, Cristina, and Schnabl, Bernd, 'The Gut Microbiota and Liver Disease', *Cellular and Molecular Gastroenterology and Hepatology*, 1/3 (2015), 275–84
- Loftus, Mark, Hassouneh, Sayf Al-Deen, and Yooseph, Shibu, 'Bacterial Community Structure Alterations within the Colorectal Cancer Gut Microbiome', *BMC Microbiology*, 21 (2021), 98
- Louis, Petra, and Flint, Harry J., 'Diversity, Metabolism and Microbial Ecology of Butyrate-Producing Bacteria from the Human Large Intestine', *FEMS Microbiology Letters*, 294/1 (2009), 1–8

- Louis, Petra, and Flint, Harry J, 'Formation of Propionate and Butyrate by the Human Colonic Microbiota', *Environmental Microbiology*, 19/1 (2017), 29–41
- Magne, Fabien, Gotteland, Martin, Gauthier, Lea, Zazueta, Alejandra, Pesoa, Susana, Navarrete, Paola, et al., 'The Firmicutes/Bacteroidetes Ratio: A Relevant Marker of Gut Dysbiosis in Obese Patients?', *Nutrients*, 12/5 (2020), 1474
- Mahmud, Md Rayhan, Akter, Sharmin, Tamanna, Sanjida Khanam, Mazumder, Lincon, Esti, Israt Zahan, Banerjee, Sanchita, et al., 'Impact of Gut Microbiome on Skin Health: Gut-Skin Axis Observed through the Lenses of Therapeutics and Skin Diseases', *Gut Microbes*, 14/1 (2022), 2096995
- Mancera-Hurtado, Yesenia, Lopez-Contreras, Blanca E., Flores-Lopez, Regina, Villamil-Ramirez, Hugo, Del Rio-Navarro, Blanca E., Canizales-Quinteros, Samuel, et al., 'A Dysregulated Bile Acids Pool Is Associated with Metabolic Syndrome and Gut Microbial Dysbiosis in Early Adolescence', *Obesity (Silver Spring, Md.)*, 31/8 (2023), 2129–38
- Martínez-Ruiz, Manuel, Robeson, Michael S., and Piccolo, Brian D., 'Fueling the Fire: Colonocyte Metabolism and Its Effect on the Colonic Epithelia', *Critical Reviews in Food Science and Nutrition*, 2025, 1–20
- Matchado, Monica Steffi, Lauber, Michael, Reitmeier, Sandra, Kacprowski, Tim, Baumbach, Jan, Haller, Dirk, et al., 'Network Analysis Methods for Studying Microbial Communities: A Mini Review', *Computational and Structural Biotechnology Journal*, 19 (2021), 2687–98
- Mazmanian, Sarkis K., Liu, Cui Hua, Tzianabos, Arthur O., and Kasper, Dennis L., 'An Immuno-modulatory Molecule of Symbiotic Bacteria Directs Maturation of the Host Immune System', *Cell*, 122/1 (2005), 107–18
- Morrison, Douglas J., and Preston, Tom, 'Formation of Short Chain Fatty Acids by the Gut Microbiota and Their Impact on Human Metabolism', *Gut Microbes*, 7/3 (2016), 189–200
- Mosca, Alexis, Leclerc, Marion, and Hugot, Jean P., 'Gut Microbiota Diversity and Human Diseases: Should We Reintroduce Key Predators in Our Ecosystem?', *Frontiers in Microbiology*, 7 (2016) <<https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2016.00455/full>> [accessed 22 August 2025]
- Mousa, Walaa K., Chehadeh, Fadia, and Husband, Shannon, 'Microbial Dysbiosis in the Gut Drives Systemic Autoimmune Diseases', *Frontiers in Immunology*, 13 (2022), 906258
- Musso, Giovanni, Gambino, Roberto, and Cassader, Maurizio, 'Obesity, Diabetes, and Gut Microbiota: The Hygiene Hypothesis Expanded?', *Diabetes Care*, 33/10 (2010), 2277–84
- Ohkubo, T., Tsuda, M., Tamura, M., and Yamamura, M., 'Impaired Superoxide Production in Peripheral Blood Neutrophils of Germ-Free Rats', *Scandinavian Journal of Immunology*, 32/6 (1990), 727–29
- Opazo, Maria C., Ortega-Rocha, Elizabeth M., Coronado-Arrázola, Irenice, Bonifaz, Laura C., Boudin, Helene, Neunlist, Michel, et al., 'Intestinal Microbiota Influences Non-Intestinal Related Autoimmune Diseases', *Frontiers in Microbiology*, 9 (2018) <<https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2018.00432/full>> [accessed 22 August 2025]
- Ou, Suwen, Wang, Hufei, Tao, Yangbao, Luo, Kangjia, Ye, Jinhua, Ran, Songlin, et al., 'Fusobacterium Nucleatum and Colorectal Cancer: From Phenomenon to Mechanism', *Frontiers in Cellular and Infection Microbiology*, 12 (2022), 1020583

