University of Leeds

Development of an in vitro co-culture biofilm model to decipher human-oral microbiome interactions and investigate mechanisms for prevention of dysbiosis and antimicrobial resistance

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

Jack Jonathon-Paul Lynch

10-31-2025

Title: *Development of an in vitro co-culture biofilm model to decipher human-oral microbiome interactions and investigate mechanisms for prevention of dysbiosis and antimicrobial resistance*

Short Title: *In Silico* and *In vitro* Modelling of Oral Microbiome Dysbiosis

Jack Jonathon-Paul Lynch

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

School of Dentistry, University of Leeds

October 2025

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Thuy Do. Her mentorship over the last four years was the cornerstone of this research. Her unwavering support, insightful critiques, and expert guidance helped shape the direction of this project and were instrumental in my development as a microbiologist. She challenged me to become a more well-rounded researcher and person.

My sincere appreciation is also extended to my supervisory team, Dr Richarda Hawkins, Dr Rob Howlin, Dr David Bradshaw, and Professor Sue Pavitt. Their diverse perspectives and constructive feedback were invaluable, pushing me to refine my ideas and ensuring this work was as rigorous and polished as possible.

This research would not have been possible without the generous financial support of the BBSRC and Haleon Consumer Healthcare. I am also deeply grateful to the academic and technical staff of the Oral Biology department at the University of Leeds for their professionalism, advice, and assistance. My thanks also go to the entire School of Dentistry's Oral Microbiome group for their camaraderie. In particular, I must thank Dr Jonathan Vernon for the essential scientific conversations and the equally essential welcome distractions, which were often centred on the misfortunes of Leeds United. Additionally, I am grateful to Dr Xia Yu for her enduring support and friendship.

On a personal note, I could not have completed this journey without the support of my friends and family. Thank you for your endless patience and encouragement over these four years. I owe special thanks to my partner, Amy, for her unwavering belief in me every single day, my sister, Libby, for her invaluable advice, often accompanied by mind-clearing walks and coffees, and my lifelong friend, Harry, for teaching me the true meaning of problem-solving and perseverance in my other lab, the Climbing Lab.

Finally, my last thanks go to my dog, Barnaby, for being the best and scruffiest boy in my life.

Ethics

Ethical approval for this project was granted by the University of Leeds' Dental Research Ethics Committee for the collection of oral samples from healthy volunteers (DREC 150322/JL/349) (Appendix 0.1)

I acknowledge the use of Gemini (Google, www.google.gemini.com) for code troubleshooting, summarising articles, and to proofread my final draft.

Abstract

Background: Shifts in the oral microbiome from a healthy state to a dysbiotic one are a key driver in oral diseases such as caries and periodontitis. While shotgun metagenomics has enabled taxonomic profiling of these diseases, a comprehensive, multifaceted analysis comparing their taxonomic, functional, resistance, and mobile genetic element profiles is needed. Furthermore, a lack of sophisticated *in vitro* models that capture host-biofilm interactions limits the ability to study the longitudinal progression of oral dysbiosis.

Aims: This project aimed to 1) conduct a comprehensive bioinformatic analysis of public metagenomic data to define and compare the taxonomic, functional, resistome, and mobilome signatures of dental caries and periodontitis compared to healthy controls. 2) Utilise machine learning to determine the predictive power of these signatures for disease classification. 3) Validate a novel host-biofilm co-culture model to observe the microbial and functional changes during an induced dysbiotic shift over 14 days.

Materials and Methods: Publicly available shotgun metagenomic data for caries and periodontitis cohorts were downloaded and processed through a custom bioinformatic pipeline, including assembly, gene prediction, and annotation against custom and public databases. This generated counts for taxonomic, functional, antibiotic resistance gene (ARG), and horizontal gene transfer (HGT) profiles. These counts were used to train classical and deep learning models for disease classification. Concurrently, a stratified epithelial collagen gel model was co-cultured with a complex, multi-species oral biofilm derived from healthy human donors. Dysbiosis was induced using a high serum medium, and samples were collected over a 14-day period for metagenomic and metatranscriptomic sequencing.

Results: Bioinformatic analysis of public data revealed distinct taxonomic and functional profiles for caries and periodontitis compared with health. The resistome and mobilome analyses identified differentially abundant ARGs and HGT events associated with each disease state. Machine learning models accurately classified periodontitis samples from healthy samples but failed to distinguish between caries and healthy samples. The *in vitro*

co-culture model successfully demonstrated a longitudinal shift in both microbial composition and gene expression following the dysbiotic media use, containing key markers of a known periodontitis state.

Conclusion: The findings suggest that dental caries and periodontitis possess unique profiles beyond taxonomy to the functional, resistome, and mobilome level. These signatures are robust enough to serve as predictive biomarkers for disease classification in periodontitis. Furthermore, the developed host-biofilm model provides a novel platform for investigating mechanisms driving the shift from oral health to disease and for testing new therapeutics.

Contents

List of Figures

List of Tables

List of Abbreviations

- ACTB: Actin Beta
- AMR: Antimicrobial Resistance
- ANK2: Ankyrin 2
- ANOVA: Analysis of Variance
- ANXA2: Annexin A2
- ARGs: Antimicrobial Resistance Genes
- AUC-ROC: Area Under the Receiver Operating Characteristic Curve
- BECD: Buoyant Epithelial Culture Device
- bp: Base Pair
- CARD: Comprehensive Antibiotic Resistance Database
- cciMGEs: Conjugative and Chromosomally Integrated MGEs
- CLDN4: Claudin 4
- CLSM: Confocal Laser Scanning Microscopy
- CPM: Counts Per Million
- CTLP: C-terminal processing-like protease
- CXCL: C-X-C motif chemokine ligand
- DB: Database
- DMEM: Dulbecco's Modified Eagle Medium
- DNA: Deoxyribonucleic Acid
- DOK: Dysplastic Oral Keratinocyte
- DUFs: Domains of Unknown Function

- EGFR: Epidermal Growth Factor Receptor
- ENA: European Nucleotide Archive
- EPS: Extracellular Polymeric Substance
- FBS: Foetal Bovine Serum
- FDR: False Discovery Rate
- GO: Gene Ontology
- HGT: Horizontal Gene Transfer
- HOM: Human Oral Microbiome
- HOMD: Human Oral Microbiome Database
- HPC: High Performance Computing
- HSD: Honestly Significant Difference
- hTERT: Human Telomerase Reverse Transcriptase
- IBD: Inflammatory Bowel Disease
- ICEs: Integrative and Conjugative Elements
- IMEs: Integrative Mobilisable Elements
- ITS: Internal Transcribed Spacer
- KRT6B: Keratin 6B
- LEfSe: Linear discriminant analysis Effect Size
- LIME: Local Interpretable Model agnostic Explanations
- LPS: Lipopolysaccharide
- LRT: Likelihood Ratio Test
- MEM: Minimum Essential Medium

- MGEs: Mobile Genetic Elements
- ML: Machine Learning
- MLP: Multi-layer Perceptron
- MLS: Macrolides, Lincosamides, Streptogramins
- MRTFB: Myocardin-related Transcription Factor B
- Msp: Major Outer Sheath Protein
- NCBI: National Centre for Biotechnology Information
- OSCC: Oral Squamous Cell Carcinoma
- PAD: Publicly Available Data
- PCA: Principal Component Analysis
- PCoA: Principal Coordinate Analysis
- PERMANOVA: Permutational Multivariate Analysis of Variance
- PI: Peri-implantitis
- PTS: Phosphoenolpyruvate:sugar phosphotransferase system
- qPCR: Quantitative Polymerase Chain Reaction
- RNA: Ribonucleic Acid
- RPKM: Reads Per Kilobase of transcript per Million mapped reads
- RPS: Ribosomal Proteins
- rRNA: Ribosomal RNA
- RT: Resistotypes
- SGD: Stochastic Gradient Descent
- SHAP: SHapley Additive exPlanations

- SMOTE: Synthetic Minority Over-sampling Technique
- SRA: Sequence Read Archive
- SVC: Support Vector Classifier
- TNFRSF21: TNF Receptor Superfamily Member 21
- TPE: Tree-structured Parzen Estimator
- TXNIP: Thioredoxin Interacting Protein
- XAI: eXplainable AI

Chapter 1: General Introduction

1.1 The Oral Microbiome

The human body is often described as a complex holobiont, an ecosystem comprising host and microbial communities, including bacteria, fungi, viruses, archaea, and protists (van de Guchte, Blottière and Doré, 2018, Hou *et al.*, 2022). These microbial communities, known as microbiomes, aid host functions such as nutrient metabolism, immune homeostasis and defence against pathogens, playing a crucial role in supporting the physiology and health of the host (Jandhyala et al., 2015). Within the human microbiome, individual niches offer unique environments for specific populations of microbes (Mark Welch et al., 2020).

One of the most important of these niches is the human oral microbiome (HOM). It is recognised as a dynamic and complex environment, second only to the gut in terms of both microbial diversity and density (Baker et al., 2024). The oral cavity comprises multiple niches, each with distinct characteristics that support different microbial communities. A balanced oral microbiome contributes to pathogenic colonisation resistance, preventing dysbiotic shifts associated with disease (Kilian *et al.*, 2016). The oral microbiome's complex interactions and metabolic activities suggest that it functions as an integrated system vital to host health (Kilian *et al.*, 2016). Key sites within the mouth, such as the teeth, tongue, palate, and cheeks, offer differing factors, including salivary flow, dietary pH changes, and environmental exposure (Faran Ali and Tanwir, 2012). These factors support the rich microbial community in the oral cavity. For instance, just 1 mL of saliva contains approximately 1×10^8 microbes, with over 700 different taxa identified in the HOM, demonstrating high diversity (Deo and Deshmukh, 2019, Willis and Gabaldón, 2020). Bacteria dominate the HOM, particularly phyla such as *Firmicutes, Bacteroidetes, Proteobacteria, Spirochaetes*, and *Fusobacteria*, which together account for about 96% of the observed taxa (Dewhirst *et al.*, 2010). However, their composition changes through health and disease, with some more associated with one than the other (Table 1.1). Although bacteria make up the majority, other organisms, including fungi, archaea, and viruses, are also essential to maintaining balance in the microbial environment. Fungi, primarily *Candida* species, constitute

a smaller fraction but exist commensally in health (Ghannoum *et al.*, 2010, Montelongo-Jauregui and Lopez-Ribot, 2018). Archaea, like methanogens such as *Methanobrevibacter oralis*, are present in lower numbers and are generally considered non-pathogenic in healthy individuals, although their role in disease is under investigation (Belmok *et al.*, 2020, Dame-Teixeira *et al.*, 2025). Finally, the oral virome contains eukaryotic viruses (e.g., Herpesviridae) and bacteriophages, which can shape bacterial communities with infection and lysis (Martínez et al., 2021).

Microbial communities in the mouth are often found within biofilms, commonly known as dental plaque, with the formation of oral biofilms following a well-documented process (Rath et al., 2021). Early coloniser species, such as *Streptococcus* species, are the first to adhere to surfaces covered by a salivary pellicle, a layer formed by the aggregation of salivary proteins (Li et al., 2004). These early colonisers are vital as they initiate the production of extracellular polymeric substance (EPS) (Chenicheri et al., 2017). Subsequently, secondary colonisers attach via coaggregation, a specific cell-to-cell recognition process that often involves lectin-carbohydrate adhesion. *Fusobacterium nucleatum* acts as a key bridging organism, facilitating the binding of late colonisers, often Gram-negative anaerobes like *Porphyromonas gingivalis* (Kolenbrander *et al.*, 2002, Rath, Bal and Dubey, 2021). Throughout this process, EPS surrounds microbial cells, enhancing the structure of the biofilm and promoting cell-to-cell adhesion. The growth of biofilms offers significant advantages over planktonic cells, as they protect microbes from host immune responses, antimicrobial agents, and mechanical forces such as tooth brushing (Zhao et al., 2023). This resistance is crucial for the persistence of these microbes in the oral cavity.

Furthermore, these characteristics are essential in developing dysbiosis, which refers to an imbalance in the microbial composition of a microbiome (DeGruttola et al., 2016). In the oral microbiome, dysbiosis is linked to various oral diseases, including dental caries and periodontitis, and is also linked to systemic diseases like rheumatoid arthritis and cancer (Peng *et al.*, 2022). External pressures and changes result in fluctuations in species abundance and diversity in the HOM, which may trigger dysbiosis. Factors influencing these shifts include host genetics, diet, oral hygiene practices, antibiotic use, and preexisting systemic diseases. (Li et al., 2022).

Despite this dynamic nature, the HOM in health is usually resilient, being able to withstand environmental changes and prevent the transition to dysbiosis and maintain a state of microbial balance, known as eubiosis (Wade, 2000). Several factors from the host are vital for this stability. Saliva, for example, plays multiple roles, including buffering changes in pH, supplying essential nutrients to commensal microbes, and containing antimicrobial proteins and peptides that selectively target pathogens (Lynge Pedersen and Belstrøm, 2019). Additionally, the immune system in the oral cavity is adapted to its environment, employing distinct strategies to manage the continuous exposure to various microbes. Central to this system is its ability to differentiate between commensal microorganisms and pathogenic ones, through the use of Pattern Recognition Receptors, host proteins that can detect pathogens and trigger an immune response (Şenel, 2021). This distinction is essential for maintaining the health and balance of the HOM.

Moreover, complex interactions among microbes, such as synergism, commensalism, and antagonism, are essential for the HOM's stability and ability to prevent pathogens' colonisation, through competition for nutrients and production of antimicrobials (Baty et al., 2022). Interestingly, some studies indicate that certain oral bacteria influence the host's taste perception and dietary choices, potentially creating a feedback loop that helps maintain their presence in the mouth (Cattaneo et al., 2019). This illustrates a closely linked, coevolved relationship in which the host provides a habitat and nutrients, while the microbial community offers beneficial services. Recognising the various factors from both the host and microbes contributing to this resilience is crucial for developing methods to prevent dysbiosis or restore a healthy microbial balance when disrupted.

Table 1.1: Oral Species and their key characteristics

Phylum	Key Genera	Gram Stain (Typical)	Oxygen Requirement (Typical)	General Role in Health
Firmicutes	*Streptococcus, Veillonella, Lactobacillus*	Positive (*Streptococcus, Lactobacillus*), Negative (*Veillonella*)	Facultative/Anaerobic	Primary colonisers, acid production/utilisation, pH balance (some *Strep,*), biofilm structure, competitive exclusion
Actinobacteria	*Actinomyces, Corynebacterium, Rothia*	Positive	Facultative/Anaerobic	Primary/secondary colonisers, biofilm formation, metabolic interactions, some contribute to pH homeostasis
Bacteroidetes	*Prevotella, Porphyromonas, Tannerella*	Negative	Anaerobic	Low abundance in health, can be pathobionts, involved in

				complex metabolic networks within biofilms
Proteobacteria	*Neisseria, Haemophilus, Aggregatibacter*	Negative	Facultative/Aerobic	Early colonisers, nitrate reduction (some *Neisseria*), and immune modulation
Fusobacteria	*Fusobacterium*	Negative	Anaerobic	Bridging organism in biofilm coaggregation, structural role, can be a pathobiont
Spirochaetes	*Treponema*	Negative (Gram-indeterminate normally)	Anaerobic	Low abundance in health, associated with periodontal pockets, motile

1.2 Major Oral Diseases Associated with Dysbiosis: Dental Caries and Periodontal Disease

Dental caries and periodontal diseases represent two of the most prevalent non-communicable diseases affecting the human population and are closely tied to the dysbiosis of the oral microbiome (Rösing et al., 2023). While both conditions arise from overall dysbiotic oral biofilm formation, they are characterised by distinct disease-associated species, mechanisms, and clinical outcomes (Table 1.2).

Table 1.2: Caries and Periodontitis and their key characteristics

Feature	Dental Caries	Periodontitis
Primary Etiological Driver	Frequent dietary sugar intake leading to prolonged low pH	Host inflammatory/immune response to dysbiotic subgingival biofilm
Key Microbial Signatures/Pathogens	Acidogenic/aciduric: *Streptococcus mutans, Lactobacillus spp., Actinomyces spp*.	Proteolytic/anaerobic: "Red complex" *(P. gingivalis, T. forsythia, T. denticola*), *F. nucleatum, Prevotella spp., F. alocis*
Nature of Dysbiosis	Shift towards an acid-producing and acid-tolerant microbial community	Shift towards a community thriving in an inflammatory, anaerobic, proteolytic environment
Primary Site of Biofilm	Supragingival tooth surfaces (enamel, dentine)	Subgingival sites, within the periodontal pocket
Key Pathogenic Mechanisms	Demineralisation of tooth hard tissues by bacterial metabolic acids	Host-mediated inflammatory destruction of periodontal ligament and alveolar bone

Reversibility	Initial lesions (white spots) are reversible, advanced lesions require restoration	Gingivitis is reversible, Periodontitis involves irreversible loss of attachment and bone
Major Clinical Outcomes	Cavities, pain, pulpal infection, tooth loss	Periodontal pocket, bone loss, gingival recession, tooth mobility, tooth loss

1.2.1 Dental Caries

As aforementioned, HOM dysbiosis can lead to disease within the oral cavity and systemically. Dental caries involves the demineralisation of tooth enamel, caused by a dysbiotic shift toward acidogenic and aciduric bacteria, primarily due to frequent sugar consumption (Marsh, 1999, Dewhirst *et al.*, 2010). The metabolism of sugars by bacteria produces organic acids, such as lactic acid, which lowers the plaque pH. When the pH drops below 5.5, demineralisation of enamel begins. Cavities develop if this process persists and outpaces the natural restoration of minerals within enamel, called remineralisation (Bradshaw and Marsh, 1998).

Key pathogenic bacteria in caries include *Streptococcus mutans*, thought to be the primary pathogen due to its ability for sugar metabolism, acid production (acidogenicity), and acid tolerance (aciduricity), and its frequent presence in those with caries (Matsui and Cvitkovitch, 2010, Dinis *et al.*, 2022). *S. mutans* adheres to tooth surfaces through specific adhesins, creating EPS from sucrose (Koo *et al.*, 2010, Bowen and Koo, 2011). This EPS matrix captures acids, resulting in a localised decrease in pH (Koo et al., 2013). Mechanisms that confer acid tolerance include proton pumps (F_1F_o-ATPase) and DNA repair systems (Liu *et al.*, 2015, Baker, Faustoferri and Quivey, 2017). Other significant species include *Lactobacillus spp.* (linked to lesion progression due to their high acidogenicity/aciduricity), *Actinomyces spp.* (particularly relevant in root caries) (Caufield *et al.*, 2015, Dame-Teixeira *et al.*, 2016). Moreover, *Scardovia wiggsiae* is often associated with severe early childhood caries, even found frequently in those with caries in the absence of traditional caries pathogens like *S. mutans* (Prabhu Matondkar et al., 2020).

The HOM in caries exhibits a transition from health-associated non-mutans streptococci and *Actinomyces* to a predominance of aciduric species, such as *S. mutans*, particularly in initial lesions, with a reduction in overall species diversity (Spatafora et al., 2024). In contrast, dentin caries may exhibit increased diversity compared to enamel caries (Spatafora et al., 2024). Clinically, caries initiates as a reversible white spot lesion. If left untreated, it can progress past the enamel into the dentin, potentially reaching the pulp and causing pulpitis, pain, and eventual tooth loss (Rathee and Sapra, 2025)

1.2.2 Periodontal Diseases

Periodontal diseases encompass inflammatory conditions that affect the structures supporting teeth, caused by dysbiotic biofilms and host immune responses. Gingivitis, characterised by reversible inflammation of the gums with no attachment loss, is primarily driven by non-specific plaque accumulation (Haffajee and Socransky, 2005). In this condition, the microbial community is marked by an increased abundance of Gram-negative anaerobic and facultative species, such as *Prevotella* and *Fusobacterium*, respectively (Iniesta et al., 2024). If gingivitis remains untreated, it can progress to periodontitis.

Periodontitis is a chronic, irreversible inflammatory disease that leads to destruction of the periodontal ligament and alveolar bone, resulting in the formation of periodontal pockets, gum recession, increased tooth mobility, and eventual tooth loss (Chen et al., 2018). This arises from the interplay between the dysbiotic subgingival microbiome and a dysregulated immune response in a susceptible host (Silva *et al.*, 2015). The microbial profile of periodontitis is characterised by an enrichment of predominantly Gram-negative, anaerobic, and proteolytic bacteria (Sandholm, 1986). “Red complex” bacteria, *Porphyromonas gingivalis, Tannerella forsythia*, and *Treponema denticola*, are strongly implicated in its pathogenesis (Suzuki et al., 2013). Additional key species include *Fusobacterium nucleatum, Prevotella intermedia, Aggregatibacter actinomycetemcomitans*, and the emerging periodontal pathogen *Filifactor alocis* (Aruni *et al.*, 2015, Oscarsson *et al.*, 2019).

The pathogenesis of periodontitis involves various virulence factors from these microbes. *P. gingivalis*, identified as a keystone pathogen, a pathogen in low abundance that can initiate pro-inflammatory responses, drives dysbiosis and proliferates in pro-inflammatory environments (Hajishengallis et al., 2012; Hao et al., 2023). Crucial to its pathogenesis are *P. gingivalis’s* gingipains, which degrade host proteins, impair neutrophil function, and disrupt epithelial barriers, while *P. gingivalis* LPS promotes chronic inflammation by activating host immune cells (Imamura, 2003, Nativel *et al.*, 2017). Additionally, *Treponema denticola* utilises motility for tissue penetration, Major Outer Sheath Protein (Msp) for adhesion and pore formation, and dentilisin (CTLP protease) for tissue degradation (Goetting-Minesky et al., 2021; Puthenveetil et

al., 2017). *Fusobacterium nucleatum* functions as a coaggregation bridge, with the FadA adhesin facilitating host cell binding and invasion and enhancing endothelial permeability by disrupting cell-cell junctions (Xu et al., 2007). Lastly, *Filifactor alocis* demonstrates adhesion and invasion capabilities, resists oxidative stress, and produces proteases that may contribute to the overall pathology of periodontal diseases (Aruni et al., 2015).

1.3 The Oral Resistome and Mobilome: Emerging Interest

In addition to its taxonomic interest, the HOM has highlighted the systemic and environmental impacts of oral microbes, including the dissemination of antimicrobial resistance (AMR). The problem of AMR is becoming an ever-growing threat to public health, with an estimated 10 million deaths per annum attributed to AMR by 2050 (K. W. K. Tang et al., 2023). Additionally, it causes increased healthcare costs, demands, and lost work hours due to complications in treating previously treatable infections. The primary driver of this crisis is the misuse of antimicrobials, not just in human healthcare, but also in veterinary medicine and agriculture (Salam et al., 2023).

The term resistome includes all antimicrobial resistance genes (ARGs), as well as their precursors, that are found within a given microbiome (Wright, 2007). ARGs' primary function allows microbes to survive antibiotics in the environment through numerous mechanisms, such as target alteration or antibiotic metabolism, against environmental antibiotics (Reygaert, 2018). Although typically referred to in the context of healthcare, these mechanisms are often a result of the inherent defence of microbes against interspecies warfare with other microbes (Olivares Pacheco et al., 2013). The evolution of the resistome is significantly influenced by HGT, the spread of mobile genetic elements (MGEs), both intra- and inter-species. The rate of HGT is further enhanced by antibiotics and environmental pressures (Barlow, 2009). Therefore, it is essential to monitor this development, with the One Health perspective highlighting the interconnected nature of human, animal and ecological microbiomes, and in this case, the resistome, to understand how different resistomes influence others (Atlas, 2013).

Resistomes are ubiquitous across diverse environments, including soil, water systems, agriculture, and the complex microbiomes of animals and humans, such as the gut and oral cavity (Bargheet

et al., 2023; Li et al., 2023) Environmental microbiomes are considered reservoirs where ARGs can transfer to human pathogenic bacteria (Forsberg *et al.*, 2012, Delgado-Baquerizo *et al.*, 2022). Local ecological conditions and selective pressures influence the prevalence and variety of ARGs. These local resistomes are linked through microbial dissemination and subsequent HGT, supporting the idea of a 'meta-resistome' (Stebliankin et al., 2022). Additionally, environmental pollution, including antibiotic residues and heavy metals, can directly influence the evolution and spread of resistance by increasing HGT rates (Larsson and Flach, 2022).

Acquired resistance occurs via several mechanisms, including limited drug uptake, reduced porin channels or biofilm formation, modification of drug targets, such as changes in antibiotic-binding proteins or ribosomal components, drug inactivation, for instance, via β-lactamase enzymes that break down penicillin, and active efflux of drugs through membrane pumps (Munita and Arias, 2016). The primary mechanism for the spread of acquired ARGs is HGT, which includes transformation (uptake of free DNA), transduction (DNA transfer via bacteriophages), and conjugation (direct DNA transfer between cells) (van Hoek *et al.*, 2011, Burmeister, 2015). MGEs, such as plasmids, transposons, integrons, and bacteriophages, enable HGT (Ellabaan et al., 2021). MGEs can also carry other adaptive genes, such as those associated with virulence. This means that selection for antibiotic resistance might additionally boost pathogenicity, as demonstrated in oral streptococci. (Lee et al., 2024).

As mentioned earlier, the growth of resistomes has led to a global AMR crisis, threatening the effectiveness of antibiotics and increasing healthcare burdens. This situation should be observed from a One Health perspective that combines human, animal, and environmental health approaches (Atlas, 2013). ARGs circulate within microbiomes, with agricultural practices and sewage disposal significantly adding to environmental ARG reservoirs and increasing their movement through the food chain (Fernández-Trapote et al., 2024). Examining all resistomes across these interconnected areas is vital for understanding AMR transmission patterns, creating effective surveillance systems, and guiding antibiotic stewardship policies (Graham *et al.*, 2019, Sukumar *et al.*, 2024)

The oral cavity serves as a significant reservoir for ARGs, constituting the oral resistome (Sukumar et al., 2024). The oral mobilome comprises MGEs, including plasmids, transposons (e.g., Tn916), integrons, and bacteriophages. (Lunde et al., 2021). Oral biofilms are vital for maintaining and spreading the oral resistome. The high microbial density and closeness within biofilms create an optimal microenvironment for HGT (Madsen *et al.*, 2012, Michaelis and Grohmann, 2023). Crucially, oral bacteria and their ARGs are not limited to the mouth and can spread systemically, contributing to infections elsewhere and the wider AMR challenge (Li *et al.*, 2000).

For example, a systematic review identified 159 distinct ARGs conferring resistance to 22 different antibiotic classes across various oral sites, with saliva and supragingival biofilms showing the greatest ARG diversity. Common oral resistance genes include those against macrolides, for example, *ermB*, *msr(D)*, and *mef(A)*, tetracyclines, such as *tet(M*) and *tet(O),* and beta-lactams, like *blaTEM* (Sukumar et al., 2024). Adapted from this review, Table 1.3 gives a detailed summary of these major ARG classes commonly found in the oral cavity. Further investigation has identified a core oral resistome, with three ARGs, *tet(M), tet(O),* and *ermB*, found consistently across all sampled oral sites, indicating they form a fundamental component of the oral resistome (Sukumar et al., 2024). Interestingly, previous findings suggest that a healthy oral microbiome might paradoxically contain a richer resistome at the antibiotic class level compared to caries-affected children, potentially due to the lack of microbial diversity and therefore genes present (Sukumar et al., 2023).

Within the oral mobilome, conjugative and chromosomally integrated MGEs (cciMGEs), including Integrative and Conjugative Elements (ICEs) and Integrative Mobilisable Elements (IMEs), self-transferring and MGEs that require mobilisation, respectively, are common(Lao et al., 2020). These are particularly ubiquitous in oral streptococci and often co-carry ARGs like *tet(M)* and *ermB* alongside virulence factor genes (Lee et al., 2024). The *Tn916* transposase, an ICE, is correlated with ARGs in the developing paediatric oral resistome, and this potential for mobilisation seems to increase with age (Sukumar et al., 2023). The prevalence of MGE families like Tn916-family ICEs carrying *tet(M)* and *ermB* across varied oral bacteria suggests their high

success as vectors in the oral environment (Lunde, Hjerde and Al-Haroni, 2021, Sukumar *et al.*, 2023, Lee *et al.*, 2024).

The accumulation of ARGs in the oral microbiome can significantly reduce the effectiveness of dental treatments for conditions such as periodontitis and dental abscesses. The growing resistance to common dental antibiotics is a significant concern. This issue is compounded by mechanisms such as exposure to subinhibitory antibiotic doses, which, rather than eliminating bacteria, can promote ARG formation by inducing error-prone DNA polymerases (López and Blázquez, 2009). These enzymes can cause mutations that transform non-resistant genes or non-coding DNA into ARGs (Kåhrström, 2013). Low-level antibiotic exposure is a particular problem in healthcare settings, including dentistry, where antibiotics may be used prophylactically at low doses (Hollingshead and Brizuela, 2023). Consequently, prescribing antibiotics in dentistry without prior susceptibility testing can result in ineffective treatment and the progression of dental infections. The inherent tolerance of microbial biofilms to antimicrobials further exacerbates this, as AMR organisms increase resistance and often require more invasive treatments, such as mechanical debridement (Hollingshead and Brizuela, 2023). Despite this, dental practitioners in the UK account for approximately 10% of all antibiotic prescriptions (Sharif, 2012). These practices, particularly the overuse of broad-spectrum antibiotics such as amoxicillin, increase antibiotic exposure in the oral cavity, exerting significant selective pressure and promoting the survival and proliferation of resistant microbial strains (Uddin et al., 2021). These selective pressures directly influence the makeup and intricacy of the oral resistome, thereby contributing to the problem of AMR (Sukumar et al., 2023). For these reasons, dental antibiotic stewardship programs are essential (Teoh et al., 2025).

Furthermore, the oral cavity is not an isolated system. Oral microbes can disseminate to other parts of the body and the wider environment. Oral microorganisms can enter the bloodstream through breaks in the oral mucosal barrier, which can occur with conditions like periodontitis, during invasive dental procedures such as extractions or scaling, or even through daily activities like tooth brushing, particularly when gingival inflammation is present (Gendron, Grenier and Maheu-Robert, 2000, Lockhart *et al.*, 2008). Additionally, continuous saliva swallowing provides

a direct route for these microbes to reach the gastrointestinal tract, with around 35% of classifiable oral microbes able to colonise the gut, contributing to at least 2% of the identifiable microbes in faeces (Schmidt et al., 2019). The spread of oral bacteria to other body sites has been implicated in various systemic diseases, including cardiovascular issues like infective endocarditis, IBD, rheumatoid arthritis, and certain cancers (Li *et al.*, 2000). ARG-carrying oral bacteria can make systemic infections more challenging to treat, with documented cases, such as oral streptococci linked to infective endocarditis (Longman et al., 1991). MGEs worsen this risk by enabling the spread of ARGs between microbial communities and environmental microbes (Lunde, Hjerde and Al-Haroni, 2021, Lee *et al.*, 2024, Sukumar *et al.*, 2024).

This potential for oral bacteria, ARGs, and MGEs to disseminate systemically, coupled with external influences on the oral microbiome, firmly places the oral resistome within the One Health framework. Oral ARGs can enter wastewater systems and the broader environment through swallowing, excretion, or expectoration. Conversely, the oral cavity is constantly exposed to environmental ARGs via food, water, and air. Therefore, monitoring the oral resistome is critical to a comprehensive One Health strategy against AMR (Sukumar et al., 2023).

Table 1.3: Antibiotic classes and their corresponding genes and implications

Antibiotic Class	Common ARG Examples	Associated MGE Types	Implicated Bacterial Genera (Examples)	Potential Clinical Significance
Tetracyclines	*tet(M), tet(O), tet(Q), tet(W)*	Transposons (e.g., Tn916-family ICEs), Plasmids, IMEs	*Streptococcus, Actinomyces, Veillonella, Fusobacterium*	Resistance to commonly used broad-spectrum antibiotics in medicine and dentistry.
Macrolides, Lincosamides, Streptogramins (MLS)	*erm(B), erm(F), mef(A), msr(D)*	Plasmids, Transposons, ICEs, IMEs	*Streptococcus, Actinomyces, Staphylococcus*	Resistance to necessary antibiotics for respiratory, skin, and oral infections.
Beta-lactams	*blaTEM, blaCTX-M, mecA*	Plasmids, Transposons	*Streptococcus, Neisseria, Haemophilus*, ESKAPE pathogens	Resistance to penicillins, cephalosporins, and other beta-lactam antibiotics, critical AMR concern.

Aminoglycosides	*aac(6')-aph(2''), aph(3')-IIIa*	Plasmids, Integrons	*Enterococcus*, Gram-negative bacteria e.g. ESKAPE pathogens	Resistance to antibiotics used for serious Gram-negative and some Gram-positive infections.
Fluoroquinolones	*qnr genes, mutations in gyrA/parC*	Plasmid-mediated (qnr), Chromosomal mutations	Various Gram-negative and Gram-positive bacteria	Resistance to broad-spectrum synthetic antibiotics used for a variety of infections.

1.4 Studying Oral Dysbiosis: Methodological Approaches and Their Limitations

To investigate the influence of the HOM and its connection to disease, there has been a need for new methodologies, and it has benefited significantly from the ongoing development from traditional culture-based methods to more complex *in silico* analyses (Do, Devine and Marsh, 2013). However, a common theme is their inherent limitations, meaning that no single approach is perfect and must be carefully considered when designing study protocols, underlining the need for multifaceted approaches.

1.4.1 Traditional Culture-Based Approaches

Traditional microbiology relied on culture-based methods to understand microbes within the human body, including the HOM. Culture-based methods encompass a wide range of protocols, from using non-selective and selective media for microbial growth to developing anaerobic techniques for isolating anaerobic microbes from the oral cavity (Attebery and Finegold, 1969, Krishnan, Chen and Paster, 2017). These allowed for early studies to develop an understanding of oral bacteria's morphological, physiological and biochemical characteristics, as well as the susceptibility of isolates to antimicrobials (Khelaifia et al., 2023). These techniques were crucial in identifying many of the microbes we now know as oral commensals and pathogens, and gave the basis for oral microbiome research (Loos *et al.*, 1993, Lemos *et al.*, 2013).

However, these approaches have many significant limitations. One of the most important limitations is the number of unculturable microbes harboured in the mouth, with estimates that around 34% of oral bacteria cannot be cultured using traditional methods (Wade, 2000, Chen *et al.*, 2010). This is primarily attributed to the fastidious nature of these microbes, requiring specialised media and growth conditions that cannot be easily replicated *in vitro* (Kilian et al., 2016). Consequently, culture-dependent studies of the HOM provided a biased and incomplete representation of the microbial diversity and composition.

Importantly, when deciphering the role of specific microbes in health and disease, the phenotypic expression of these organisms is essential, enabling researchers to identify potential virulence factors and assess host risk. However, the isolation and growth of microbes *in vitro* can alter these

phenotypes, often leading to the expression of genes different from those found *in vivo*, especially those commonly associated with biofilm formation (Donlan and Costerton, 2002). This relates to the lack of community context, which reduces the complexity of microbiomes to a single dimension (Yu et al., 2017). This reduction limits the ability to study inter-species interactions, such as synergistic or antagonistic relationships, as well as the structure of biofilms formed, all of which are crucial for understanding the implications of oral biofilms in health and disease (Darrene and Cecile, 2016). Finally, these traditional methods are often labour-intensive and time-consuming, reducing the scalability for high-throughput analysis (Lagier et al., 2015). The limitations of culture-based methods, particularly the inability to capture the full extent of microbial diversity, directly motivated the development and widespread adoption of culture-independent molecular techniques (Deo and Deshmukh, 2019).

1.4.2 Molecular and Sequencing-Based Approaches

The acknowledged limitations of traditional culture-based methods necessitated the development and refinement of culture-independent techniques. Molecular approaches, primarily targeting microbial nucleic acids (DNA and RNA), have revolutionised the study of the oral microbiome. Key techniques such as the 16S ribosomal RNA (rRNA) gene for bacteria and archaea, or Internal Transcribed Spacer (ITS) region for fungi, allow for assessment of microbial diversity and community composition (Przymus *et al.*, 2025). Shotgun metagenomics provides a further advancement by sequencing all DNA present in a sample, providing insights not only into taxonomic composition but also into the collective functional potential of the community by identifying genes encoding metabolic pathways (Utter et al., 2020). Alongside these DNA-based approaches, metatranscriptomics, which involves sequencing the RNA, has allowed for the study of gene expression within the community, indicating active gene pathways within microbiomes, rather than just insights into potential functions given by DNA sequencing (Solbiati and Frias-Lopez, 2018)

The increased availability of whole genome sequencing, as well as the establishment of digital databases for curation of this sequencing data, such as SRA, ENA and the Human Oral Microbiome Database, has led to a significant increase in the number of studies moving away

from traditional lab-based methodologies to *in silico* techniques (Katz *et al.*, 2022, Yuan *et al.*, 2024). The creation of these repositories has made research easier for researchers. For example, the HOMD, established by Dewhirst *et al.*, (2010) provides a curated, oral-specific database of over 619 oral taxa, based on 16S rRNA gene sequences, offering genomic information for each and analytic tools for other researchers to use, as well as a repository for new studies to upload their data. This standardisation of the HOMD has been crucial for comparing key findings across studies, leading to improved hypothesis generation and biomarker identification. For example, Wang *et al.*, (2025) performed a meta-analysis of 1,255 samples from OSCC-related 16S rRNA gene datasets gathered from the SRA and ENA databases using the Illumina MiSeq platform. They showed that in biopsy samples, the microbiomes of cancerous tissues and adjacent normal tissues were more similar to one another than those of fibroepithelial polyps. They identified key genera and pathways to develop a diagnostic model for OSCC using random forest algorithms on swab samples, with an AUC of 0.918 within these samples and 0.849 when compared to data gathered from the NextSeq platform (Section 2.7.2) (Wang et al., 2025). This demonstrates the potential for using publicly available data for large-scale cohort analysis, reducing the effect of localised variables such as diet, while providing new scope for analysis, including correlations that the primary researcher did not previously conduct, such as those related to the resistome and mobilome.

Despite this, publicly available data from published sources and databases is challenging to access. Many publications demonstrate suboptimal quality and standardisation of associated metadata. Kim *et al.*, (2025) found that within gut microbiome studies, nearly half of the reviewed 3000 studies did not meet the minimum standard for sequence data availability and had poor metadata, lacking key descriptors such as host age and sex for individual samples. This creates a barrier for those wanting to re-analyse data, as key comparisons are unable to be made, due to the lack of transparency and repeatability in these studies, which limits confidence in cross-study comparisons (Dorst et al., 2024). Additionally, as with the work by Wang *et al.*, (2025), the comparisons made between different sequencing platforms may be comparable. However, there are still apparent differences, as quantified by the reduction in confidence when deploying the

same random forest algorithm. This is controllable by selecting datasets only obtained from the same sequencing platform, but this limits the cohort size. At the same time, other parameters are less easily controlled and selected for. For example, DNA/RNA extractions, while similar in methodology, are heavily dependent on the researcher performing them (Nearing et al., 2021). Similarly, when collecting patient samples, it is heavily dependent on the participant and the researcher involved (Roca et al., 2024). These confounding variables can influence results and provide a skewed representation of the true HOM.

Furthermore, whilst the availability of this data is of great importance and usefulness to the oral microbiome research community, resources such as high-performance computer nodes greatly limit the plausibility of large-scale data sets (Do, Devine and Marsh, 2013). Additionally, bioinformatic pipelines require expertise that is not readily apparent to most, with a multitude of pipelines available and little homogeneity between studies, which hampers the ease of utilising this data and its comparability once analysed (Davis-Turak et al., 2017; Do et al., 2013).

1.4.3 Experimental Models

While sequencing-based approaches provide a comprehensive cross-sectional overview of taxonomic and genetic potential, experimental models are essential for longitudinal analysis of interactions with the host, contributing to the understanding of dysbiosis development between health and disease.

1.4.3.1 In Vivo Models

In vivo studies of the oral microbiome encompass both human and animal studies, allowing insights into real systemic host responses. However, they come with significant scientific and ethical hurdles.

Animal models, including rodents, dogs and primates, have been particularly effective in periodontal and caries research. A key example of this is within caries research. Jiang *et al.*, (2023) were able to utilise rat models to demonstrate the ability of *Streptococcus mutans* UA159 to proliferate after inoculation with a three-day high sucrose diet, resulting in damage to deep dentin layers. Animal models, often constructed by inoculating a single pathogenic organism, have

allowed researchers to study specific aspects of oral disease and the effects of individual species on disease progression in live hosts (Li et al., 2025). These models allowed contextualising host immune responses and therefore efficiency of targeted interventions that would be impossible to test directly in humans. However, their use is significantly constrained by several factors. For example, the cost of maintaining these models is significant, requiring specialised animal housing facilities, dedicated veterinary care, and time spent adhering to stringent ethical review processes. Ethical practices in animal models are becoming increasingly pressing, with initiatives such as the 3Rs (Replacement, Reduction, and Refinement) placing pressure on researchers to move away from animal testing where possible (Hubrecht and Carter, 2019). More importantly, there is a significant difference in the makeup of the oral microbiome across host species, limiting the direct applicability of target interventions and findings in these models (Joseph *et al.*, 2021, Šakarnytė *et al.*, 2023). Additionally, the common practice of using a single defined species to represent polymicrobial diseases reduces their complexity and cannot accurately reflect the intricate nature of human oral diseases (Li et al., 2025). Furthermore, as with any *in vivo* study, there is greater inter-individual variability than in vitro studies, making it difficult to fully standardise experiments and to understand the roles of specific microbial or host factors in disease development (Kriaa et al., 2024).

In addition to animal *in vivo* studies, human studies have been conducted, and by their nature, they offer direct relevance to human health and disease. Observational studies have played a crucial role in identifying associations between the microbiome and oral and systemic clinical diagnoses. For example, Al-Maweri *et al.*, (2025) conducted a systematic review of real patient samples from the oral microbiome and hypertension and discovered a significant difference in the oral microbiome taxa between those who were hypertensive and normotensive, despite heterogeneity between the studies. Some microbial taxa, such as *Prevotella* and *Veillonella*, were consistently enriched in individuals with hypertension, as observed in studies that controlled for false discovery rates. Similarly, connections have been established between oral cancer, OSCC, and colorectal cancer through studies conducted within the human oral cavity, providing substantial evidence for the impact of the oral microbiome on systemic diseases (Peng *et al.*,

2022). These have provided essential information for generating hypotheses on the importance of oral health and the effects of poor oral hygiene, both locally and systemically. However, these studies also face limitations. Similar to animal models, stringent ethical review frameworks limits plausibility of these experiments, with ethical conflicts arising from inducing oral dysbiosis in otherwise healthy individuals. The result is that many of these studies are cross-sectional, focusing on the effects of disease after onset, and are descriptive or correlational, lacking the ability to distinguish cause from effect for oral dysbiosis and its associated diseases. This means that these studies often cannot definitively establish causality without additional mechanistic investigations in controlled environments, as is possible in longitudinal, induced dysbiosis studies (Kodikara, Ellul and Lê Cao, 2022, Grieneisen, Blekhman and Archie, 2023). Human studies are further limited by their lack of controlled subjects, with variability present between participant's genetics, diets and other factors (Healey *et al.*, 2017, Vujkovic-Cvijin *et al.*, 2020). Whilst inclusion/exclusion criteria for factors such as smoking status or antibiotic use can help reduce these effects, the reproducibility of these studies remains challenging, with studies often requiring large cohorts to control for these effects and reveal true biological significance, thereby increasing the complexity and overall cost of such studies.

1.4.3.2 In vitro Models

To mitigate the limitations of *in vivo* studies, particularly regarding variable control and repeatability, numerous studies have attempted to recreate the oral microbiome *in vitro*. These studies have been invaluable for discovering pathogenic mechanisms and virulence factors of disease-associated species, such as red complex species (*Porphyromonas gingivalis, Tannerella forsythia*, and *Treponema denticola*), in controlled environments, providing early insights into their potential role in periodontitis (Mohanty et al., 2019). Additionally, these *in vitro* models have allowed insights into the formation of dental biofilms. Sánchez *et al.*, (2011) created simplified models of oral biofilms using six key oral taxa (*Streptococcus oralis*, *Actinomyces naeslundii*, *Veillonella parvula*, *Fusobacterium nucleatum*, *P. gingivalis*, and *Aggregatibacter actinomycetemcomitans*) grown on hydroxyapatite discs coated in salivary pellicles. They demonstrated a pattern of microbial colonisation and biofilm maturation similar to what is

observed in subgingival *in vivo* plaque. Using CLSM, they demonstrated biofilm structures, species distribution within biofilms, and key differences between the upper and lower biofilm, illustrating the sequential colonisation of bacteria as biofilms mature. A key aspect of these models, such as the six-species model, compared to in vivo models, is the defined consortium, which shows high reproducibility, allowing other researchers to utilise it in various studies, including studies of the effects of antibiotic treatment. These studies demonstrate the utility of traditional models in addressing specific questions about oral biofilms and provide reproducible platforms for further investigation.