- Papoutsoglou, Georgios, Tarazona, Sonia, Lopes, Marta B., Klammsteiner, Thomas, Ibrahimi, Eliana, Eckenberger, Julia, et al., 'Machine Learning Approaches in Microbiome Research: Challenges and Best Practices', *Frontiers in Microbiology*, 14 (2023), 1261889
- Pham, Van T., Dold, Susanne, Rehman, Ateequr, Bird, Julia K., and Steinert, Robert E., 'Vitamins, the Gut Microbiome and Gastrointestinal Health in Humans', *Nutrition Research*, 95 (2021), 35–53
- Pimentel, B. a. S., Martins, C. a. S., Mendonça, J. C., Miranda, P. S. D., Sanches, G. F., Mattos-Guaraldi, A. L., et al., 'Streptococcus Agalactiae Infection in Cancer Patients: A Five-Year Study', *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology*, 35/6 (2016), 927–33
- Plaza-Díaz, Julio, Manzano, Manuel, Ruiz-Ojeda, Francisco Javier, Giron, Maria D., Salto, Rafael, López-Pedrosa, Jose M., et al., 'Intake of Slow-Digesting Carbohydrates Is Related to Changes in the Microbiome and Its Functional Pathways in Growing Rats with Obesity Induced by Diet', *Frontiers in Nutrition*, 9 (2022) <<https://www.frontiersin.org/journals/nutrition/articles/10.3389/fnut.2022.992682/full>> [accessed 23 August 2025]
- Pozuelo, Marta, Panda, Suchita, Santiago, Alba, Mendez, Sara, Accarino, Anna, Santos, Javier, et al., 'Reduction of Butyrate- and Methane-Producing Microorganisms in Patients with Irritable Bowel Syndrome', *Scientific Reports*, 5/1 (2015), 12693
- Que, Yanyan, Cao, Man, He, Jianquan, Zhang, Qiang, Chen, Qiongyun, Yan, Changsheng, et al., 'Gut Bacterial Characteristics of Patients With Type 2 Diabetes Mellitus and the Application Potential', *Frontiers in Immunology*, 12 (2021) <<https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2021.722206/full>> [accessed 15 August 2025]
- Ramezani, Ali, and Raj, Dominic S., 'The Gut Microbiome, Kidney Disease, and Targeted Interventions', *Journal of the American Society of Nephrology : JASN*, 25/4 (2014), 657–70
- Renaud, Estelle, Baudry, Emmanuelle, and Bessa-Gomes, Carmen, 'Influence of Taxonomic Resolution on Mutualistic Network Properties', *Ecology and Evolution*, 10/7 (2020), 3248–59
- Rhee, Sang Hoon, 'Lipopolysaccharide: Basic Biochemistry, Intracellular Signaling, and Physiological Impacts in the Gut', *Intestinal Research*, 12/2 (2014), 90–95
- Ríos-Covián, David, Ruas-Madiedo, Patricia, Margolles, Abelardo, Gueimonde, Miguel, de Los Reyes-Gavilán, Clara G., and Salazar, Nuria, 'Intestinal Short Chain Fatty Acids and Their Link with Diet and Human Health', *Frontiers in Microbiology*, 7 (2016), 185
- Roediger, W. E., 'Role of Anaerobic Bacteria in the Metabolic Welfare of the Colonic Mucosa in Man', *Gut*, 21/9 (1980), 793–98
- Rozera, Tommaso, Pasolli, Edoardo, Segata, Nicola, and Ianiro, Gianluca, 'Machine Learning and Artificial Intelligence in the Multi-Omics Approach to Gut Microbiota', *Gastroenterology, Shaping the Future of Gastroenterology and Hepatology With Artificial Intelligence*, 169/3 (2025), 487–501
- Ruiz-Limón, Patricia, Mena-Vázquez, Natalia, Moreno-Indias, Isabel, Manrique-Arija, Sara, Lisbona-Montañez, Jose Manuel, Cano-García, Laura, et al., 'Collinsella Is Associated with Cumulative Inflammatory Burden in an Established Rheumatoid Arthritis Cohort', *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie*, 153 (2022), 113518
- Sanos, Stephanie L, Bui, Viet L, Mortha, Arthur, Oberle, Karin, Heners, Charlotte, Johner, Caroline, et al., 'ROR $\gamma$ t and Commensal Microflora Are Required for the Differentiation of Mucosal Interleukin 22-Producing NKp46+ Cells', *Nature Immunology*, 10/1 (2009), 83–91