However, a major drawback of these models is their significant oversimplification, which limits their ability to reflect the complex nature of the HOM accurately. A clear example of this is their drastic lack of microbial diversity compared to *in vivo* samples, which contain just six key bacteria in the model by Sánchez *et al.*, (2011). In contrast, over 700 species are observed in the oral cavity (Chen et al., 2010). Not incorporating nonbacterial species reduces the likelihood of detecting the effects of low-abundant organisms on disease prognosis. This microbial simplification also reduces the ability to study and decipher inter-species interactions such as metabolic cross-feeding, competition and cell-cell signalling, which are key in all microbiomes, not just the oral cavity. However, new developments have incorporated the use of patient-derived inoculum for the production of orally relevant biofilm. Naginyte *et al.*, (2019) developed a model to enrich for periodontal pathogens from samples taken from dentally healthy adults by growing pooled saliva, supragingival plaque, and tongue scrapings as a biofilm under conditions simulating an inflamed periodontal pocket. Metagenomic analysis revealed a significant shift from a health-associated microbial profile to one enriched in taxa implicated in periodontitis, such as *Porphyromonas gingivalis* and *Tannerella forsythia*, supporting the polymicrobial synergy and dysbiosis hypothesis (Naginyte et al., 2019). This study, however, also showed significant differences between the diversity and species present in the samples between inoculum and after incubation, though crucially this included a number of fastidious organisms commonly found in the mouth that were undetectable in the initial inoculum, suggesting increased representation of the oral cavity, and the ability to culture traditionally unculturable oral microbes.

However, the importance of crosstalk should not be omitted, and the model still lacks incorporation of host factors. Within the oral cavity, fluid dynamics play a crucial role in biofilm formation, as saliva and gingival crevicular fluid generate shear forces, buffer pH, and establish nutrient and oxygen gradients. Most *in vitro* studies use static media and conditions for reproducibility. Additionally, studies lack key host factors, such as the absence of host cells in these models, the exclusion of epithelial or immune cells, and the omission of key salivary constituents secreted by these cells, including antimicrobial proteins, mucins, and mechanisms to instigate a host response (Linden *et al.*, 2008, Vila *et al.*, 2019). These host factors are crucial to the formation of dental biofilms *in vivo*, modulating microbial adhesion and pathogenesis, and therefore significantly affect disease outcomes. Nevertheless, the inclusion of these cells also has significant drawbacks for the plausibility of *in vitro* models, as the rapid growth of oral microbes co-cultured with mammalian cells can lead to the accumulation of endotoxins due to prolonged exposure in static conditions, resulting in unrepresentative interpretations of the effect of oral dysbiosis on host responses or even cell death (Kim, 2023). While the insights from these simplified *in vitro* models remain useful, their oversimplification restricts their use as predictive models for *in vivo* comparisons, with their use in testing therapeutic interventions limited without proper physiological context.

To bridge the gap between oversimplified traditional *in vitro* systems and the complex, yet less controllable, *in vivo* environments, there is a clear need to advance *in vitro* models of the oral microbiome. New models should incorporate greater physiological relevance, microbial complexity, and dysbiosis modelling, enabling a deeper understanding of microbiome dynamics in health and disease within a controlled laboratory environment while remaining relevant to real-world applications.

1.4.4 Computational Biology and Machine Learning:

With these sequencing studies, computational bioinformatics is essential for analysing complex ‘omics data. However, the development of machine learning (ML) has led to new approaches being utilised for the discovery of new relationships and interactions in microbiome research. Readily available tools, such as scikit-learn, enable new interpretations of existing microbiome

data, with classifications and regression facilitating the discovery of biomarkers for both health and disease (Pedregosa et al., 2011).

For example, Ma *et al.* (2025) were able to identify the tongue microbiome as a potential marker for the diagnosis of digestive system tumours using models such as Generalised Linear Models, Random Forest, and XGBoost, based on 16S rRNA sequencing from samples of over 1200 patients. They were able to demonstrate the superior performance of classifiers like XGBoost, with an AUC of 0.926, compared to others, such as Random Forest (0.848), suggesting that tongue scrapings may be a non-invasive early biomarker for digestive system tumours. Through traditional analysis, such as LEfSe on taxonomic datasets, they were able to decipher *Alloprevotella* and *Prevotella* as enriched in disease. At the same time, *Neisseria*, *Haemophilus*, and *Porphyromonas* were dominant in healthy controls. However, this was not readily apparent in the machine learning outputs. This study highlights the utility of machine learning (ML) in identifying key microbiomes associated with disease states. However, it often involves the "black box" nature of some complex models, where it is difficult to decipher why or how decisions were made by the algorithms, their biological importance, and the features (species) that were crucial to the decision-making process.

Similar approaches have already been used in the oral cavity. Pais *et al.*, (2025) utilised machine learning for the classification of healthy and peri-implantitis samples. The research benchmarked a variety of ML algorithms, including two AI-driven auto ML approaches that generated 100 distinct models. The results demonstrated that these ML models significantly outperformed traditional single microorganism-based classification methods in predicting PI. Notably, models derived from the saliva microbiome, highlighting *Prevotella salivae*, *Streptococcus sanguinis*, and *Fusarium fujikuroi* as key biomarkers, exhibited superior predictive capabilities compared to those from biofilm samples, where *Neurospora crassa*, *Prevotella nigrescens*, and *Alloprevotella tannerae* were prominent. This work underscores the potential of ML in oral microbiome research by demonstrating its ability to develop more accurate, cost-effective diagnostic tools for complex microbiome-related diseases. The study's findings, such as the effectiveness of classifiers based on combinations of 2-4 microbes and the viability of saliva sampling, demonstrate improved and

earlier disease detection, showing the usefulness of advanced AI techniques with practical clinical applications in fields like dentistry and providing a solid foundation for future investigations (Pais et al., 2025). By limiting the models to key organisms, the study avoided the ‘black box’ by predetermining the species involved and basing interpretations on the best-performing models.

Beyond identifying taxonomic biomarkers, machine learning can also uncover functional signatures within microbial communities. For instance, Veseli *et al.* (2025) developed a classifier for IBD based on 33 metabolic modules related to high metabolic independence in the gut microbiome. This classifier effectively distinguished between healthy individuals and those with IBD, and monitored recovery post-antibiotic treatment, suggesting that high metabolic independence could be a key indicator of microbial stress adaptation. Such an approach, using ML to identify functional microbial markers, is highly relevant to the oral microbiome, where metabolic processes are crucial in determining health and disease states. However, a significant challenge remains the interpretability of these complex ML models. To address this, eXplainable AI (XAI) techniques like SHAP (SHapley Additive exPlanations), which uses game theory to explain the output of any machine learning model by attributing importance values to each feature for a particular prediction (Lundberg & Lee, 2017) and LIME (Local Interpretable Model agnostic Explanations), which explains the predictions of any classifier or regressor by approximating it locally with an interpretable model (Ribeiro et al., 2016) are gaining traction. However, their application in microbiome research remains unexplored. Despite their growing use, a considerable gap remains in the quantitative assessment and validation of explanations generated by these XAI techniques, underscoring the need for further research in this area to ensure the reliability of ML-driven insights in microbiome studies (Saarela & Podgorelec, 2024). This present the challenges in utilising ML for research on the oral microbiome, which coupled with the "black box" nature of some ensemble models, which can be challenging to interpret and the high dimensional nature of microbiome data that can lead to model overfitting highlighting the need for better methodologies in explainable artificial intelligence evaluation (Saarela & Podgorelec, 2024, Przymus *et al.*, 2025)

The multitude of methodological approaches in oral microbiome research underscores that while each technique has significantly advanced our understanding of the HOM, no single method is without limitations. Traditional culture-based methods laid the essential groundwork but were fundamentally restricted by the ability to culture all species and their inability to capture community dynamics. Molecular techniques, particularly high-throughput sequencing, vastly expanded the view of microbial diversity and functional potential, yet brought forth considerable challenges in data management, analysis, and interpretation, alongside issues of standardisation and comparability. Experimental *in vivo* and *in vitro* models, which are crucial for dissecting mechanisms and understanding host-microbe interactions, do not fully account for the complex issues of translatability to the human condition, ethical considerations, and the persistent challenge of balancing experimental control with physiological relevance. Finally, computational and machine learning approaches offer powerful analytical capabilities for deciphering the complexity of microbiome data but also display challenges, particularly concerning the interpretability of models, the need to validate their output (including explanations from XAI), and addressing the unique characteristics of microbiome datasets.

These individual shortcomings, when considered collectively, highlight a critical overarching theme: the imperative need for integrated, multifaceted research strategies. To capture the true complexity of the HOM in states of both health and disease, and to translate fundamental discoveries into meaningful clinical applications, future research must increasingly adopt approaches that combine the strengths of these diverse methodologies to mitigate their respective weaknesses.

Chapter 2: Materials and Methods

2.1 Search of Public Data Repositories

The identification of publicly available data (PAD) was conducted via a search for shotgun metagenomic sequencing from published articles relating to the oral microbiome in health and disease, with two cohorts, one for the comparison of caries and caries-free (health), as well as the comparison of periodontitis and periodontitis-free (health). Data was found and selected for via manual searches of PubMed for studies corresponding to the following key words- "((‘caries’) OR ('periodontitis')) AND (('shotgun') OR ('metagenome') OR ('microbiome') OR (‘metagenomics’)" as well as the European Nucleotide Archive with search terms ‘caries oral microbiome’ and ‘periodontitis oral microbiome’ (Burgin *et al.*, 2022; *The Sequence Read Archive (SRA)*, 2025). Studies appearing in search results were then manually filtered for the following criteria: Oral Sample Type: plaque (either supra– and sub- gingival plaque) AND Sequencing Platform: Illumina. Additionally, demographic metadata filtering was applied to ensure the caries cohort and the periodontitis cohort, along with their matching controls, consisted of children (<18 years old) and adults (>18 years old) respectively. Studies which utilised multiple cohorts e.g. systemic disease or disease intervention, only samples from those with systemic health and no interventions were selected. This allowed for the control of variables arising from differences in oral cavity niches as well as types of sequencing technologies. Additionally, studies that passed these criteria must have had sequencing data readily available from the SRA repository, with the minimum metadata available set at the identification of disease status for individual samples.

2.2 Download and Preprocessing of PAD

Subsequent retrieval and alignment of PAD was conducted on the High Performance Computing (HPC) facilities at the University of Leeds, using ARC4/AIRE nodes (Advanced Research Computing, Leeds.) Once studies were identified, PAD was acquired using the SRA toolkit (version 3.0.1-ubuntu64), with prefetch used to download sequencing data as SRA files. Following this, paired-end fastq files were extracted using fasterq-dump. The paired-end files were then processed with cutadapt (v3.5). Initially, Illumina Universal adapter sequences were

5’-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3’ and 3’-TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG-5’. After adapter trimming, quality filtering was performed using a Phred score threshold of 10, trimming reads below this from both ends (Martin, 2011). Additionally, any reads shorter than 30 bp were discarded after adapter trimming. The output paired-end reads from cutadapt were subsequently assembled de novo into contigs using MEGAHIT (v1.2.9) with default parameters (Li et al., 2015).

2.3 Taxonomic Analysis

2.3.1 Custom Oral Database

To establish a baseline for the taxonomic profiling, MetaPhlAn4 (v4.0.6) was run on the cleaned paired end fastq files, against the inbuilt reference database. The analysis was run with default parameters, generating a taxonomic abundance table for each sample. This initial analysis provided a robust, bacteria-focused community profile, which served as a baseline before the more comprehensive, multi-domain analysis was performed with the custom in-house database. To facilitate this comprehensive taxonomic profile beyond the typical bacterial analysis, an in-house pipeline was developed. The primary objective of this approach was to overcome the inherent prokaryote limitations of standard microbiome databases. This was achieved by constructing and using a custom protein database that explicitly includes viral, archaeal, and eukaryotic organisms previously identified as constituents of the human oral cavity. This bespoke protein database was constructed utilising the makedb function within the DIAMOND software (v2.0.15) (Buchfink et al., 2015). The database comprises a curated collection of protein sequences from multiple key sources, using whole reference genomes from NCBI. The database consists of all microbial protein sequences from the Human Oral Microbiome Database (HOMD) (https://github.com/ll16jjpl/PhD_scripts_JL) (Chen et al., 2010). This was supplemented with protein sequences from archaea, viruses, and eukaryotes demonstrated to colonise the human oral cavity, as detailed in previous reports (Ghannoum *et al.*, 2010, Baumgardner, 2019, Belmok *et al.*, 2020, Dame-Teixeira *et al.*, 2025). To ensure a thorough representation of the oral virome, the viral dataset incorporated sequences from both human-infecting viruses and bacteriophages,

thereby capturing the full spectrum of viruses within the HOM (Martínez, Kuraji and Kapila, 2021, Ruiz-Mojica and Brizuela, 2025).

2.3.2 Contig Annotation and Filtering

Following sequence alignment of metagenomic contigs against the custom multi-domain database, a post-processing filtering step was implemented to enhance the accuracy of taxonomic assignments. For this purpose, a custom Python script was developed (https://github.com/ll16jjpl/PhD_scripts_JL) to filter the alignment results and select the most representative hits for any given region of a contig. To mitigate overestimation of taxonomic assignments arising from redundant or non-specific alignments, the script first sorts all alignments for a single contig by start and end coordinates, then sorts them by E-value. This process grouped all co-localised hits. The script then iterated through these sorted alignment groups. For any set of alignments covering a similar region, defined as having start and end coordinates within a 10-base-pair tolerance, the script retained only the single alignment possessing the lowest associated E-value, indicative of the highest alignment confidence. Alignments that did not exhibit positional overlap with other hits were preserved as unique annotations. Following this, gene names, species, and full taxonomic classifications for the retained DIAMOND gene alignments were retrieved using Entrez Direct command-line tools (Kans, 2025). This step provided comprehensive annotation for subsequent analysis of both taxonomy and function.

2.3.3 Quantification and Count Matrix Production

A subsequent Python script was utilised to process the filtered alignment files for each sample and generate a unified count matrix. Initially, alignments were subjected to a stringent high-confidence filter, retaining only those that met or exceeded a 99% sequence identity and 97% query coverage threshold. For each sample, a relative abundance value for every unique gene (sseqid) was calculated based on the cumulative length of its high-confidence alignments relative to the total length of all high-confidence alignments in that sample. The script processed all samples in parallel and merged the results into a single data matrix, with genes, strain and full taxonomy as rows and samples as columns. To generate a count table suitable for downstream statistical analysis with count-based models, these relative abundance proportions were scaled by

a factor of one million and rounded to the nearest integer. The final output used a comprehensive integer count table representing the abundance of each gene, as well as summed species and taxonomy across all samples.

2.4 Resistome Analysis

2.4.1 Resistome Profiling

To characterise the oral resistome, cleaned paired-end FASTQ files were aligned against the Comprehensive Antibiotic Resistance Database (CARD, v4.0.0) using the DIAMOND aligner (v2.0.15) (Alcock et al., 2023). A multi-step filtering process was applied, beginning with the same contig-level script used for taxonomic analysis (Section 2.3.2) to resolve multiple alignments. Subsequently, alignments were further filtered to retain only those meeting criteria of ≥90% sequence identity and ≥80% query coverage. The abundance of the remaining ARGs was quantified and normalised to Reads Per Kilobase of transcript per Million mapped reads (RPKM) using a custom script. A final count table was annotated with gene name, resistance mechanism, and drug class information obtained from the CARD database.

2.4.2 Resistotyping Analysis

To investigate whether ARGs organised into discrete profiles, a stepwise resistotyping analysis was performed. The core method for each involved converting ARG counts to a binary presence/absence matrix, calculating Jaccard distances, and performing PCoA. Resistotypes were then defined by applying hierarchical clustering (Ward's method) to the Euclidean distances of the resulting PCoA scores. The optimal number of clusters for each analysis was determined by maximising the average silhouette width. To visualise these defined clusters, the samples were plotted on a PCoA plot, with 95% confidence ellipses constructed around each assigned cluster using the stat_ellipses function in ggplot. The analysis was conducted in several stages. First, independent disease-specific analyses were performed on the periodontitis cohort versus its matched healthy controls, and separately on the dental caries cohort versus its controls. Next, to assess potential confounding effects, a cohort validation step was conducted via a direct resistotyping comparison between the healthy from the caries cohort and the healthy from the

periodontitis cohort. Following this validation, a comprehensive analysis was performed by merging all samples from the health, caries, and periodontitis groups into a single dataset to identify global resistotypes across the entire study population. For each of these comparisons, the statistical significance of the association between the identified resistotypes and clinical status was evaluated using a Chi-squared test or Fisher's Exact Test where appropriate. Pairwise PERMANOVA was also used to assess beta-diversity differences between the groups.

2.5 Mobilome Analysis

2.5.1 Plasmid and HGT Profiling

Plasmid content within the assembled contigs was characterised using PLASMe (v1.1), a tool leveraging an extensive database of known plasmids (X. Tang et al., 2023). To ensure high confidence in plasmid identification, PLASMe incorporates stringent filtering criteria, requiring 90% query coverage and 90% sequence identity for alignments. The resultant outputs were subsequently processed with custom Python scripts (https://github.com/ll16jjpl/PhD_scripts_JL) to format the data appropriately for downstream analyses. Following plasmid identification, the protein-coding gene content of these plasmids and their putative species of origin were determined by querying the NCBI nucleotide database via Entrez Direct.

To investigate HGT events, the assembled contigs were analysed with the WAAFLE tool (v1.1)meta, which utilises the ChocoPhlan pangenome database (Hsu et al., 2023). It identified sections of single contiguous DNA sequences where genes in the database belonging to species B have transferred into species A, e.g. a sequence containing genes AABBAA would be inferred as species B having donated two genes to species A. The resulting WAAFLE outputs were processed using a custom Python script (https://github.com/ll16jjpl/PhD_scripts_JL) to quantify HGT events at both taxonomic and functional levels. The script systematically processed each event, extracting genus and species-level taxonomy for the participating clades. For events with a resolved transfer direction (A>B or B>A), the script tallied counts for donor and recipient taxa. Functional annotation was performed by extracting UniRef90 gene identifiers associated with each event. WAAFLE provides annotation using both UniRef90 and UniRef50, however,

UniRef90 was chosen for analysis due to its more stringent criteria, which more closely reflect the analysis done elsewhere in this project and give higher confidence in species and genes involved. For directed transfers, the script utilised the synteny information provided by WAAFLE to attribute each gene to either the donor or recipient, allowing for the specific quantification of transferred genes. This process generated multiple distinct count matrices, which were subsequently used for differential abundance analysis. To further interpret the functional implications of these transferred genes, Gene Ontology (GO) pathway analysis was performed on the UniRef90 gene identifiers, using the UniProt id mapping service to gather GO pathways from these gene ids to annotate previously created count tables, to elucidate potential functional alterations conferred by HGT (Apweiler et al., 2004).

2.6 Statistical Analysis

2.6.1 Diversity Analysis

All subsequent statistical analyses and data visualisation were performed in the R environment (version 4.2.1) using the integer count tables generated in the previous steps. The phyloseq package (v1.50.0) was utilised for the microbiome data (McMurdie and Holmes, 2013). Alpha diversity was quantified using four standard ecological metrics. Observed species and Chao1 were used to estimate taxonomic richness, while the Shannon and Simpson indices were calculated to assess both richness and evenness, measures of how many species and how equally distributed these species are, respectively. Beta diversity was evaluated to compare the compositional structure between samples. This was achieved by calculating Bray-Curtis dissimilarity, a metric sensitive to taxon abundance, and Jaccard distance, a metric based on the presence or absence of taxa. The resulting distance matrices were visualised using PCoA to ordinate the samples in a two-dimensional space. Samples in the PCoA plots were coloured by their respective disease status, with 95% confidence ellipses constructed around each group using multivariate t-distributions to illustrate clustering using the stat_ellipses function in ggplot2.

2.6.2 Differential Abundance Analysis

To identify features with statistically significant differences in abundance between clinical conditions, differential abundance analysis was performed using the DESeq2 package (v1.46.0) in R (v4.4.1) (Love et al., 2014). The core analysis was conducted using the DESeqDataSetFromMatrix function, with a design formula based on the clinical status. Features with zero counts across five or more samples were removed. The DESeq function was then run, utilising the poscounts size factor estimation method, which is robust for sparse metagenomic datasets. Results for the contrast of interest were extracted, and p-values were adjusted for multiple testing using the Benjamini-Hochberg method. Features were considered significantly differentially abundant if they exhibited an adjusted p-value (FDR) < 0.05 and a log2 fold change > 1. The results were visualised using volcano plots generated with the ggplot2 (v4.0.0) and ggrepel (v0.9.6) packages. This consistent analytical approach allowed for direct comparison of diversity patterns and differentially abundant features across the taxonomic, resistome, and mobilome components of the oral microbiome.

2.7 Machine Learning for Disease State Classification

2.7.1 Data Preparation for Machine Learning

The count tables generated from the taxonomic, resistome, and mobilome analyses served as the feature sets for machine learning model development (https://github.com/ll16jjpl/PhD_scripts_JL). A multi-faceted approach was designed to compare the predictability of the two distinct disease states, dental caries and periodontitis. For each data type, such as Species or ARGs, three separate analyses were conducted, a binary classification for Caries vs. Health using data from the caries cohort, a binary classification for Perio vs. Health using data from the periodontitis cohort, and a multi-class classification for Caries vs. Perio vs. Health using a merged dataset. This joint analysis was created by combining data from both projects, and features were aligned by taking the features of those present in both datasets. For each analysis, the corresponding count table was transposed so that samples were represented as rows and features as columns and then merged with the study metadata file to assign the clinical status as the target variable. Categorical target variables were numerically encoded, with "Health"

consistently assigned as the negative class (0) in binary classifications. Each complete dataset was then partitioned into a training set (80%) and a test set (20%) using a stratified split to ensure that the class distribution was preserved in both sets. This test set was reserved for the final evaluation of the best-performing models and was not used during any part of model training or selection.

2.7.2 Model Training and Evaluation

A comprehensive evaluation of various machine learning classifiers was conducted. The classifiers included Logistic Regression, Stochastic Gradient Descent (SGD) Classifier, Support Vector Classifier (SVC), K-Neighbors Classifier, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier, AdaBoost Classifier, Gaussian Naive Bayes, and a Multi-layer Perceptron (MLP) Classifier, all implemented using Scikit-learn (1.4.1.post1) in Python (Pedregosa et al., 2011). The primary metric for model evaluation was the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), calculated using a 5-fold cross-validation on the training dataset. AUC are calculated from the area under the ROC, a graph used to measure the true positive rate against the false positive rate, with a score of 1.0 being perfect classification (Mathew et al., 2023) For multi-class problems, such as the joint analysis, a one-vs-rest approach with weighted averaging was used for AUC-ROC calculation. This method train individual binary classifiers for each class, e.g periodontitis against the rest of the classes, with a final AUC-ROC calculated from the weighted averages of these classifiers (Mathew et al., 2023).

2.7.3 Preprocessing Scenarios

To optimise performance, each classifier was evaluated under several data preprocessing scenarios constructed as pipelines. The baseline 'Native' scenario involved only standard feature scaling. A 'FeatureFilter' scenario applied a variance threshold to remove low-variance features. To address class imbalance, a 'SMOTE' scenario was implemented using the Synthetic Minority Over-sampling Technique (Chawla *et al.*, 2002). A 'PolyFeatures' scenario was also evaluated to capture non-linear relationships by generating second-degree polynomial and interaction features. Additional scenarios consisted of combinations of these techniques, such as 'FeatureFilter and SMOTE'.

2.7.4 Classifier Optimisation and Selection

The evaluation was conducted in two phases. First, an initial screening was performed in which all classifiers were evaluated across all preprocessing scenarios, using both default and predefined 'optimised' parameters to establish a broad performance baseline. Following this, a rigorous hyperparameter tuning phase was conducted for the top five performing model-scenario combinations, as ranked by their mean cross-validated AUC-ROC. GridSearchCV with 5-fold cross-validation was used to search a comprehensive parameter grid for each candidate. A key step in our methodology was the final selection process. The performance of the best model found during the grid search was compared with that of its pre-tuning counterpart. The model with the higher mean cross-validated AUC-ROC was designated as the definitive "best model" for that configuration and was carried forward for final assessment.

2.7.5 Final Model Assessment and Statistical Comparison

The performance of the selected best models was conclusively assessed on the 20% test set to provide an unbiased estimate of generalisation capability, quantified by the test set AUC-ROC. Confusion matrices and ROC curves were also generated. To rigorously compare the top-performing models within each analysis, a formal statistical framework was employed. A one-way ANOVA was first used to determine if any significant performance differences existed among the top models. Subsequently, a Tukey's HSD post hoc test was performed to identify which models were statistically better than others ($p < 0.05$). Finally, to directly address whether one disease state was more predictable than the other, a cross-project statistical comparison was performed. For each data type, the single best-performing model from the Caries analysis was compared with the best-performing model from the periodontitis analysis using a Wilcoxon signed-rank test on their cross-validation scores.

2.7.6 Deep Learning Model Development and Optimisation

In addition to classical machine learning models, a deep learning approach using a neural network was developed and optimised for disease classification. The entire deep learning pipeline, from data preprocessing to hyperparameter tuning and feature importance analysis, was run in Python

(v3.11.9) using the PyTorch library (2.8.0+cu126), using GPU functions (https://github.com/ll16jjpl/PhD_scripts_JL) (Paszke et al., 2019). A fixed random seed (42) was used throughout all stages to ensure full reproducibility of the results.

2.7.6.1 Deep Learning Model

The neural network consisted of an input layer with a size corresponding to the number of features, followed by two fully connected hidden layers. The first hidden layer was followed by a batch normalisation layer, which standardises the inputs to a layer for each batch to stabilise and accelerate training. This was followed by a ReLU function, which introduces non-linearity by outputting the input directly if positive and zero otherwise (Przymus et al., 2025). This, in theory, helps mitigate the vanishing gradient problem (Oh et al., 2023). Finally, a dropout layer was included, which acts as regularisation by randomly setting a fraction of outputs to zero during training to prevent overfitting (Srivastava et al., 2014). The second hidden layer had a similar structure of batch normalisation, ReLU activation, and dropout. The final output layer was a single linear neuron that produced raw logits. The model was trained using a binary loss function with logits to calculate errors in binary classifications.

2.7.6.2 Model Training and Optimisation Workflow

The model's performance was optimised through a multi-stage process. An initial phase evaluated the model's baseline performance with and without two key preprocessing steps, as with the classical ML classifiers, using low-variance feature filtering (threshold of 0.05) and SMOTE to address class imbalance. The model was trained using the Adam optimiser, an adaptive learning rate algorithm well-suited for a wide range of problems, with a 5-fold cross-validation (Kingma and Ba, 2017). Training was conducted over multiple epochs, where each epoch consisted of a full pass through the entire training dataset. To prevent overfitting and select the best-performing model, early stopping with 10-15 epochs was employed, halting training when the validation set AUC-ROC did not improve.

Following the initial screening, hyperparameter optimisation was performed using Optuna (v4.5.0) (Akiba et al., 2019). This framework employs a TPE sampler, a Bayesian optimisation

algorithm that explores the hyperparameter space by building a probabilistic model to predict which parameters are most likely to yield better performance.The search was conducted for 100 trials to identify the optimal combination of learning rate, the number of neurons in each of the two hidden layers, and the dropout rate, with the objective of maximising the validation AUC-ROC on a 20% split of the training data.

2.7.6.3 Final Evaluation and Feature Importance

The best set of hyperparameters identified by Optuna was used to train a final model on the full training dataset, again using a 5-fold cross-validation approach to ensure robustness. The performance of this fully optimised model was then evaluated on the test set. To identify the most influential features for the model's predictions, permutation feature importance was calculated. This technique assesses a feature's importance by measuring the decrease in model performance (AUC-ROC) after randomly changing the values of a single feature in a validation set. Features that cause a larger drop in performance are considered more important to the model's predictions. The top 20 most impactful features were then identified and reported.

2.8 Host-Biofilm Co-Culture Model

2.8.1 Stratified Epithelial Collagen Gel Model

A stratified epithelial collagen gel model was established, adapted from the methods of Hewitt *et al.*, (2022), who used a custom-fabricated Buoyant Epithelial Culture Device (BECD), floating culture devices to keep cells at the air liquid interface. In brief, a collagen solution was prepared on ice by mixing rat tail collagen I with 10x DMEM (Sigma, Missouri, US) and a reconstitution buffer (1.1 g $NaHCO_3$ (Sigma) and 2.385 g HEPES (Sigma) in 50 mL dH2O) and neutralised to pH 6.9 with 1M NaOH (Sigma). hTERT-immortalised gingival fibroblasts (ATCC, Virgina, US) were suspended in Fibroblast Basal Medium (ATCC) supplemented with a Fibroblast Growth Kit-Low Serum (ATCC) and incorporated into the collagen solution to a final concentration of 5×10^4 cells/mL. 1.1mL of this cell collagen mixture was placed into sterile polypropylene BECDs and allowed to polymerise for 90 minutes at 37°C. The resulting fibroblast gels were cultured submerged using submerging rings in the fibroblast growth medium for 24 hours to allow for

initial cell growth. Following this, dysplastic oral keratinocyte (DOK) cells were seeded onto the gel surface at a concentration of 8×10^4 cells/mL (Chang et al., 1992). Upon seeding, the culture medium was changed to DMEM supplemented with 2mM L-glutamine (Merck, Darmstadt, Germany), 5µg/ml hydrocortisone (Gibco, Massachusetts, US), and 10% Foetal Bovine Serum (FBS) (Sigma). The models were initially cultured in a submerged state for 72 hours at 37°C and 5% CO2. Following this submerged culture period, the models were raised to the air-liquid interface, by removing submerging rings, to promote epithelial stratification and differentiation. The constructs were cultured at the air-liquid interface for a further 10 days with the media changed every two days to ensure cell viability and promote epithelial development. Due to staggered timepoints in the co-culture, to ensure the epithelial layers were of consistent maturity, the initial seeding of fibroblast cells and DOK cells was staggered. This staggered workflow guaranteed each stratified epithelial layer was exactly 10 days post-air-liquid interface establishment when introduced to the specific complex biofilm.

2.8.2 *In vitro* Biofilm Model

Ethics approval for oral sample collection was granted by the University of Leeds Dental Research Ethics Committee (DREC 150322/JL/349), and consent was obtained from all participants. Supragingival plaque, biofilm from the tongue dorsum, and unstimulated saliva were collected from four dentally healthy volunteers (Donors A, B, C and D) and subsequently anonymised. To generate three distinct biological replicates, these samples were pooled into unique combinations utilising three donor samples per replicate (e.g ABC. BCD, ACD) for each inoculum. Volunteers were required to have no current treatment for dental caries or periodontal disease, and to have not used antibiotics in the previous three months. Participants refrained from tooth brushing for 16 hours and from eating or drinking for 1.5 hours prior to sample collection. Plaque was collected into a sterile, pre-reduced protein-rich medium, and at least 5 mL of saliva was collected by drooling into a sterile tube. All samples were transferred to an anaerobic workstation within 60 minutes. The plaque and saliva samples were homogenised and pooled under anaerobic conditions to create inoculums.

Biofilms were cultivated on 7mm hydroxyapatite discs within 48-well plates (Corning, New York. Prior to inoculation, the discs were preconditioned for 5 hours with 200 µL of sterile human saliva to form a salivary pellicle. Each well was then inoculated with a total volume of 500 µL, consisting of 227.5 µL pooled saliva, 136.25 µL pooled supragingival plaque, and 136.25 µL pooled tongue biofilm and allowed to grow for 24 hours prior to the first media change. The plates were incubated anaerobically (10% H_2, 10% CO_2, 80% N_2) at 37°C in an anaerobic workstation (Don Whitley, Shipley, UK). To establish a 'healthy representative' biofilm, the cultures were maintained for an initial period of 14 days in a serum-free, protein-rich medium, with the medium changed twice a week.

2.8.3 Co-Culture Design

Following the initial 14-day growth period to establish a mature, healthy state biofilm, a dysbiotic shift was induced by replacing the serum-free medium with the same protein-rich medium supplemented with 20% (v/v) foetal bovine serum to simulate conditions associated with periodontitis. Biofilm hydroxyapatite discs were then collected at multiple time points: Day 0 (representing the mature healthy biofilm), Day 2, Day 5, Day 8, Day 11, and Day 14. At each time point, two discs from each biological replicate were transferred into separate stratified epithelial collagen gel cultures and co-cultured for 24 hours in a 5% CO_2 atmosphere in a 24-well plate containing 2.5mL DOK growth media. The overall model timeline is outlined in Figure 2.1.

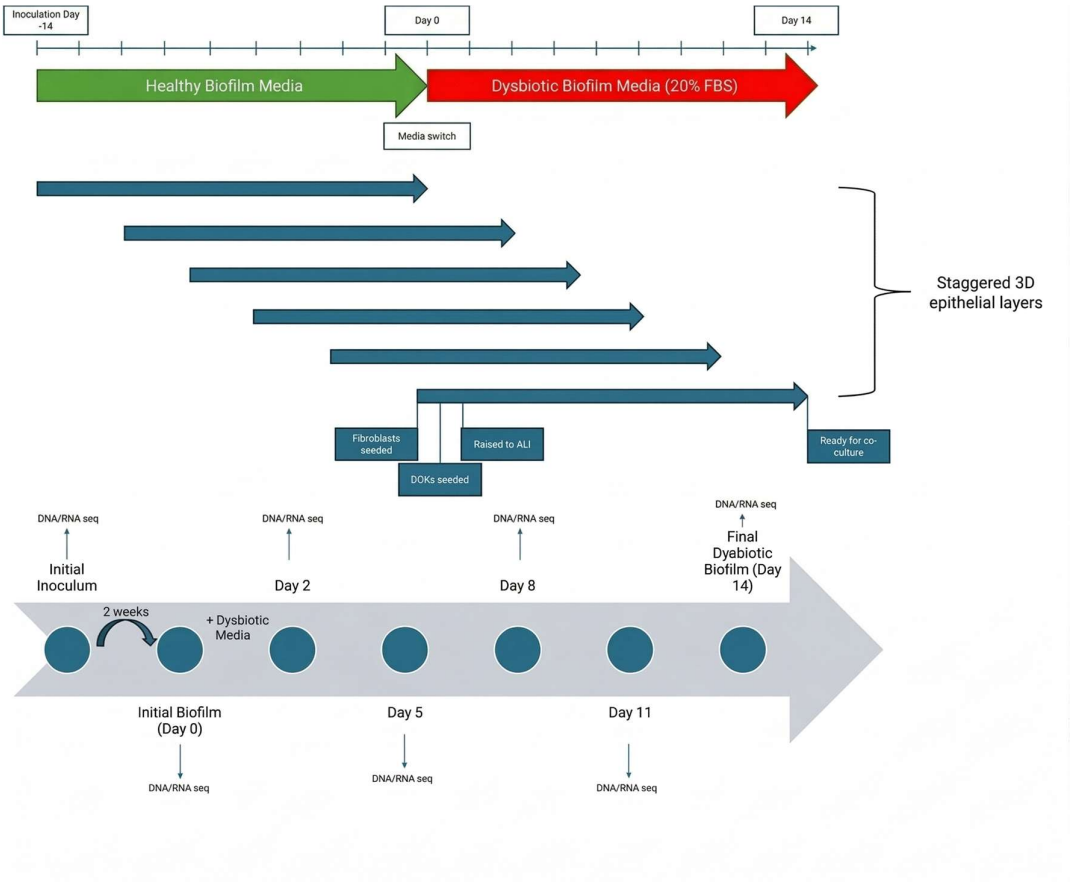


Figure 2.1 Experimental timeline detailing co-culture model design outlining biofilm dysbiosis and staggered epithelial cell culture.

Diagram of the experimental design model of periodontitis dysbiosis progression and host response. Outline of continuous biofilm growth, with a 14-day health associated media, followed by a switch at D0 to dysbiosis inducing media. The staggered preparation of 3D epithelial tissues ensures all host tissues reach the same physiological age for the 24-hour co-culture. DNA and RNA sequencing was performed at all timepoints of co-culture, D0, D2, D5, D8, D11 and D14.

2.8.4 Nucleic Acid Extraction and Sequencing

After the co-culture period, both collagen and biofilms were stored separately in 1mL RNAlater (Sigma) until the end of the full experimental procedure. Subsequently, total DNA and RNA were extracted from the samples. For DNA extraction from biofilms, the PureLink™ Microbiome DNA Purification Kit (Invitrogen, Massachusetts, US) was used following the manufacturer's saliva protocol. This procedure uses a triple lysis approach, combining heat, chemical lysis, and mechanical disruption via bead beating to ensure efficient breakdown of cell walls and maximise DNA extraction. Purified DNA was bound to a silica spin column, washed, and eluted in 100 µL elution buffer.

For RNA extraction, the PowerFecal RNA extraction kit (Qiagen, Hilden, Germany) was used. This protocol used mechanical lysis via bead beating in a lysis buffer containing β-mercaptoethanol to denature and inactivate RNases, preventing RNA degradation. To ensure high purity, the protocol included an on-column DNase I digestion step to remove any contaminating genomic DNA before the final RNA elution in 100 µL RNase-free water. RNA extracted from the host collagen gel and the corresponding biofilm were pooled at a 19:1 host-to-biofilm ratio prior to sequencing.

Both the extracted DNA and the pooled RNA samples were sent to BMKGENE (https://www.bmkgene.com, Beijing, China) for next-generation sequencing. DNA samples underwent library preparation, including fragmentation, end-repair, A-tailing, and adapter ligation, for shotgun metagenomic sequencing. RNA samples were first subjected to rRNA depletion to enrich for mRNA, which was then converted to cDNA for library preparation. All libraries were sequenced on the Illumina NovaSeq platform (Illumina, California, US) using PE150, with a target data output of 10 Gb per sample and a quality standard of ≥85% of bases with a Phred score of 30 or higher.

2.8.5 Downstream Bioinformatic and Machine Learning Analysis

The metagenomic and metatranscriptomic data generated from the co-culture model were processed using the same bioinformatic pipelines established for the analysis of the publicly

available data, as detailed in sections 2.3, 2.4, and 2.5. This included taxonomic, resistome, and mobilome profiling to generate corresponding count tables. For RNA sequencing, adapter removed paired end fastq files we processed to separate host and microbial data using Bowtie2 (v2.4.4.). Reads were aligned to host reference genome GRCh38, reads that aligned were retained for host analysis while unaligned reads were retained for analysis as outlined in sections 2.3, 2.4, and 2.5. For the host reads, STAR aligner (v2.7.9) was used with default parameters with quantMode GeneCounts to generate count table of Ensembl gene IDs.

To annotate host IDs, R package biomaRt (v2.61.1) was used to map Ensembl IDs to corresponding HGNC gene symbols, gene descriptions, and functional pathway information from the Gene Ontology (GO) and Reactome databases, with a consolidated count table generated for each. For Reactome pathways, Reactome.db R package (v1.89.0) was utilised to convert pathway IDs to their descriptive names. This created summary count tables that reflected the overall transcriptional activity of entire biological pathways, which were then used for subsequent analysis.

To identify features with significant changes in abundance, two differential abundance approaches were employed using DESeq2. First, pairwise comparisons were conducted between specific time points (e.g., Day 14 vs. Day 0) to identify differentially abundant features, defined as those with an adjusted p-value (*padj*) < 0.05 and log2 fold change > 1.0. Second, to capture longitudinal trends, a Likelihood Ratio Test (LRT) was performed across the entire experimental time course to identify any feature that changed significantly over time. The results were used to generate heatmaps of the top changing features.

.

Chapter 3: Analysis of Publicly Available Sequencing Data

3.1 Introduction

The human oral cavity, first described as microbial by van Leeuwenhoek in the 17th century when he viewed dental plaque under a microscope, is now understood to be a vast and dynamic ecosystem, second only to the gut in its diversity (Christersson, Zambon and Genco, 1991, Deo and Deshmukh, 2019). The HOM, comprising a complex community of bacteria, archaea, fungi, and viruses, exists in a state of balance, or eubiosis, in healthy individuals (Rajasekaran et al., 2024). It plays a crucial role in maintaining homeostasis, contributing to host defence by preventing the colonisation of potentially pathogenic microbes and modulating local immune responses (Lamont et al., 2018). This equilibrium, however, can be disturbed with disruptions to the community, when influenced by factors such as diet, host genetics, and lifestyle choices, which can lead to a state of microbial imbalance known as dysbiosis (Maier, 2023). Dysbiosis is recognised as a primary etiological factor in the development of the world's most prevalent oral pathologies, namely dental caries and periodontal diseases (Santacroce et al., 2023).

Dental caries and periodontitis represent two distinct ecological shifts. Caries is mainly caused by a metabolic change in which frequent exposure to fermentable carbohydrates, such as sucrose, creates an acidic environment that selects for a limited number of acidogenic and aciduric species, such as Streptococcus mutans, leading to further acid production and enamel demineralisation (Zhang et al., 2022). Periodontitis, conversely, is an inflammatory disease marked by a shift towards a proteolytic, anaerobic consortium of pathogens, including the keystone "red complex" bacteria (*Porphyromonas gingivalis, Tannerella forsythia*, and *Treponema denticola*), which thrive in the inflammatory environment they established (Mohanty et al., 2019). A new principle emerging in the stability of the healthy oral microbiome is functional redundancy, where multiple, phylogenetically distinct taxa can perform similar metabolic functions, providing a buffer against environmental changes (Benítez-Páez *et al.*, 2014, Wei *et al.*, 2021). The transition to dysbiosis can be viewed as an overcoming of this redundancy, making the ecosystem more susceptible to pathogenic colonisation (Rajasekaran et al., 2024).

While the taxonomic and broad functional shifts associated with oral dysbiosis are well-documented, a more focused understanding requires moving beyond simple identification of the microbes involved. The collective genetic potential of the microbiome, particularly its capacity for AMR and gene mobility, represents a critical but often understudied dimension of oral health and disease. Biofilms are a known hotspot for HGT, the process by which bacteria exchange genetic material, facilitated by a diverse array of mobile genetic elements such as plasmids and conjugative transposons (Abe, Nomura and Suzuki, 2020, Michaelis and Grohmann, 2023). This exchange enables the dissemination of MGEs, allowing for gain of new functions, most notably AMR genes, which can increase the clinical challenge of treating oral infections (Lerminiaux and Cameron, 2019). Within the HOM, this is of paramount importance, as common oral diseases are plaque-driven and dental plaque consists of microbial biofilms.

Furthermore, the oral cavity is not an isolated ecosystem. The One Health concept, which recognises the interconnectedness of human, animal, and environmental health, provides an essential framework for understanding the broader implications of oral dysbiosis (Moussa et al., 2025). There is now compelling evidence that the oral microbiome serves as a significant reservoir for the gut microbiome, with constant translocation of oral microbes occurring via swallowing (Kunath et al., 2024). Similarly, routine daily activities like toothbrushing can cause transient bacteremia, providing a direct route for oral microbes and their MGEs to enter systemic circulation and potentially colonise distant body sites (Lockhart et al., 2008). These connections position the oral cavity as an important area of research, capable of both receiving genetic material from the wider environment and disseminating its own resistome and mobilome throughout the host, with potential implications for systemic health and the global AMR crisis.

Despite this growing recognition, a comprehensive, integrated analysis of how the oral microbiome, resistome, and mobilome are collectively reshaped between health and the distinct dysbiotic states of caries and periodontitis is lacking. This research study was undertaken to address this gap by collating and standardising publicly available oral metagenomic data. The primary aims were to: (1) Find, curate, and standardise public oral microbiome sequencing data,

(2) Compare the microbial community in health versus the dysbiotic states of dental caries and periodontitis, and finally, (3) identify specific taxa, ARGs, and MGEs associated with dysbiosis.

3.2 Results

3.2.1 Collation of an Oral Metagenomic Cohort

To construct a large and diverse dataset for this investigation, a systematic search for publicly available data (PAD) was conducted. Public repositories, including PubMed and the ENA, were searched using keywords such as (‘caries’ OR 'periodontitis') AND ('shotgun' OR 'metagenome') to identify relevant studies. The search was designed to assemble two primary cohorts, with one comparing dental caries and another comparing periodontitis against their respective healthy controls. Studies identified through this search were then manually filtered for inclusion, requiring that they utilised Illumina shotgun sequencing of dental plaque (supragingival or subgingival) and had raw data available in the SRA with sufficient metadata to assign a clinical status to each sample. This filtering process yielded a total of 10 studies that met the inclusion criteria. Following quality control and processing, a final cohort was assembled for downstream analysis, comprising 66 caries-free, 95 caries, 151 periodontitis, and 80 healthy samples. A summary of the included studies is provided in Table 3.1. A review of the metadata from the selected studies highlights the variations in reporting and data availability within the cohorts. Within the caries dataset, disease classification methodologies vary, employing different clinical metrics such as ICDAS or DMFT. Available metadata varied also, with some studies such as dietary sugar frequency and oral hygiene habits only recorded in some studies (e.g. PRJNA766357). Differences in DNA extraction was also apparent, ranging from commercially available kits, CTAB and phenol/chloroform extraction. Similarly, within the periodontitis cohort these variations could be observed, particularly for disease classification. Whilst most studies used pocket depth, attachment loss and bleeding on probing to some degree, the threshold for disease classification of these was often not reported (e.g. PRJNA528558). Unlike the caries cohort, DNA extraction methods were more uniform, with the commercially available kits being utilised such as QIAamp, PowerSoil and DNeasy. A comprehensive breakdown of all available metadata for each data set, including specific inclusion and exclusion criteria, age demographics, disease

classification metrics, and DNA extraction protocols is detailed in Appendix 3.1. Although more studies contained public metagenomic sequencing files, it was not possible to include them in the analysis because they did not meet the same strict inclusion criteria, such as plaque origin, Illumina shotgun sequencing, and sufficient metadata, primarily enough data for the identification of health status for each participant. Many other studies, in contrast, employed saliva or other sequencing platforms, while others lacked useful metadata.