Seekatz, Anna M., and Young, Vincent B., 'Clostridium Difficile and the Microbiota', *The Journal of Clinical Investigation*, 124/10 (2014), 4182–89

Shin, Jae Hyun, Tillotson, Glenn, MacKenzie, Tiffany N., Warren, Cirle A., Wexler, Hannah M., and Goldstein, Ellie J. C., 'Bacteroides and Related Species: The Keystone Taxa of the Human Gut Microbiota', *Anaerobe*, 85 (2024), 102819

Song, Guangyao, Xie, Yuxiao, Yi, Lanlan, Cheng, Wenjie, Jia, Huijin, Shi, Wenzhe, et al., 'Bile Acids Affect Intestinal Barrier Function through FXR and TGR5', *Frontiers in Medicine*, 12 (2025) <<https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2025.1607899/full>> [accessed 27 August 2025]

Sun, Wen, Zhang, Yue, Guo, Ruochun, Sha, Shanshan, Chen, Changming, Ullah, Hayan, et al., 'A Population-Scale Analysis of 36 Gut Microbiome Studies Reveals Universal Species Signatures for Common Diseases', *Npj Biofilms and Microbiomes*, 10/1 (2024), 96

Sze, Marc A., and Schloss, Patrick D., 'Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome', *mBio*, 7/4 (2016), e01018-16

Tarracchini, Chiara, Lugli, Gabriele Andrea, Mancabelli, Leonardo, van Sinderen, Douwe, Turroni, Francesca, Ventura, Marco, et al., 'Exploring the Vitamin Biosynthesis Landscape of the Human Gut Microbiota', *mSystems*, 9/10 (2024), e0092924

Thursby, Elizabeth, and Juge, Nathalie, 'Introduction to the Human Gut Microbiota', *Biochemical Journal*, 474/11 (2017), 1823–36

Ubeda, Carles, Taur, Ying, Jenq, Robert R., Equinda, Michele J., Son, Tammy, Samstein, Miriam, et al., 'Vancomycin-Resistant Enterococcus Domination of Intestinal Microbiota Is Enabled by Antibiotic Treatment in Mice and Precedes Bloodstream Invasion in Humans', *The Journal of Clinical Investigation*, 120/12 (2010), 4332–41

Vujkovic-Cvijin, Ivan, Dunham, Richard M., Iwai, Shoko, Maher, M. Cyrus, Albright, Rebecca G., Broadhurst, Mara J., et al., 'Dysbiosis of the Gut Microbiota Is Associated with HIV Disease Progression and Tryptophan Catabolism', *Science Translational Medicine*, 5/193 (2013), 193ra91

Wang, Caihua, Zhu, Chunpeng, Shao, Liming, Ye, Jun, Shen, Yimin, and Ren, Yuezhong, 'Role of Bile Acids in Dysbiosis and Treatment of Nonalcoholic Fatty Liver Disease', *Mediators of Inflammation*, 2019 (2019), 7659509