Table 3.1 Overview of Publicly Available Sequencing data collected

NCBI SRA Accession Number	Authors	Year	Country	Sampling	Samples- Disease (n= (Disease type)	Samples-Healthy	Sequencing Platform
PRJNA712952	Carda-Diéguez et al.	2022	Spain	Dental Plaque	18 (Caries)	17	Illumina NextSeq
PRJNA383868	Espinoza et al.	2022	USA	Dental Plaque	57 (Caries)	30	Illumina NextSeq
PRJNA766357	Pang et al.	2021	China	Dental Plaque	20 (Caries)	19	Illumina NovaSeq
PRJNA255922	Shi et al.	2019	USA	Dental Plaque	24 (Periodontitis)	0	Illumina HiSeq
PRJNA528558	Shi et al.	2015	USA	Dental Plaque	6 (Periodontitis)	3	Illumina
PRJNA625082	Shaiber et al.	2020	USA	Dental Plaque	0	35	Illumina NextSeq
PRJNA230363	Wang et al.	2016	China	Dental Plaque	10 (Periodontitis)	10	Illumina HiSeq
PRJNA932553	Koohi-Moghadam et al.	2023	Hong Kong	Dental Plaque	21 (Peritonitis)	7	Illumina NovaSeq

PRJNA508385	Altabtbaei et al.	2021	USA	Dental Plaque	68 (Periodontitis)	25	Illumina HiSeq
PRJEB42701	Plachokova et al.	2021	Netherlands	Dental Plaque	22 (Periodontitis)	0	Illumina HiSeq

3.2.2 The Oral Microbiome Profile

To validate our approach of aggregating publicly available metagenomic data, we first assessed whether the dataset could reproduce established taxonomic signatures of dental caries and periodontitis. After normalising species-level taxonomic profiles to CPM, to control for sequencing depth, we evaluated microbial alpha- and beta-diversity to characterise the community structure in healthy versus diseased individuals.

For individuals with dental caries, the microbial communities exhibited no significant alterations in alpha-diversity when compared to healthy controls, as measured by Observed species, Chao1, Shannon, and Simpson indices (Fig. 3.1A). Correspondingly, the overall community structure did not differ significantly between the two groups, a finding indicated by clear overlap in the PCoA of Bray-Curtis dissimilarities (Fig. 3.1C). This was statistically confirmed by a PERMANOVA, which showed that disease status accounted for a small and non-significant portion of the community variance ($R^2 = 0.008$, $p = 0.221$).

A different pattern emerged for the periodontitis-associated microbiome samples, which displayed increased dysbiosis. While species richness (Observed species) was significantly decreased, the Shannon and Simpson diversity indices were significantly increased in the periodontitis cohort relative to healthy controls ($p<0.001$) (Fig. 3.1B), pointing to a shift of ecological balance and more equal distribution of taxa. This community-wide structural shift was further evidenced by a substantial and significant separation in beta-diversity between the periodontitis and healthy groups (PERMANOVA: $R^2 = 0.078$, $p = 0.001$) (Fig. 3.1D).

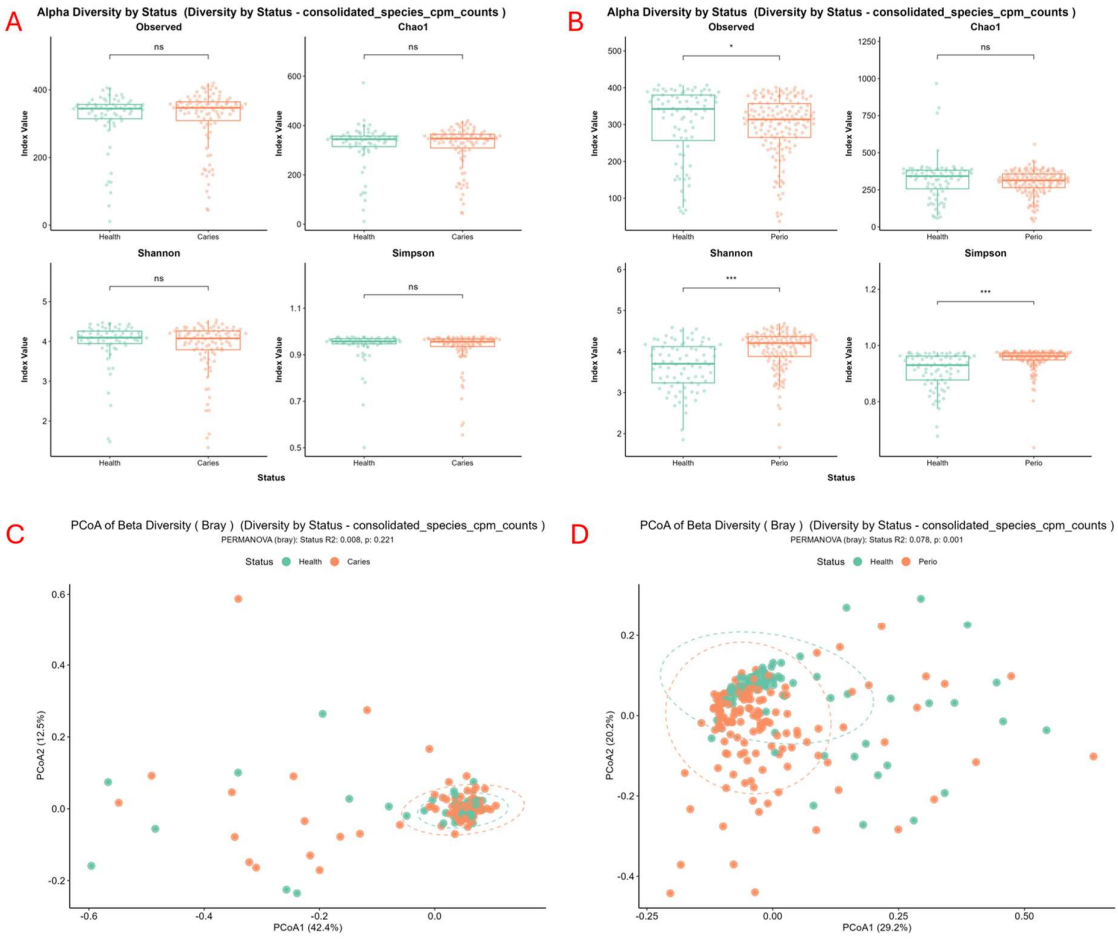


Figure 3.1- Alpha and beta species diversity changes across publicly available datasets comparing caries and periodontitis cohorts to their healthy controls.

(A/C) The caries cohort showed no significant differences in alpha diversity metrics (Observed, Chao1, Shannon, Simpson, all $p > 0.05$) compared to healthy controls (A). Similarly, beta diversity analysis (PCoA) showed no significant clustering by disease status (PERMANOVA $R^2 = 0.008$, $p = 0.221$) (C).

(B/D) In contrast, the periodontitis cohort showed significantly different alpha diversity for three metrics (Observed, Shannon, Simpson, all $p < 0.05$) (B). Beta diversity analysis also showed a significant separation in community structure, although the variance explained by a health status was low (PERMANOVA $R^2 = 0.078$, $p = 0.001$) (D).

3.2.3 Taxonomic Differential Abundance in Caries and Periodontitis Dysbiosis

To identify the specific taxa influencing these community-level changes, we employed a two-pronged analytical approach. An initial analysis using the MetaPhlAn database provided a profile of the bacteriome, which was then complemented by a deeper analysis using a custom-built DIAMOND database to resolve a broader spectrum of oral taxa, including fungi and viruses.

In the comparison of caries versus health, the MetaPhlAn analysis confirmed a significant enrichment of the cariogenic bacterium *Streptococcus mutans* and depletion of health-associated commensals (Fig. 3.2A). Our custom database analysis corroborated the enrichment of an acidogenic bacterial community, identifying significant increases in several species of *Lactobacillus* and *Limosilactobacillus* (Fig. 3.2C). More importantly, this broader analysis showed a significant cross-kingdom component to caries dysbiosis, highlighted by a substantial enrichment of the fungal pathogen *Candida albicans*. The analysis also captured shifts in the oral virome, including an increase in *Lactobacillus* phages in the caries state.

Analysis of the periodontitis cohort uncovered a similarly complex, multi-kingdom dysbiosis. Here, the initial MetaPhlAn analysis identified the enrichment of the keystone "red complex" pathogens, including *Porphyromonas gingivalis* and *Tannerella forsythia* (Fig. 3.2). The custom database analysis substantiated this, confirming a significant increase in periodontal pathogens such as *Treponema* and *Fusobacterium* species (Fig. 3.2B). Deeper analysis using the custom DB also revealed a significant viral signature involving the enrichment of Human betaherpesvirus ($\log2FC = 9.80$, $padj < 0.001$) and Human gammaherpesvirus ($\log2FC = 6.08$, $padj < 0.001$) in the periodontitis-affected oral cavity (Fig. 3.2D). These taxonomic findings imply a polymicrobial cause for both diseases, marked by potential interactions among bacteria, fungi, and viruses.

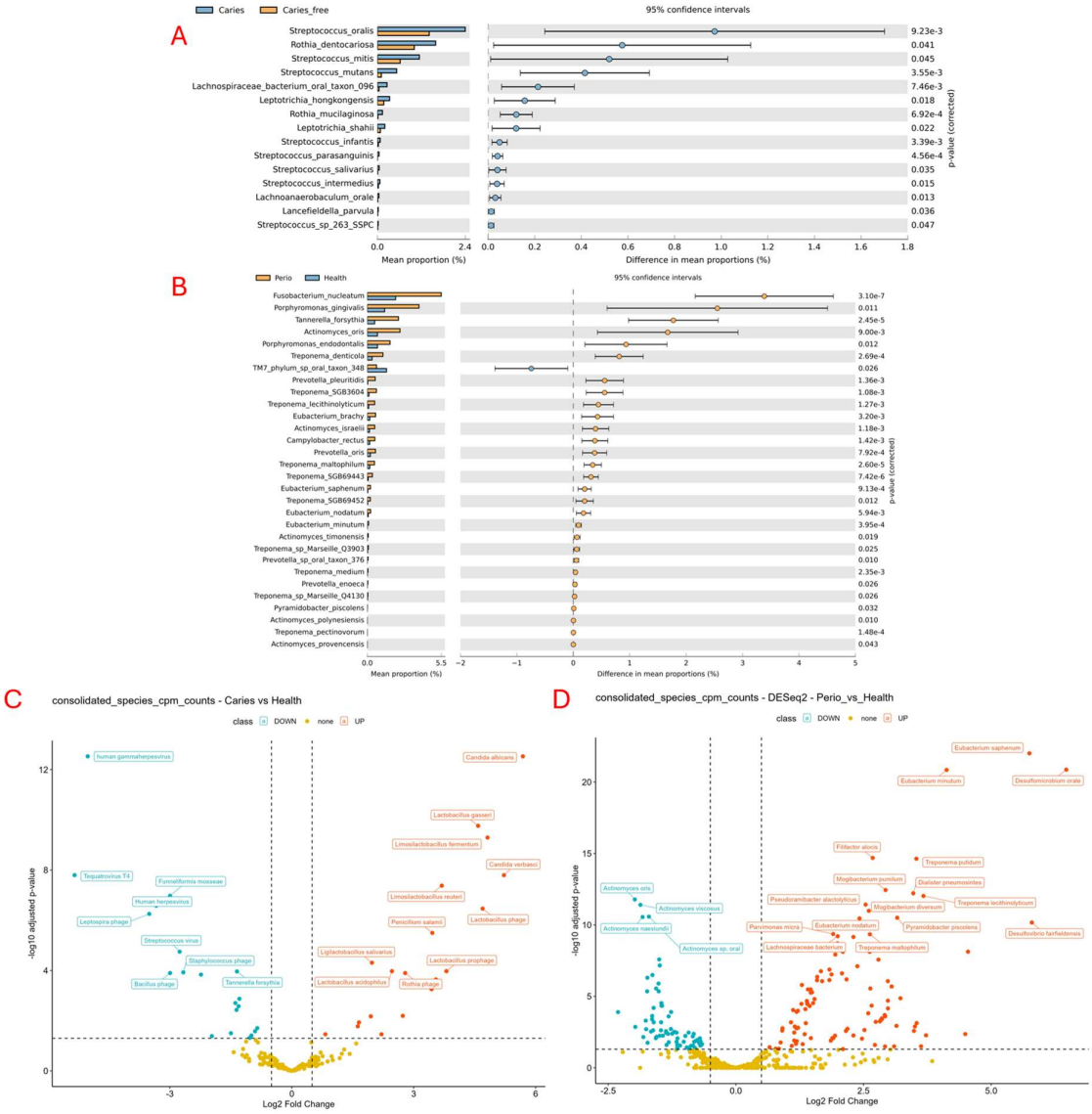


Figure 3.2- Box and Volcano plots displaying significant differential abundance changes measured by DESEQ2, using MetaPhlan and custom DB, respectively, across publicly available datasets comparing caries and periodontitis cohorts to their healthy controls.

(A/B) Results from the MetaPhlAn database analysis, showing mean proportions (left) and the difference in mean proportions with 95 percent confidence intervals (right). Panel (A) compares Caries to Health, and panel (B) compares Periodontitis to Health.

(C/D) Volcano plots from the custom database analysis, showing log2 Fold Change versus -log10 adjusted p-value. Panel (C) compares Caries to Health, and panel (D) compares Periodontitis to Health.

3.2.4 Functional Metagenomic Profiles Exhibit Stability and Redundancy

Finally, to understand the functional consequences of these taxonomic shifts, we analysed the diversity and composition of the annotated gene pool. For the caries comparison, the functional analysis mirrored the taxonomic findings, showing no significant differences in either the alpha-diversity (Fig. 3.3A) or the beta-diversity of gene profiles between healthy and caries-affected individuals (PERMANOVA: $R^2 = 0.007$, $p = 0.232$) (Fig. 3.3C).

More complex patterns emerged from the periodontitis metagenome, unlike the species-level findings, the alpha-diversity of the gene pool was not significantly different between the healthy and periodontitis groups across any metric (Fig. 3.3B). This suggests a high degree of redundancy in the richness of functions within the oral microbiome, where the overall diversity of functional capabilities is maintained even during major taxonomic collapse. While the functional richness was stable, the functional composition was not, as evidenced by beta diversity changes, meaning both groups contain a similar set of genes, but their relative abundances differ in disease. The overall functional profile of the periodontitis microbiome was significantly different from that of the healthy state (PERMANOVA: $R^2 = 0.029$, $p = 0.001$) (Fig. 3.3D). The magnitude of this functional shift ($R^2 = 0.029$) was considerably smaller than the taxonomic shift ($R^2 = 0.078$), further supporting the concept of functional redundancy despite species-level dysbiosis.

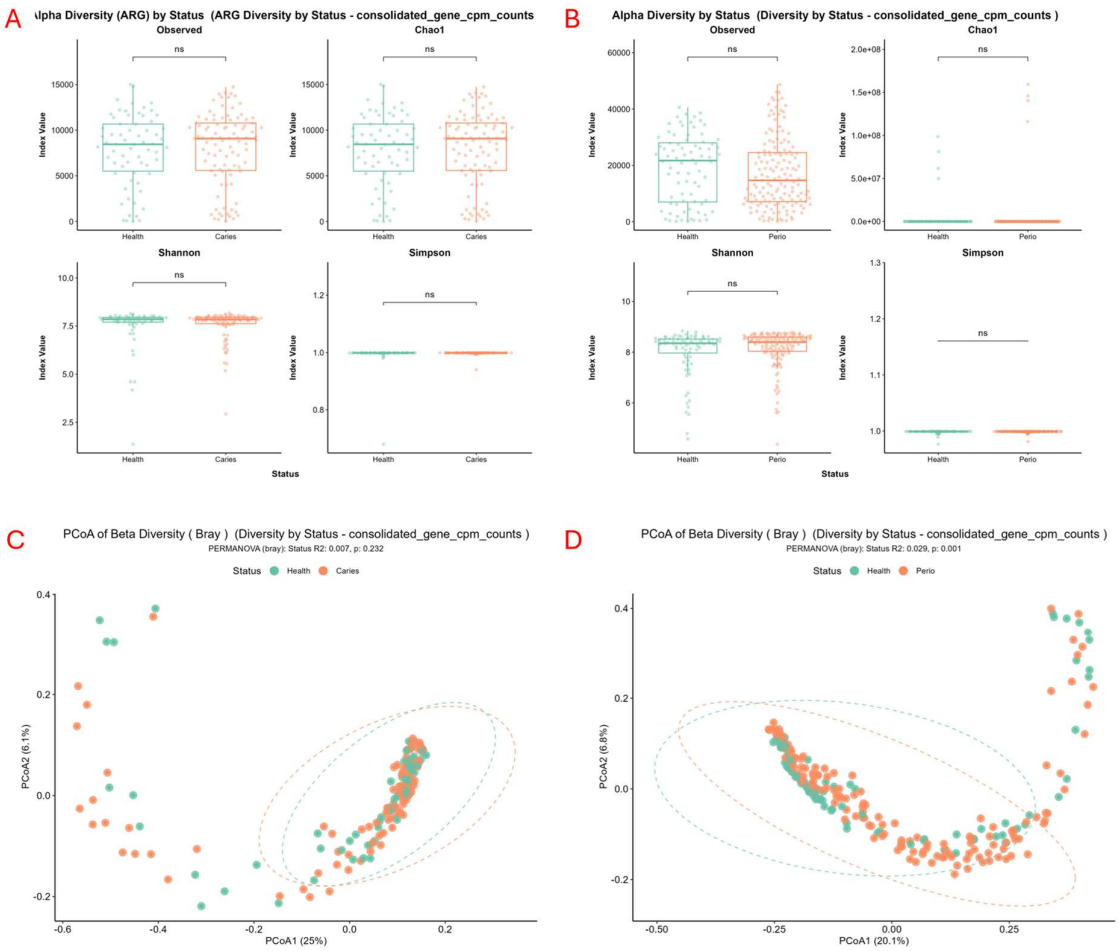


Figure 3.3- Alpha and beta gene diversity changes across publicly available datasets comparing caries and periodontitis cohorts to their healthy controls.

(A/C) The caries cohort showed no significant differences in alpha diversity metrics (Observed, Chao1, Shannon, Simpson, all $p > 0.05$) compared to healthy controls (A). Similarly, beta diversity analysis (PCoA) showed no significant clustering by disease status (PERMANOVA $R^2 = 0.007$, $p = 0.232$) (C).

(B/D) Similarly, the periodontitis cohort showed no significant differences in alpha diversity for all four metrics (Observed, Chao1, Shannon, Simpson, all $p > 0.05$) (B). Beta diversity analysis showed a significant separation in community structure, although the variance explained by a health status was low (PERMANOVA $R^2 = 0.028$, $p = 0.001$) (D).

3.2.5 The Oral Resistome is Significantly Altered in Disease

To specifically investigate AMR potential, we characterised the oral resistome by aligning metagenomic data against the CARD database. This approach revealed distinct resistome profiles in the two disease states.

For dental caries, the resistome was remarkably stable and largely resembled that of healthy samples, with no significant differences in the richness or overall diversity of ARGs (Fig. 3.4A). A minor but statistically significant decrease was observed in the Simpson index ($p < 0.05$), indicating a subtle shift in the evenness of the ARG profile. The overall composition of the resistome, however, did not differ significantly between caries and healthy individuals (PERMANOVA: $R^2 = 0.009$, $p = 0.145$) (Fig. 3.4C).

The resistome associated with periodontitis, on the other hand, underwent a significant change. A key feature of this transformation was a depletion in ARG diversity. Both the richness of ARGs (Observed and Chao1 indices, $p<0.01$) and the overall Shannon diversity ($p<0.001$) were significantly reduced in the periodontitis group compared to healthy controls (Fig. 3.4B). This loss of ARG diversity mirrors the reduction in taxonomic diversity, indicating that the ecological collapse of the commensal community results in the concurrent loss of its associated accessory genome, including a broad range of resistance determinants. Beyond this depletion, the composition of the remaining resistome was also significantly altered. Principal Coordinate Analysis revealed a distinct clustering of the periodontitis resistome away from the healthy controls (Fig. 3.4D), a shift that, although accounting for a modest portion of the total variance, was highly significant (PERMANOVA: $R^2 = 0.02$, $p = 0.001$).

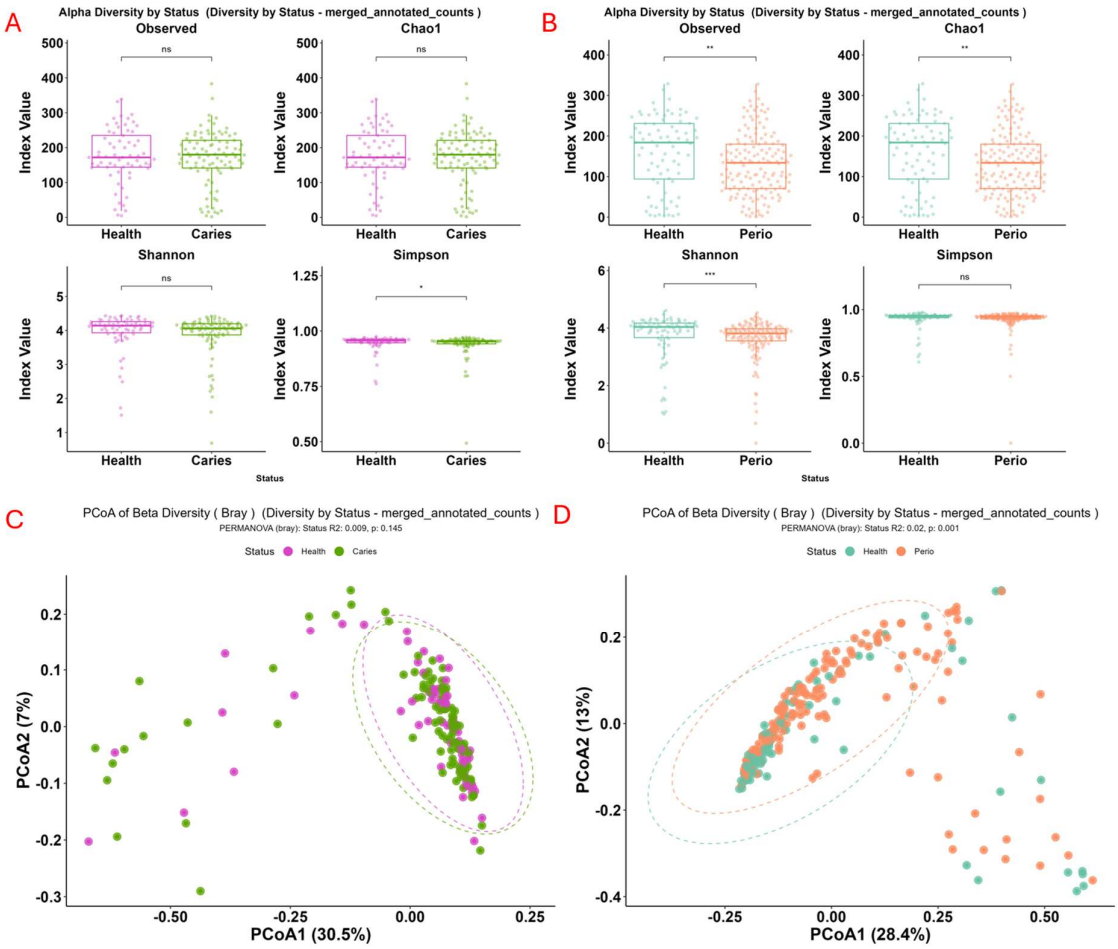


Figure 3.4- Alpha and beta ARG diversity changes across publicly available datasets comparing caries and periodontitis cohorts to their healthy controls.

(A/C) The caries cohort showed no significant differences in alpha diversity metrics (Observed, Chao1, Shannon, Simpson, all $p > 0.05$) compared to healthy controls (A). Similarly, beta diversity analysis (PCoA) showed no significant clustering by disease status (PERMANOVA $R^2 = 0.009$, $p = 0.145$) (C).

(B/D) In contrast, the periodontitis cohort showed significant differences in alpha diversity for three metrics (Observed, Chao1, Shannon, all $p < 0.05$) (B). Beta diversity analysis showed a significant separation in community structure, although the variance explained by a health status was low (PERMANOVA $R^2 = 0.02$, $p = 0.001$) (D).

3.2.6 Differential Abundance of ARGs Reveals Disease-Specific Resistance Mechanisms

To move beyond diversity metrics and identify the specific ARGs causing these shifts, we performed differential abundance analysis on individual ARG profiles in DESEQ2. This produced clear, disease-specific signatures of AMR with direct clinical relevance.

While overall resistome diversity remained stable in caries, a significant restructuring of specific ARG families, many associated with MGEs, became apparent. Notably, there was a significant enrichment of genes conferring resistance to tetracyclines (*tet*(W)) and macrolides (*erm*(B), *erm*(F)), classes of antibiotics often used as second-line treatments in dentistry (Fig. 3.5E). The prevalence of these particular genes is significant, as they are classically disseminated on MGEs like conjugative transposons, suggesting a role for HGT in shaping the caries resistome. Functionally, this translates to an enrichment of resistance mechanisms against peptide and glycopeptide antibiotics, resulting from alterations in antibiotic targets and their replacement (Fig. 3.5A/C). The enrichment of resistance to glycopeptides (e.g. vancomycin), a class of last-resort antibiotics, highlights the potential for the oral microbiome to act as a reservoir for clinically critical ARGs.

In periodontitis, the compositional shift highlighted the selection of specific, high-impact resistance determinants directly relevant to treatment. Of primary clinical interest was the significant enrichment of various beta-lactamase genes, including multiple TEM and VEB variants (Fig. 3.5F). As beta-lactams like amoxicillin represent the first line of antibiotic therapy for severe periodontitis, suggesting a mechanistic basis for potential treatment failure. The enrichment of these TEM variants, which are typically plasmid-mediated, underscores the role of MGEs in allowing the pathogens involved to develop clinically relevant resistance. This selective enrichment was contrasted by a depletion of the MGE-associated *mcr* gene family, which confers resistance to the last-resort antibiotic colistin. The specific loss of these plasmid-borne genes suggests a complex dynamic of fitness costs and selective pressures shaping the resistome during dysbiosis. The overall resistance of the periodontitis microbiome exhibited an overrepresentation of antibiotic inactivation (primarily attributed to beta-lactamases) and antibiotic efflux pathways (Fig. 3.5B/D), two mechanisms frequently mobilised by MGEs.

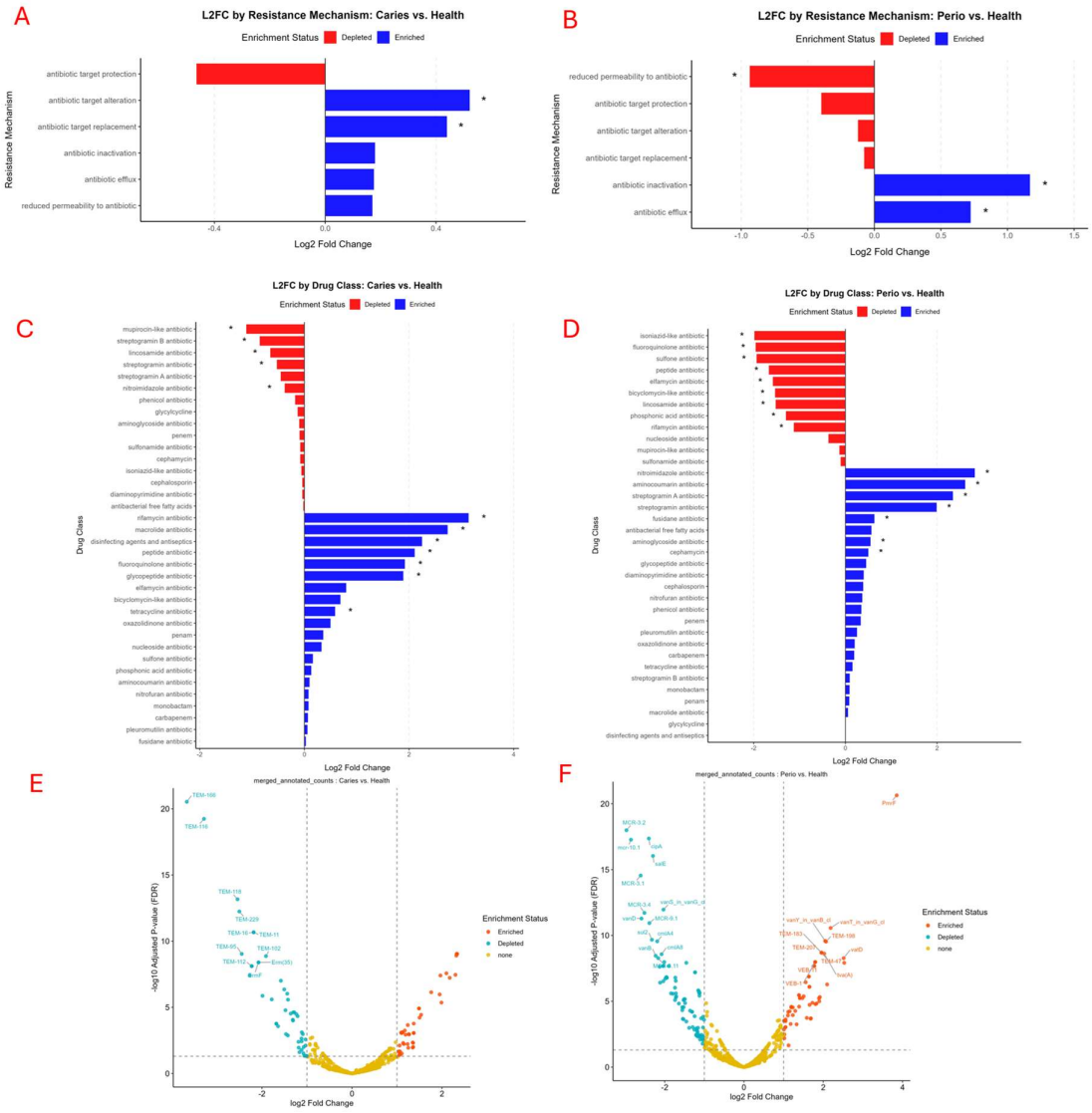


Figure 3.5. Differential abundance of Antibiotic Resistance Genes (ARGs), classes, and mechanisms in caries and periodontitis.

(A/B) Log2 fold change of ARG functional mechanisms in (A) Caries vs. Health and (B) Periodontitis vs. Health. (C/D) Log2 fold change of ARGs grouped by drug class in (C) Caries vs. Health and (D) Periodontitis vs. Health. For A-D, blue indicates enrichment in disease, and red indicates depletion. (E/F) Volcano plots showing individual ARGs in (E) Caries vs. Health and (F) Periodontitis vs. Health. Red points are significantly enriched in disease, and teal points are significantly depleted.

3.2.7 Resistotyping Defines a Core Oral Resistome and a Distinct Periodontitis-Associated Profile

To investigate whether AMR genes organise into discrete profiles across the cohort population, we performed resistotyping using hierarchical clustering on the binary presence or absence of ARGs. The analysis was conducted stepwise, first exploring disease-specific cohorts before performing a combined analysis.

Upon initial analysis, the periodontitis and its corresponding healthy adult cohort partitioned into two optimal resistotypes (RTs), RT-P1 and RT-P2 (Fig. 3.6C/D). Sample distribution was strongly and significantly associated with clinical status ($p<0.001$), with RT-P2 being overwhelmingly composed of periodontitis samples, which demonstrated the existence of a distinct, disease-specific ARG configuration in periodontitis. A similar analysis of the caries cohort, however, found no significant association with disease status ($p>0.05$) (Fig. 3.6A/B). This indicates that, unlike periodontitis, dental caries is not defined by a unique resistotype and is more similar to what is observed in healthy individuals.

Given that the healthy controls differed by age, a validation step was performed to ensure age was not a confounding factor. A direct comparison of resistomes from healthy children and adults confirmed that, although two resistotypes could be identified, their distribution was not significantly associated with age ($p > 0.39$) (Fig. 3.7A/B). This result demonstrated the potential comparability of healthy resistomes across age groups, justifying their merging into a single "Healthy" cohort for definitive comparison, however, other factors, such as diet that were not available in the metadata, may affect the resistotypes.

After validating the cohorts, we performed a comprehensive resistotyping analysis across all three groups, health, caries, and periodontitis. This separated the entire population into two optimal resistotypes, RT1 and RT2, which were divided in the PCoA (Fig. 3.7C/D). The distribution of status groups within these two resistotypes was highly structured and statistically significant ($p<0.001$). RT1 emerged as a 'core' oral resistome, comprising the vast majority of individuals from both the Healthy and Caries cohorts, as well as a smaller fraction of the Periodontitis subjects. RT2, however, was almost exclusively dominated by samples from the Periodontitis

cohort. This confirms and reinforces the initial findings, while the ARG profiles associated with health and dental caries are essentially indistinguishable within a common, core resistotype (despite significantly differentially abundant ARGs), periodontitis is uniquely defined by a fundamental shift to a distinct, disease-associated profile.

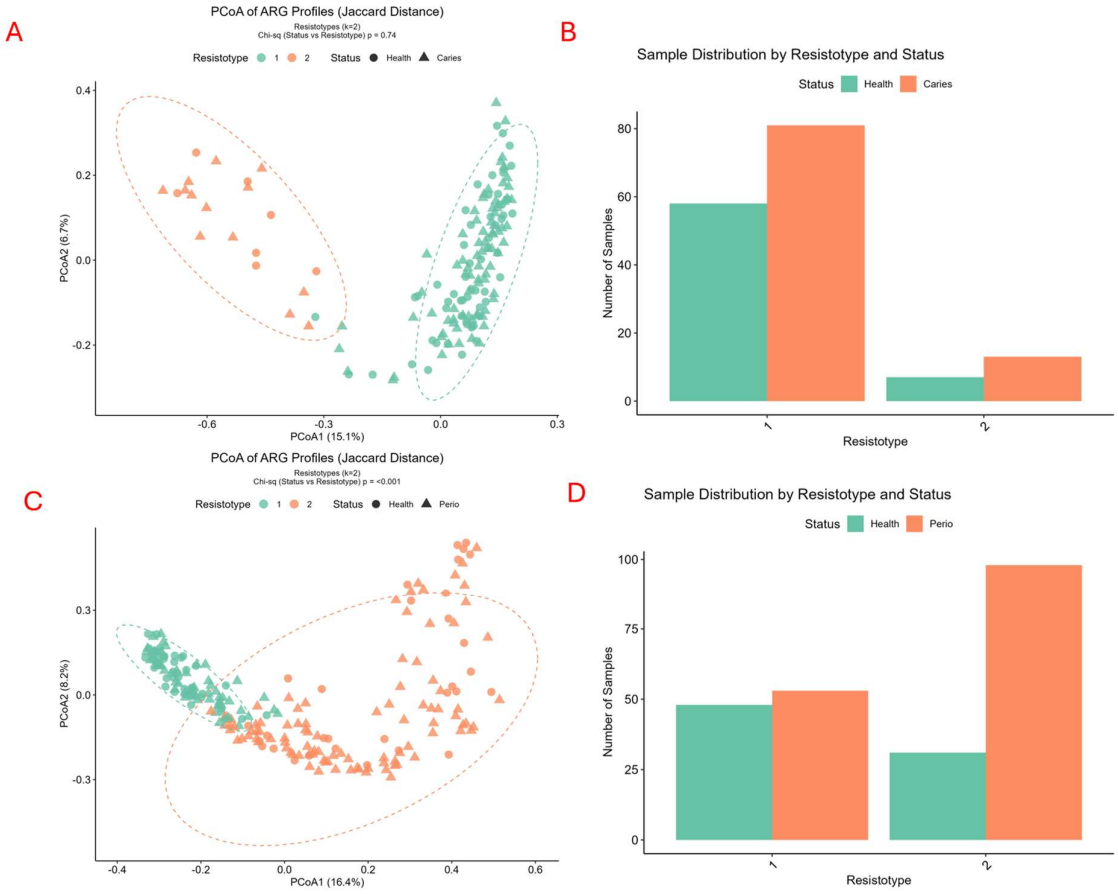


Figure 3.6. Resistotyping analysis of individual disease cohorts compared to their healthy controls.

(A/B) Analysis of the Caries cohort vs. healthy children. (A) PCoA of ARG profiles and (B) distribution of samples by resistotype. There is no significant association between resistotype and disease status (Chi-sq, $p > 0.05$). (C/D) Analysis of the Periodontitis cohort vs. healthy adults. (C) PCoA of ARG profiles and (D) distribution of samples by resistotype. There is a strong, significant association between resistotype and disease status (Chi-sq, $p < 0.001$).

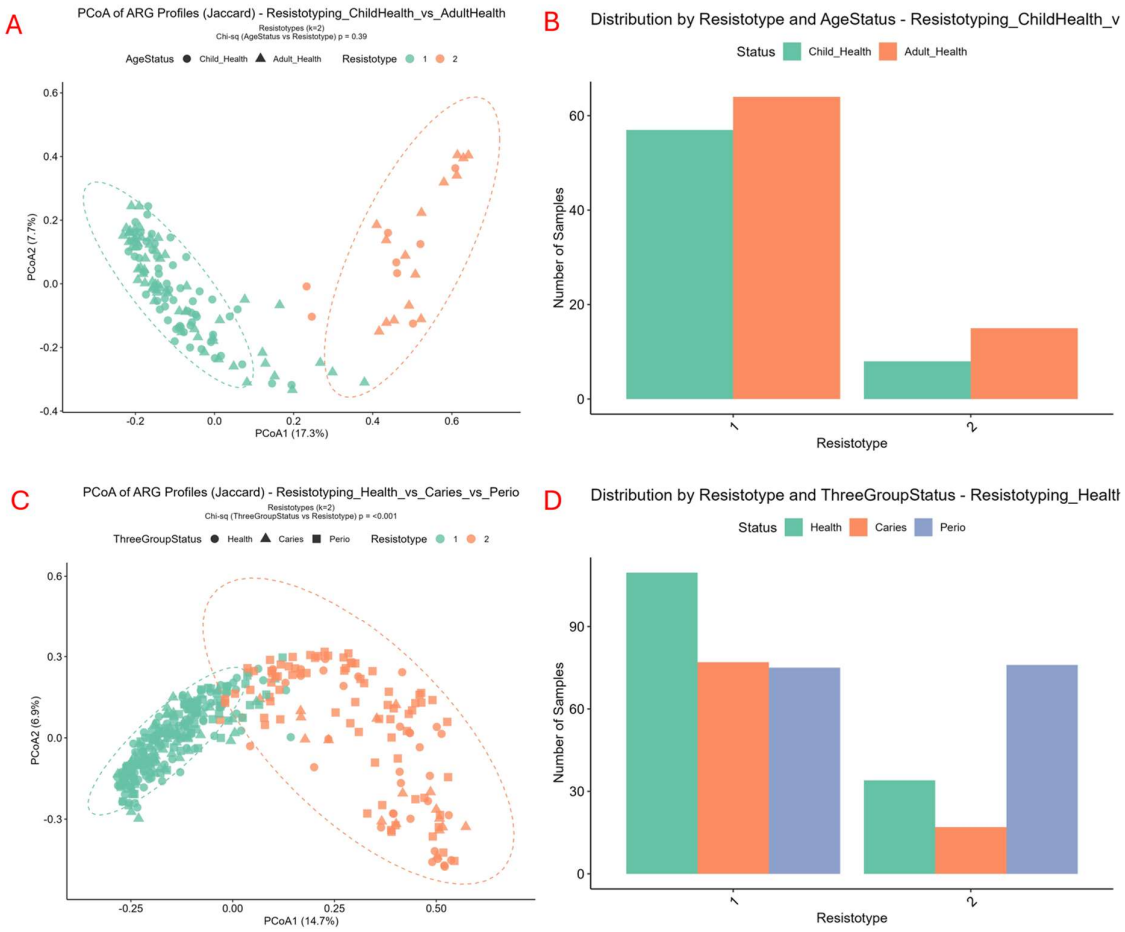


Figure 3.7. Validation of resistotypes across age and combined cohort analysis.

(A/B) Validation analysis comparing healthy children vs. healthy adults. (A) PCoA of ARG profiles and (B) sample distribution show no significant association between resistotype and age (Chi-sq, $p > 0.05$). (C/D) Comprehensive analysis of all three cohorts (Health, Caries, Periodontitis). (C) PCoA of ARG profiles and (D) sample distribution show a highly significant association between resistotype and group status (Chi-sq, $p < 0.001$), with the Periodontitis cohort forming a distinct resistotype.

3.2.8 Mobilome Analysis Uncovers a Disrupted Plasmid Profile in Periodontitis

Given the evidence for elevated HGT in oral biofilms and our findings of MGE-associated ARGs, we concluded our study by characterising the oral mobilome. Using the PLASMe tool, we first focused on the plasmid population (the plasmidome) to assess its diversity and composition in health and disease (X. Tang et al., 2023).

The findings mirrored patterns observed in the taxonomic and resistome analyses. In the dental caries cohort, the plasmidome remained remarkably stable, with no significant differences in the alpha diversity of plasmids between healthy and caries-affected individuals across any metrics evaluated (Fig. 3.8A/C). Furthermore, the overall composition did not differ significantly between the two groups (PERMANOVA: $R^2 = 0.01$, $p = 0.059$).

The periodontitis-associated plasmidome, on the other hand, was significantly altered. We observed a significant depletion in plasmid diversity, marked by lower richness (Observed, $p < 0.05$), Shannon diversity ($p < 0.01$), and Simpson diversity ($p < 0.01$) compared to healthy controls (Fig. 3.8B/D). This loss of diversity was accompanied by a significant shift in the overall plasmid composition (PERMANOVA: $R^2 = 0.028$, $p = 0.001$), indicating that the dysbiotic, pathogen-dominated community in periodontitis harbours a less diverse, yet compositionally distinct, plasmid population.

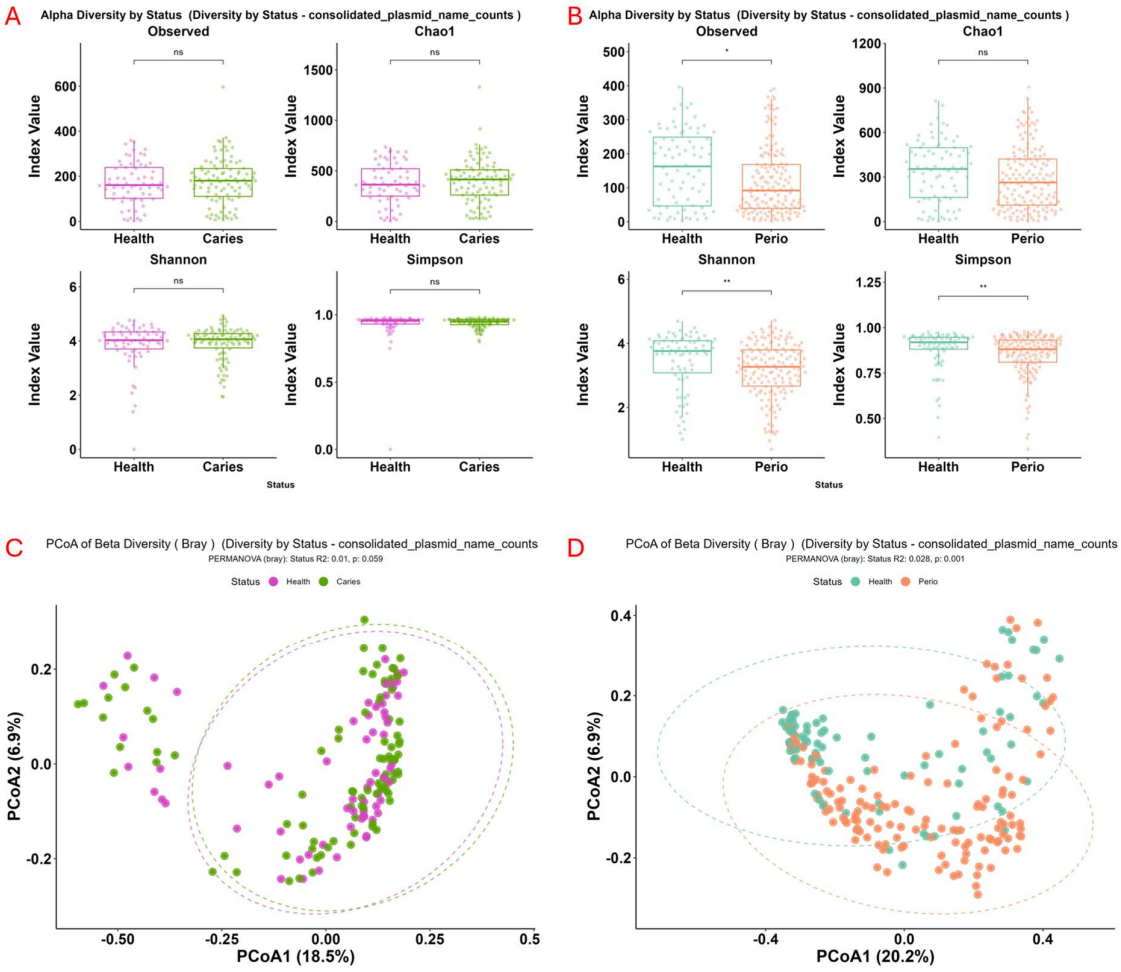


Figure 3.8 - Alpha and beta plasmid diversity changes across publicly available datasets comparing caries and periodontitis cohorts to their healthy controls.