Wang, Jing, He, Mei, Yang, Ming, and Ai, Xiaopeng, 'Gut Microbiota as a Key Regulator of Intestinal Mucosal Immunity', *Life Sciences*, 345 (2024), 122612

Wang, Siyu, Ni, Yifan, Zhou, Shuwei, Peng, Huiping, Cao, Ying, Zhu, Yue, et al., 'Effects of Choline Metabolite—Trimethylamine N-Oxide on Immunometabolism in Inflammatory Bowel Disease', *Frontiers in Immunology*, 16 (2025) <<https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2025.1591151/full>> [accessed 28 August 2025]

Wang, Tingting, Cai, Guoxiang, Qiu, Yunping, Fei, Na, Zhang, Menghui, Pang, Xiaoyan, et al., 'Structural Segregation of Gut Microbiota between Colorectal Cancer Patients and Healthy Volunteers', *The ISME Journal*, 6/2 (2012), 320–29

Wang, Yilun, Wei, Jie, Zhang, Weiya, Doherty, Michael, Zhang, Yuqing, Xie, Haibin, et al., 'Gut Dysbiosis in Rheumatic Diseases: A Systematic Review and Meta-Analysis of 92 Observational Studies', *EBioMedicine*, 80 (2022), 104055

Williams, Claire E., Hammer, Tobin J., and Williams, Candace L., ‘Diversity Alone Does Not Reliably Indicate the Healthiness of an Animal Microbiome’, *The ISME Journal*, 18/1 (2024), wrae133

Wu, Sicheng, Sun, Chuqing, Li, Yanze, Wang, Teng, Jia, Longhao, Lai, Senying, et al., ‘GMrepo: A Database of Curated and Consistently Annotated Human Gut Metagenomes’, *Nucleic Acids Research*, 48/D1 (2020), D545–53

Wu, Xinyu, Tian, Jie, and Wang, Shengjun, ‘Insight Into Non-Pathogenic Th17 Cells in Autoimmune Diseases’, *Frontiers in Immunology*, 9 (2018), 1112

Yan, Fang, and Polk, D. Brent, ‘Probiotic Bacterium Prevents Cytokine-Induced Apoptosis in Intestinal Epithelial Cells’, *The Journal of Biological Chemistry*, 277/52 (2002), 50959–65

Yang, Ruipeng, Shan, Shuhua, Shi, Jiangying, Li, Hanqing, An, Ning, Li, Songtao, et al., ‘Copro-coccus Eutactus, a Potent Probiotic, Alleviates Colitis via Acetate-Mediated IgA Response and Microbiota Restoration’, *Journal of Agricultural and Food Chemistry*, 2023

Yu, Jun, Feng, Qiang, Wong, Sunny Hei, Zhang, Dongya, Liang, Qiao Yi, Qin, Youwen, et al., ‘Metagenomic Analysis of Faecal Microbiome as a Tool towards Targeted Non-Invasive Biomarkers for Colorectal Cancer’, *Gut*, 66/1 (2017), 70–78

Yu, Wenqian, Sun, Siyuan, Yan, Yutong, Zhou, Hong, Liu, Ziyi, and Fu, Qiang, ‘The Role of Short-Chain Fatty Acid in Metabolic Syndrome and Its Complications: Focusing on Immunity and Inflammation’, *Frontiers in Immunology*, 16 (2025), 1519925

Zeng, M. Y., Inohara, N., and Nuñez, G., ‘Mechanisms of Inflammation-Driven Bacterial Dysbiosis in the Gut’, *Mucosal Immunology*, 10/1 (2017), 18–26

Zhang, Dachuan, and Frenette, Paul S., ‘Cross Talk between Neutrophils and the Microbiota’, *Blood*, 133/20 (2019), 2168

Zhu, Yingqian, Li, Qingqing, and Jiang, Hua, ‘Gut Microbiota in Atherosclerosis: Focus on Trimethylamine N-oxide’, *Apmis*, 128/5 (2020), 353–66