(A/C) The caries cohort showed no significant differences in alpha diversity metrics (Observed, Chao1, Shannon, Simpson, all $p > 0.05$) compared to healthy controls (A). Similarly, beta diversity analysis (PCoA) showed no significant clustering by disease status (PERMANOVA $R^2 = 0.01$, $p = 0.059$) (C).

(B/D) In contrast, the periodontitis cohort showed significant differences in alpha diversity for three metrics (Observed, Shannon, Simpson, all $p < 0.05$) (B). Beta diversity analysis showed a significant separation in community structure, although the variance explained by a health status was low (PERMANOVA $R^2 = 0.028$, $p = 0.001$) (D).

3.2.9 Disease States Show Unique Differential Abundance of Specific Plasmids

To identify the specific plasmids causing these patterns, we performed a differential abundance analysis. While the overall diversity of the caries plasmidome was stable, a dynamic profile of specific plasmids from an extensive range of host organisms was observed (Fig. 3.9A). Plasmids enriched in the caries state originated not only from expected oral acidogenic bacteria like *Limosilactobacillus oris*, but also from classic oral commensals, including *Streptococcus salivarius*. Notably, caries samples carried plasmids from bacteria typically associated with non-human and environmental sources, such as those found in poultry (*Campylobacter jejuni*) and soil (*Bacillus mycoides*, *Paenibacillus polymyxa*), showing that the caries-associated mobilome is a diverse mix, having contributions from both resident oral microbes and the environmentally derived community and highlighting the need for mobilome monitoring from a One Health perspective.

The compositional shift in the periodontitis plasmidome, however, showed a clear enrichment of plasmids from pathogenic host species. Plasmids harboured by the periodontal pathogen *Treponema succinifaciens* and the pathogen *Clostridioides difficile* were significantly more abundant in the periodontitis state (Fig. 3.9B). Additionally, plasmids from bacteria associated with health, such as *Streptococcus salivarius* and *Kocuria flava*, were significantly depleted.

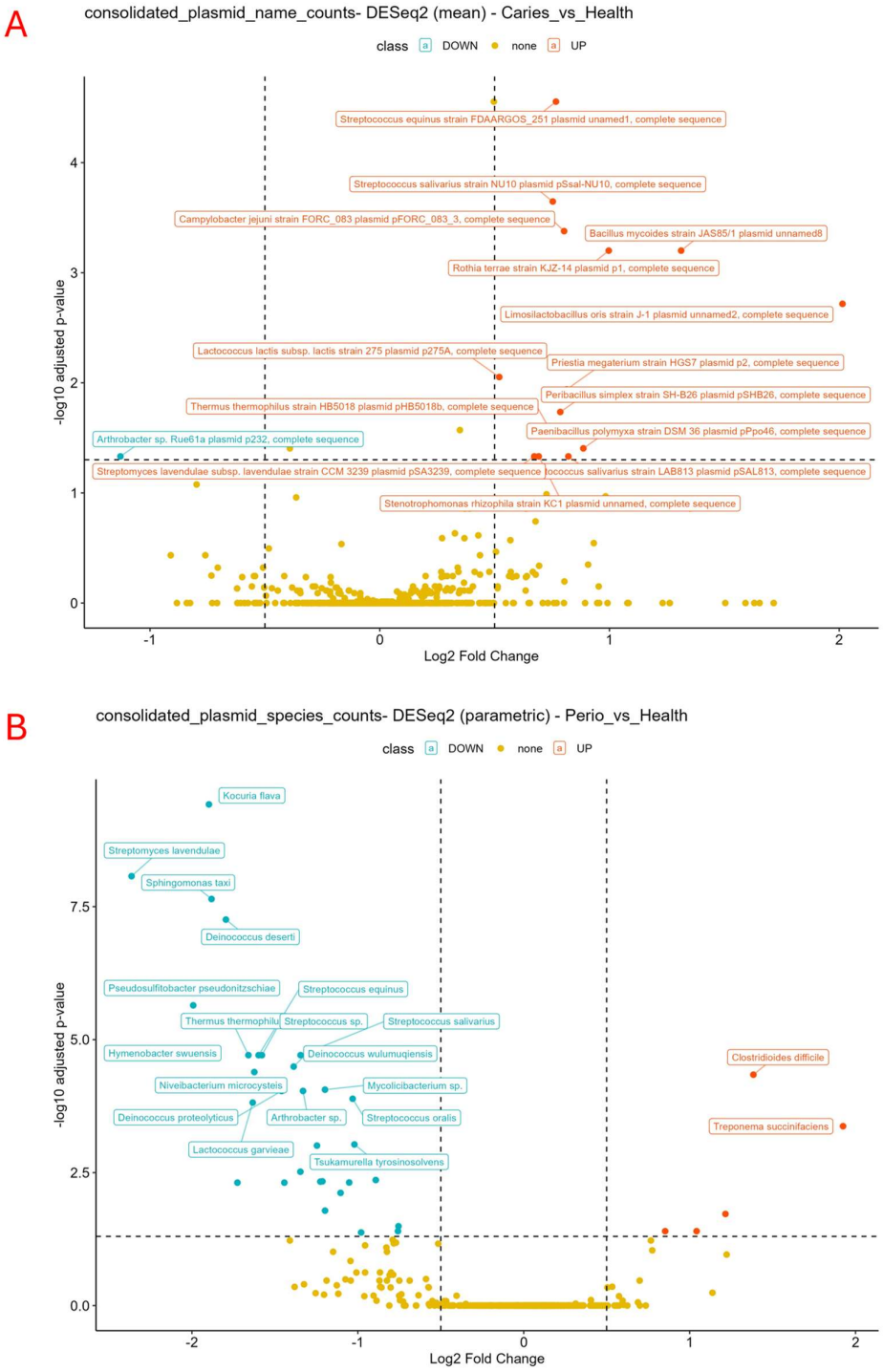


Figure 3.9. Differential abundance of plasmids in caries and periodontitis cohorts. Volcano plots show Log_2 Fold Change versus -log_{10} adjusted p-value for individual plasmids. Red points (UP) represent plasmids significantly enriched in the disease state, while blue points (DOWN) represent plasmids significantly depleted. **(A)** Caries vs. Health cohort. **(B)** Periodontitis vs. Health cohort.

3.2.10 Plasmid-Encoded Proteins Indicate Functional Specialisation in Disease

To understand the functional implications of these plasmidome shifts, we subsequently analysed the differentially abundant proteins encoded on the plasmids. In dental caries, the enriched plasmid-borne proteins were directly related to caries development, including proteins linked to adhesion and biofilm formation (Cna B-type domain-containing protein, GbpC/Spa domain-containing protein) and sugar transport (PTS sugar transporter). Furthermore, we also observed a significant enrichment of proteins essential for HGT itself, such as transposases and conjugative transfer proteins, implying that the caries plasmidome is actively involved in its own dissemination (Fig. 3.10A).

The periodontitis-associated plasmids, however, revealed a different functional signature. Here, the most significantly enriched proteins were multiple toxin systems, used in antagonistic competition between bacteria. In contrast, proteins depleted in periodontitis were primarily involved in core metabolic and biosynthetic functions, such as amino acid synthesis, iron transport, and phosphate regulation (Fig. 3.10B). This demonstrates a functional shift of the plasmidome, from contributing to metabolic stability in health to one specialised for virulence and inter-species warfare in the dysbiotic periodontal environment.

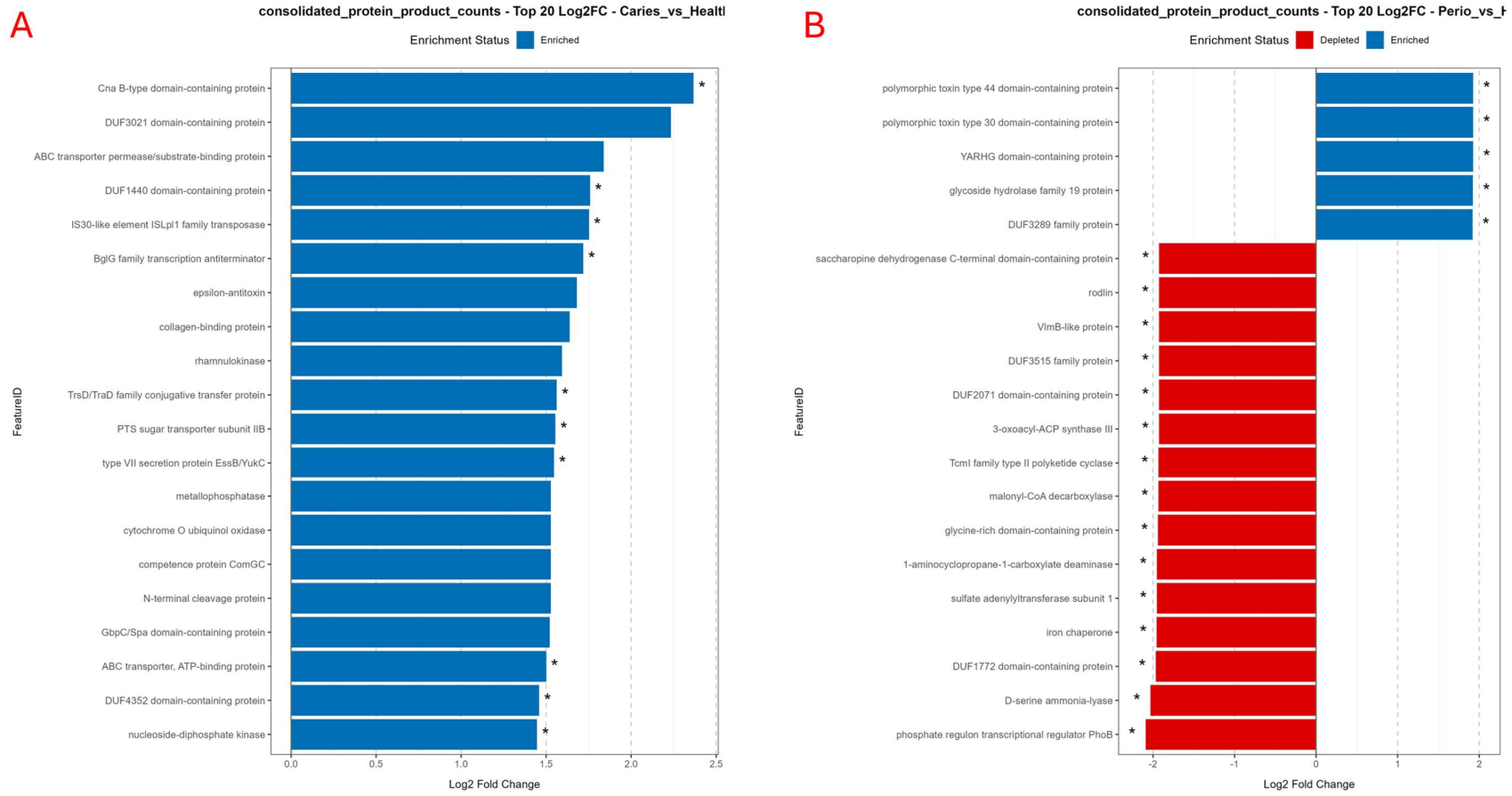


Figure 3.10. Functional analysis of differentially abundant plasmid-borne proteins. Bar charts showing the Log_2 Fold Change of the top 20 differentially abundant proteins encoded on plasmids. **(A)** Top 20 proteins enriched in the Caries vs. Health cohort. **(B)** Top 20 differentially abundant proteins in the Periodontitis vs. Health cohort. Blue bars indicate enrichment in disease, while red bars indicate depletion.

3.2.11 Disease-Specific Changes of Oral HGT Networks

Finally, to directly investigate the dynamics of HGT shaping the oral mobilome, we used the WAAFLE tool to identify active gene donors and recipients. Our analysis found that both oral diseases involve alterations to their HGT networks, but in distinct ways.

In periodontitis, the HGT network was reorganised, with clearly defined shifts in both gene donor and recipient populations. While the healthy state network was dominated by commensal oral bacteria like *Streptococcus parasanguinis* and *Haemophilus parainfluenzae*, the network in periodontitis was skewed primarily within the pathogenic consortium. Established periodontal pathogens, including multiple *Treponema* species and *Porphyromonas gulae*, became the most significantly active gene donors. Other members of the dysbiotic community, such as *Porphyromonas endodontalis* and *Prevotella nigrescens*, became the most active recipients (Fig. 3.11B), indicating a closed, pathogen-focused HGT network.

The HGT network in dental caries, however, showed a more subtle alteration. Separate analyses of gene donors and recipients did not reveal significant, overarching differences between the caries and healthy states. Yet, when all species actively participating in HGT were considered together, a clear shift in the network's active members was apparent. The caries state demonstrated increased HGT participation from several species, most notably clinically significant *Enterococcus faecium* (Fig. 3.11C). Other active participants included lactate-utilising species such as *Veillonella atypica*, various anaerobes, and several *Treponema* species were significant when compared to health, where *Atopobium sp.* was the only taxon showing significantly higher HGT involvement. This suggests that while the donor-recipient structure is not completely reorganised as in periodontitis, the caries environment is marked by a compositional shift in its HGT-active members, including the notable recruitment of opportunistic pathogens.

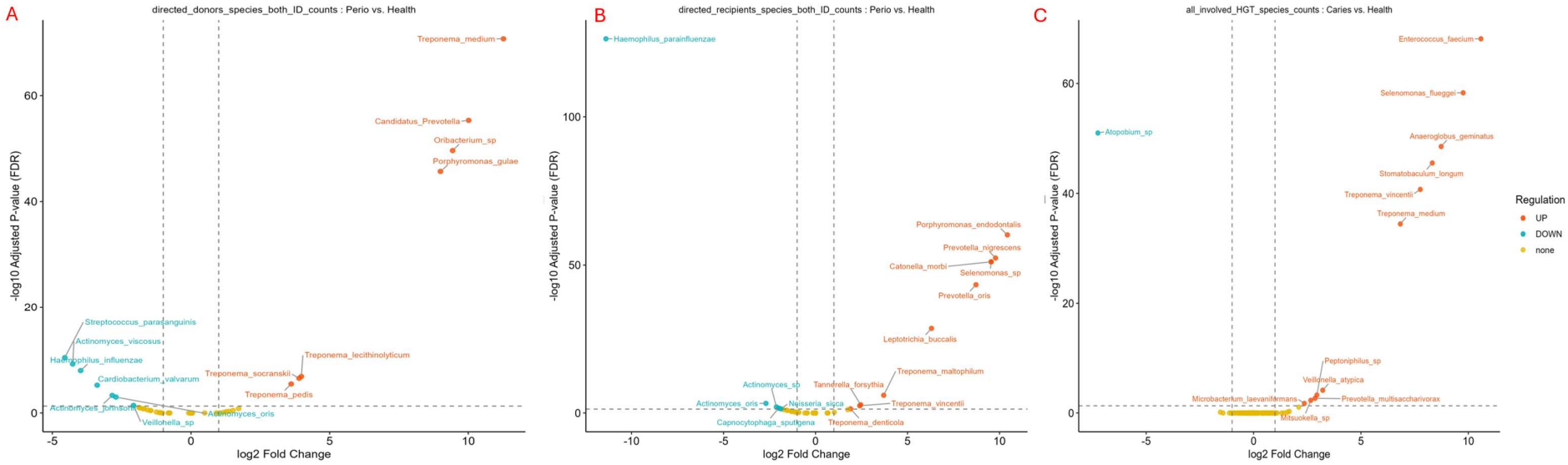


Figure 3.11. Differentially active species in Horizontal Gene Transfer (HGT) networks. Volcano plots show Log_2 Fold Change versus -log_{10} adjusted p-value for species participating in HGT. Red points (UP) represent species with significantly increased HGT activity in disease, while teal points (DOWN) represent species with decreased activity. **(A)** Active gene donors in the Periodontitis vs. Health cohort. **(B)** Active gene recipients in the Periodontitis vs. Health cohort. **(C)** All HGT-involved species in the Caries vs. Health cohort.

3.3 Discussion

This chapter aimed to perform a large-scale analysis of publicly available metagenomic data to explore the complex interplay between the microbiome, resistome, and mobilome of the human oral cavity in states of health and dysbiosis. This approach enables the synthesis of datasets from diverse populations, providing a broad perspective on the principles affecting the oral microbiome. The primary aims were to validate this dataset against established taxonomic signatures of oral disease, identify the specific signatures of dysbiosis in dental caries and periodontitis, and, most importantly, to characterise the associated AMR and gene mobility potential, relating these findings within the broader context of clinical practice and the One Health approach.

The findings of this chapter provide both agreeing and novel findings when compared to the established understanding of the human oral cavity (Deo and Deshmukh, 2019; Tan et al., 2023). In a state of health, or eubiosis, this community of bacteria, fungi, and viruses exists in a symbiotic relationship with the host. Rather than a passive collection of microbes, the HOM acts as an active, functional ecosystem that contributes to homeostasis by forming a protective barrier, preventing pathogen colonisation, and modulating host immune responses (Rajasekaran et al., 2024). Our results demonstrate that the transition from health to disease exhibits disease-specific shifts in this community structure, a state known as dysbiosis. This microbial imbalance is now firmly established as a primary etiological factor in the development of major oral pathologies and is increasingly implicated as a contributor to a broad spectrum of systemic diseases, including cardiovascular disease, rheumatoid arthritis, and inflammatory bowel disease (IBD), highlighting the connection between the oral cavity and overall systemic health (Jia et al., 2018; Tan et al., 2023). The primary mechanisms for this systemic link are bacteremia, where oral pathogens and their products enter the bloodstream, and the oral-gut axis (Jia *et al.*, 2018, Kunath *et al.*, 2024).

A key finding of this study was the confirmation of distinct signatures for dental caries and periodontitis, which both align with and challenge current understanding. Our results indicated that microbial communities in dental caries showed no significant differences in alpha-diversity metrics compared with healthy controls. Our observation contrasts with several studies that report

a trend towards decreased microbial diversity. For example, research by Bhaumik *et al.*, (2024) concluded that microbial diversity was lower in plaque from children with active caries lesions compared to their caries-free controls, while a study of Indian children with and without dental caries also found that the caries microbiome featured a lower alpha diversity (Kalpana et al., 2020). The literature, however, is divided, with other research highlighting significant increases in alpha diversity. For instance, a study performed a similar meta-analysis to this one, using 16S rRNA sequencing and reporting an increase in diversity (Al-Hebshi *et al.*, 2018, Butcher *et al.*, 2022). The finding of stable diversity in our broad meta-analysis does not contradict either of these findings but instead supports the hypothesis that caries is influenced by species-specific abundances rather than by overall diversity changes (Zhang et al., 2022). In the case of caries, frequent sugar intake creates an acidic environment that selects for a specialised, acid-tolerant consortium. Our data reflects this, while the overall number of species remains stable, there is a significant compositional shift, a form of dysbiosis, where particular species become dominant, similar to the ecological plaque hypothesis, in which there is a marked enrichment of disease-causing species (Rosier et al., 2014). This was evidenced by the significant enrichment of the known cariogenic pathogen *Streptococcus mutans* and, as revealed by deeper analysis with the custom DB, the fungus *Candida albicans* and various *Lactobacillus* species (Caufield *et al.*, 2015, Du *et al.*, 2021, Dini *et al.*, 2023). This suggests that changes in species dominance drive the disease. The highly significant enrichment of *C. albicans* in our results is particularly noteworthy, supporting the growing consensus that caries is a polymicrobial disease with a critical cross-kingdom component, where fungi can synergise with bacteria to enhance biofilm formation and acid production, with evidence of *C. albicans* promoting increased biofilm growth when co-cultured with *S. mutans* and triggering the expression of virulence factors, potentially exacerbating caries development (Falsetta et al., 2014).

In contrast, the microbiome affected by periodontitis exhibited a more pronounced dysbiosis. Our results showed a significant decrease in species richness (Observed species) but a significant increase in the Shannon and Simpson diversity indices, indicating a more equal distribution of taxa rather than the dominance of a few pathogenic species taxa. This shift in diversity contrasts

with a subset of the literature that associates periodontitis with reduction in microbial diversity. One such study demonstrated that healthy sites exhibit the highest diversity compared to healthy subjects and to those with progressing or stable periodontitis (Ai et al., 2017). Additionally, stable sites had a higher diversity than those with active periodontitis, potentially indicating that as the development of periodontitis stops, the microbial community reverts to a healthier state and thus diversity increases (Ai et al., 2017). However, our findings more closely align with other studies that have demonstrated the opposite, showing that periodontitis is a complex disease associated with increased microbial diversity. This is often attributed to sampling from deeper periodontal pockets, which represent a distinct, more anaerobic niche capable of supporting an array of pathogens. For instance, studies comparing deep (≥ 5 mm) and shallow (≤ 4 mm) pockets have found significantly higher species richness in the deeper sites (Moradi et al., 2025). The fact that our meta-analysis, which aggregated data from various sampling strategies, found a significant increase in diversity suggests that this may be the more robust, overarching signature of the disease at the community-wide level. This community-wide shift was further evidenced by the separation in beta-diversity between the periodontitis and healthy groups (PERMANOVA R^2 = 0.078), a much larger effect than that observed in caries (R^2 = 0.008). Our differential abundance analysis confirmed this shift resulted from the enrichment of keystone "red complex" pathogens, such as *Porphyromonas gingivalis* and *Tannerella forsythia*, as well as other periodontal pathogens, including *Treponema* and *Fusobacterium* species. A novel and important finding from our analysis was the significant enrichment of human beta- and gammaherpesviruses in the periodontitis cohort. This supports an emerging hypothesis where eukaryotic viruses may play a crucial role in the pathogenesis of periodontitis, potentially by impairing local host immune functions and creating a more habitable environment for bacterial pathogens to become established (Mahmood et al., 2024).

Despite these results, it is also important to note that as a meta-analysis using data from varying countries, in both caries and periodontitis cohorts, our findings on diversity could be skewed by inherent variations in diet, lifestyle, and host genetics, which could not be controlled for in this study design (Li et al., 2022). For instance, "Westernised" diets are often rich in refined

carbohydrates and are known to select for acidogenic profiles linked to caries, while lifestyle factors like smoking alter the oral environment, favouring anaerobic pathogens associated with periodontitis (Silva, 2021, Santonocito *et al.*, 2022). These geographical differences can be a dominant factor, in some cases having a larger effect on microbial composition than disease status, which may lead to discrepancies when contrasting our broad findings with more geographically restricted studies (Renson et al., 2019). This may also imply that, despite contradictory results in diversity metrics, the keystone pathogens that influence diseases, such as periodontitis and caries, are ubiquitous worldwide, and their detection and management are more vital than managing diversity.

The functional consequences of these taxonomic shifts revealed the role of resilience when observing gene profiles. In the caries comparison, the functional gene profiles mirrored the taxonomic findings, with no significant differences in alpha or beta diversity. This suggests that the functional capacity of the oral microbiome is stable during the onset of caries. It is crucial, however, to take this finding with caution as the use of metagenomics reveals the genetic potential of a community, not its real-time activity (Solbiati and Frias-Lopez, 2018). The stability of the gene profile does not necessarily mean that function is stable, as gene expression levels can change. For instance, while the genes for sugar metabolism may be present in both healthy and caries plaque, their expression levels can change. Metatranscriptomic studies have shown that in high-sugar environments, organisms like *S. mutans* significantly upregulate the transcription of genes involved in acid production and biofilm formation, driving the cariogenic process even if the overall gene content of the community remains stable (May et al., 2016). RNA-Seq data have shown that when co-cultured with *C. albicans*, *S. mutans* enhances the expression of genes related to carbohydrate transport and metabolism, leading to increased acid production and a more cariogenic environment, which could explain their apparent co-occurrence in our results (He et al., 2017).

A different pattern emerged from the periodontitis metagenome. Despite the major taxonomic collapse observed, the alpha-diversity of the gene pool did not differ significantly between the healthy and periodontitis groups. This result, again, provides strong evidence for the concept of

functional redundancy within the oral microbiome (Benítez-Páez *et al.*, 2014, Wei *et al.*, 2021). It suggests that even during a major taxonomic shift, the overall diversity of functional capabilities is maintained, likely because different pathogenic species can perform the metabolic roles of the lost commensals (Fässler et al., 2025). While the profile of functions remained stable, their overall composition did not, as evidenced by the significant shift in functional beta diversity. The magnitude of this functional shift (R^2 = 0.029) was considerably smaller than the taxonomic shift (R^2 = 0.078), further supporting the concept of functional resiliency despite species-level dysbiosis.

Beyond the taxonomic profile, a key aim of this study was to characterise the oral resistome. Our results revealed that the resistome in dental caries was remarkably stable, with no significant differences in overall diversity compared to the healthy state. This finding is novel and suggests that the ecological pressures influencing caries do not select for a wholesale change in the community's resistance potential. A more focused analysis, however, revealed a significant and clinically relevant restructuring of specific ARG families. The enrichment of genes conferring resistance to tetracyclines (*tet*(W)) and macrolides (*erm*(B), *erm*(F)) is particularly significant, as these are classes of antibiotics often used in dentistry (Ahmadi et al., 2021). The prevalence of these specific genes, which are typically disseminated on MGEs such as conjugative transposons, indicates a key role for HGT in shaping the caries resistome, even in the absence of a major diversity shift (Spigaglia, Barbanti and Mastrantonio, 2005, Akhtar, Hirt and Zurek, 2009).

The periodontitis resistome, in contrast, showed a significant alteration that mirrored its taxonomic changes. A key finding was a marked depletion in ARG diversity in the periodontitis group. This suggests that as the diverse commensal community is reduced, its associated accessory genome, including a broad range of resistance determinants, is also reduced. This is supported by research showing that the longitudinal development of the childhood resistome diversity mirrors that of the microbiome (Sukumar et al., 2023). This loss of diversity, however, was accompanied by a highly significant compositional shift, driven by the selective enrichment of specific resistance determinants. Of primary clinical importance was the considerable enrichment of various beta-lactamase genes. As beta-lactams, such as amoxicillin, represent the

first-line antibiotic therapy for severe periodontitis, accounting for 82% of all prescriptions in dentistry in Belgium, this finding provides a mechanistic basis for potential treatment failure (Mainjot et al., 2009). It highlights the importance of monitoring due to the widespread use of this antibiotic (Mainjot et al., 2009). It directly challenges the efficacy of this common clinical practice (Ahmadi, Ebrahimi and Ahmadi, 2021, Ali, 2024). This result strongly supports a call for combination therapies or the use of widescale resistome profiling to develop an understanding of the core disease resistome, guiding antibiotic selection in periodontal treatment.

To establish a core resistome and build on these findings, resistotyping analysis represents a novel approach to characterising oral disease states. The partitioning of the entire population into a core oral resistome (RT1), comprising the vast majority of healthy and caries-affected individuals, and a distinct, periodontitis-dominated resistome (RT2), is a significant finding. It demonstrates that, while the ARG profiles in health and caries are essentially indistinguishable at the community level, periodontitis is uniquely characterised by a fundamental shift to a pathogenic ARG configuration. This has implications for diagnostics and could guide future antibiotic policy. The identification of a core resistome associated with disease could inform the development of targeted surveillance programs and evidence-based prescribing guidelines, shifting from empirical treatment to a more precise, data-driven approach to antimicrobial stewardship in dentistry (Sukumar et al., 2024). Understanding the specific ARGs that constitute this core resistome enables more targeted therapeutic strategies and helps predict treatment outcomes, ultimately improving patient care.

The characterisation of the oral microbiome provided further mechanistic insights into disease development. Analysis of differentially abundant plasmid-encoded proteins revealed a functional change. In caries, the enrichment of proteins for adhesion, sugar transport (specifically PTS sugar transporters), and, critically, the machinery for HGT itself (e.g., transposases) shows that the mobilome is actively involved in promoting cariogenic development. The phosphoenolpyruvate:sugar phosphotransferase system (PTS) is the predominant and efficient route for the uptake of dietary sugars by oral streptococci like *S. mutans* (Ajdic and Chen, 2012). The enrichment of PTS components on plasmids suggests a mechanism for enhanced sugar

metabolism, which is directly linked to increased acid production and could be a key factor in determining disease severity and progression (Postma, Lengeler and Jacobson, 1993, Ajdic and Chen, 2012). This plasmid-mediated enhancement of sugar transport provides a significant competitive advantage in the high-sugar environments that influence caries.

For periodontitis-associated plasmids, the findings indicated a shift towards inter-bacterial warfare, with a substantial enrichment of toxin systems. This suggests that in the dysbiotic periodontal environment, the mobilome's primary role shifts from maintaining metabolic stability to facilitating virulence and conflict. Finally, our direct investigation of HGT dynamics revealed a change in the gene transfer network in periodontitis. The finding that established periodontal pathogens became the most active gene donors, primarily transferring genes to other members of the dysbiotic community, indicates a closed, pathogen-driven HGT network. This insight helps explain how the dysbiotic state is actively maintained and reinforced within a biofilm, which is known to facilitate higher rates of HGT compared to planktonic states due to the proximity of cells (Abe et al., 2020). Interestingly, this increased HGT can occur even as taxonomic diversity decreases under intense selective pressure. This relationship has been observed in other ecosystems where environmental stress, such as nitrogen addition, can lead to an increase in HGT events and functional gene diversity despite a decline in bacterial taxonomic diversity (Yang et al., 2024). This reduced diversity accompanied by intensified HGT highlights the adaptive strategies used by microbial communities under stress, however, it may also be attributed to more mature biofilms seen in oral disease leading to an increased rate of HGT as previously described (Kurtzman *et al.*, 2022, Michaelis and Grohmann, 2023). The activity of Enterococcus faecium in the caries HGT network is of considerable importance, as it demonstrates how a common oral disease can create an environment that selects for increased HGT activity by a medically important pathogen, highlighting the significant risk posed by the oral ARG/MGE reservoir. The constant systemic dissemination of oral bacteria via transient bacteremia, a common consequence of daily activities such as chewing, means that these mobilised resistance genes can be transferred to pathogens at distant body sites, jeopardising the treatment of extra-oral infections (Kunath *et al.*, 2024, Sukumar *et al.*, 2024). This positions the oral cavity not only as a source of local

infection but also as a critical hub for the stockpiling and dissemination of AMR, posing a broader public health threat (Moussa et al., 2025).

While this study provides significant new insights, it is essential to acknowledge the inherent limitations of a meta-analysis of publicly available, sequence-based data and to outline directions for future research. A primary limitation is the heterogeneity across available studies, including variations in study design, sample collection, sequencing methods (e.g. different 16S rRNA variable regions), and bioinformatic pipelines, all of which can introduce noise and affect comparability (Sukumar et al., 2024). To mitigate this, the selected studies met specified inclusion criteria that allowed for similar sequencing methods and the use of plaque rather than the salivary microbiome. Furthermore, the analysis is constrained by incomplete metadata variables such as detailed dietary habits, lifestyle factors like smoking, host genetics, and precise metrics of disease severity were not uniformly available across all datasets, limiting our ability to control for their influence fully (Kim et al., 2025). Additionally, the simple identification of disease classification in samples remains limited, with journals often requiring the upload of sequencing data but not corresponding metadata, which significantly reduces the reproducibility and further use of the data, as highlighted by Kim *et al.*, (2025). That study highlighted that almost 50% of the 3,000 studies examined fell short of basic requirements for sequence data accessibility and were accompanied by inadequate metadata. Future work should aim for larger cohorts of both diseases (this study comprising 66 caries-free, 95 caries, 151 periodontitis, and 80 healthy samples) with more homogenous metadata gathered, though this is limited by the availability of samples and corresponding data.

Additionally, the reliance on metagenomics reveals the genetic potential of a community but cannot confirm real-time gene expression or functional activity, a critical distinction when studying dynamic disease processes (Aguiar-Pulido et al., 2016). Finally, our understanding is constrained by the completeness of reference databases for ARGs and MGEs, and the inherent bioinformatic challenges of accurately assembling these complex mobile elements from short-read sequencing data can lead to an underestimation of the mobilome's accurate scale and diversity (Carr *et al.*, 2021, Papp and Solymosi, 2022). Future research should aim to address

these limitations. There is a pressing need to standardise methodologies for data collection, sequencing, and bioinformatic processing to enhance the reliability and comparability of future meta-analyses. Additionally, longitudinal studies are required to move beyond cross-sectional snapshots, allowing us to capture the dynamics of ecological shifts, HGT events, and disease progression, which is essential for inferring causality. The integration of multi-omics approaches, combining metatranscriptomics, metaproteomics, and metabolomics, is crucial for linking genetic potential with functional activity and metabolic output, thereby providing a more comprehensive understanding of disease pathogenesis. These sequencing-based approaches should be complemented by advanced *in vitro* models to isolate and functionally characterise novel taxa, as well as to integrate microbial biofilms with host cells to unravel the complex host-microbe interactions that mediate the transition from health to disease.

In conclusion, this study presents a comprehensive analysis of the microbiome, resistome, and mobilome in oral health and disease. By drawing specific examples, this chapter has shown how the ecological and genetic dynamics of dental caries and periodontitis are distinct from health. The novel findings regarding resistotyping, the functional specialisation of the mobilome, and the reorganisation of HGT networks provide new avenues for a deeper understanding of oral disease pathogenesis. These results demonstrate the critical importance of the oral cavity as a reservoir for AMR and a hub for gene transfer, highlighting the need for stringent antimicrobial stewardship and a One Health approach to managing oral and systemic health.

Chapter 4: Predictive Modelling of Oral Dysbiosis using Machine Learning

4.1 Introduction

The use of metagenomic sequencing in microbiome research has provided new and complex insights into microbial communities, such as the HOM. As demonstrated in Chapter 3, the transition from health to the dysbiotic states of dental caries and periodontitis involves shifts not only in taxonomic composition but also within the community's collective resistome and mobilome. The resulting datasets are inherently high-dimensional, often comprising thousands of microbial, genetic, or functional features from a comparatively small number of samples. While this data contains the signatures of disease, evidenced by changes in diversity and differential abundance, they can often be subtle, with unseen correlations and patterns that may emerge with the use of more advanced computational techniques (Topçuoğlu et al., 2020).

Machine learning may offer a suitable framework for this challenge, providing a range of algorithms capable of building predictive classification models from complex datasets. The use of these methods in oral microbiome research could mean the potential for clinical translation, particularly in the development of non-invasive diagnostic tools based on microbial profiles (Kim et al., 2020). The success of any given model, however, is not guaranteed and depends heavily on the chosen algorithm and the specific data preprocessing steps implemented. The no free lunch theorem in machine learning suggests that no single classifier is universally optimal for all problems, highlighting the need for a systematic evaluation of multiple models and data manipulation strategies (Wolpert and Macready, 1997). This requires the use of a thorough methodology, utilising principles such as cross-validation, to ensure that model performance is reliably measured and that the classifiers can perform well on unseen data, a particular challenge in microbiome studies, which often have limited sample sizes (Vabalas et al., 2019).

While machine learning has been successfully used to build predictive models for specific oral diseases, a systematic comparison of the predictability of multiple different dysbiotic states remains unstudied (Kim *et al.*, 2020, Wu *et al.*, 2021). It is not yet clear whether the taxonomic

shift towards a distinct community profile and resistotype seen in periodontitis offers a more predictable signature than the more subtle compositional shift that characterises dental caries. Furthermore, the question of whether the strongest predictive signal belongs to the taxonomic profile or in the accessory gene pools (the resistome and mobilome) is a key question that could help steer future biomarker discovery. Answering this requires a standardised analytical framework that can be applied consistently across different diseases and data profiles, enabling a fair and direct comparison of their predictive potential.

For this reason, this study was designed to use the curated metagenomic cohorts with their corresponding datasets from the previous chapter to investigate these questions. The primary aims of this chapter were, first, to systematically develop and evaluate a wide range of machine learning models for classifying health versus caries and periodontitis. Secondly, to determine the optimal combination of data preprocessing pipelines and classification algorithms for the taxonomic, resistome, and mobilome data. Thirdly, this study aimed to statistically compare the predictive power of these different 'omes in distinguishing between health and disease. Finally, the research was designed to formally assess whether dental caries or periodontitis represents a more computationally predictable dysbiotic state.

4.2 Results

4.2.1Traditional Machine Learning Models

4.2.1.1 Taxonomic Profiles show Disease-Specific Prediction

The initial analysis focused on the species-level taxonomic profiles to assess their ability to distinguish between health and disease. For the dental caries cohort, the models achieved moderate predictive performance (Fig 4.1A). The top-performing combinations, which included Gradient Boosting and Random Forest classifiers paired with preprocessing scenarios involving SMOTE and feature filtering, yielded mean cross-validated AUC-ROC scores ranging from approximately 0.64 to 0.68. Following an extensive hyperparameter tuning process on these leading models, a statistical comparison using ANOVA and Tukey's HSD test revealed no significant performance differences between the top-performing classifiers (*padj* = 0.7731). This

indicates that while the taxonomic signature of caries contains a consistent predictive signal, it is not particularly strong, and no single model-scenario combination emerged as statistically superior.

A contrasting different outcome was observed for the periodontitis cohort, where the taxonomic data demonstrated a higher degree of prediction (Fig 4.1B). The best models, which included AdaBoost and Logistic Regression classifiers, consistently achieved excellent performance, with mean cross-validated AUC-ROC scores around 0.91. An interesting finding from the hyperparameter tuning phase was that the original, pre-tuned models frequently performed as well as, or better than, their optimised versions, suggesting that the default parameters were already highly effective for this classification. Despite the high performance, the subsequent statistical analysis confirmed that, similar to the caries analysis, there were no significant differences among the top-performing model configurations ($padj = 0.9203$). These results strongly suggest that the taxonomic shift associated with periodontitis provides a clear signature that is correlated with disease status.

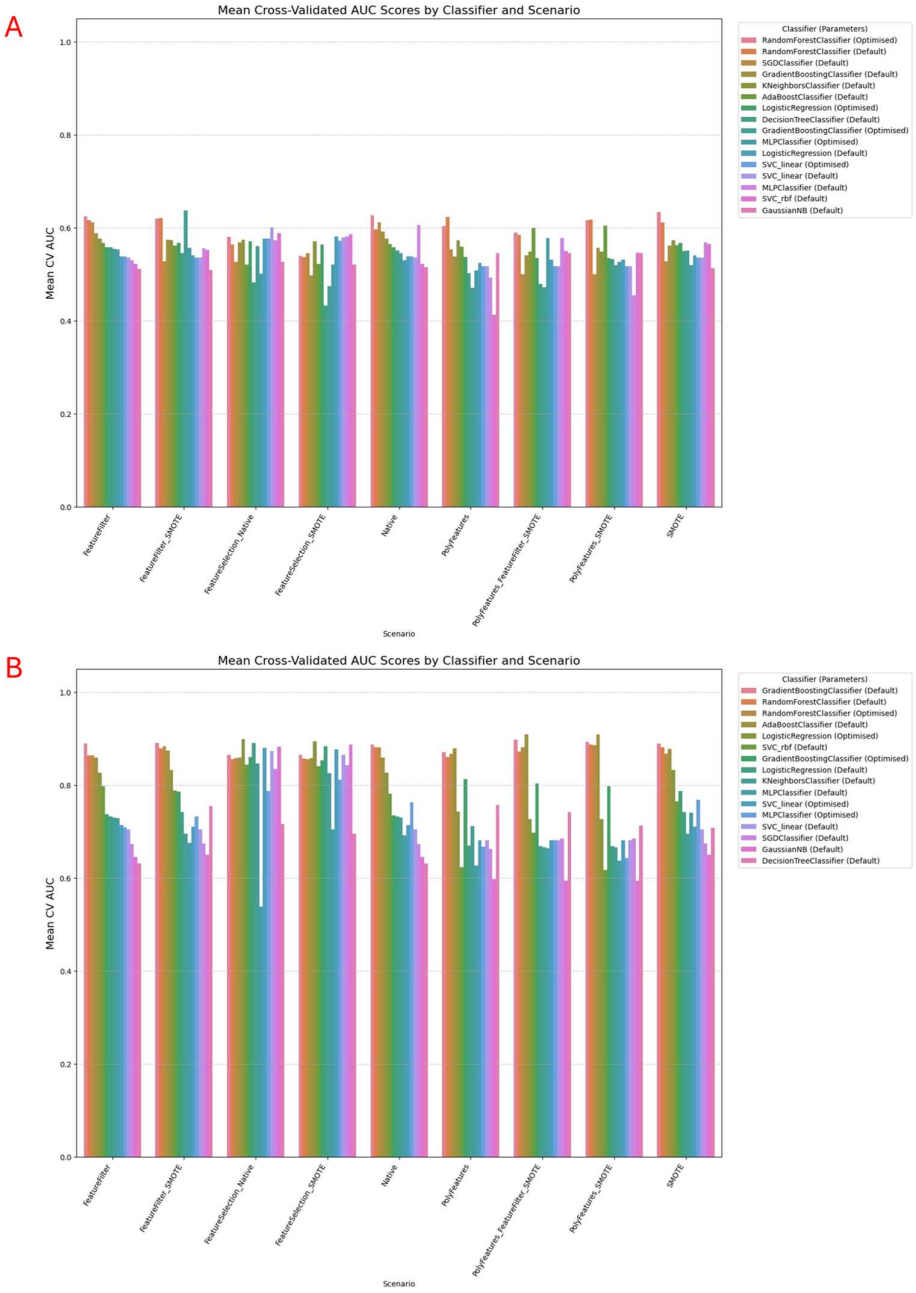


Figure 4.1. Performance of machine learning models for disease classification based on species-level taxonomic profiles. Bar charts show the mean cross-validated Area Under the Curve (AUC-RUC) scores for various classifiers and preprocessing scenarios. **(A)** Caries vs. Health cohort, showing moderate predictive performance with top models achieving AUC scores of ~0.68. **(B)** Periodontitis vs. Health cohort, showing high predictive performance with top models achieving AUC scores of ~0.91.

4.2.1.2 Final Model Assessment and Feature Importance of Top-Performing Models

To further improve the models in the initial assessment, the best-performing model from each analysis was evaluated on the held-out test set, which was not used during training or hyperparameter tuning.

4.2.1.2.1 Caries Taxonomic Model Performance

For the dental caries-species analysis, the optimal model was a hyperparameter-tuned Random Forest Classifier, utilising a feature filtering preprocessing step. On the unseen test data, this model achieved a final AUC-ROC of 0.69, confirming a moderate but clear predictive signal within the caries taxonomic profile (Fig. 4.2A). The model's performance, detailed in the confusion matrix, showed a greater ability to correctly identify caries samples (15 out of 19) than healthy samples (7 out of 14), indicating a stronger signal for the disease than health (Fig. 4.2A).

Analysis of the features driving the caries model's predictions showed a diverse set of taxa (Fig. 4.2). The most influential species was *Hoylesella enoeca*, a member of the *Actinomycetales*, which is increasingly recognised for its presence in the oral cavity. The model also identified several bacteriophages, such as *Streptococcus phage IPP61*, as important predictors, highlighting the potential role of the oral virome in dysbiosis. Notably, the model's predictive features also included well-known oral inhabitants like *Tannerella forsythia* and *Rothia dentocariosa*, demonstrating its ability to capture signals from both established and emerging members of the oral microbial community.

4.2.1.2.2 Periodontitis Taxonomic Model Performance

For the periodontitis analysis, the top model was an AdaBoost Classifier within the pipeline that included the generation of polynomial features, feature filtering, and SMOTE for class balancing. After evaluation, this final model achieved an AUC-ROC of 0.86 on the unseen test data, confirming its strong predictive capability (Fig. 4.2C). The confusion matrix showed a high degree of accuracy, with the model correctly classifying 26 out of 31 periodontitis samples and 11 out of 16 healthy samples (Fig. 4.2C). This demonstrates the model's ability to effectively

generalise to new data, with a clear capacity to distinguish the microbial signature of periodontitis from that of health.

Analysis of the model's feature importances provided critical insight into the biological drivers of predictions (Fig. 4.2D). The most influential features were interaction terms generated by the polynomial features step, such as the co-occurrence of *Actinomyces oris* and *Prevotella fusca*. This highlights that the predictive power of the model lies not just in the presence of individual species, but in the interplay between them. Notably, several well-documented periodontal pathogens were identified as highly predictive features, including *Aggregatibacter actinomycetemcomitans* and the interaction between *Fusobacterium sp.* and *Tannerella forsythia*, a member of the periodontitis red complex. The identification of these known pathogens validates the model's biological relevance and demonstrates its ability to learn clinically important species from the taxonomic data, whilst also providing new insights.

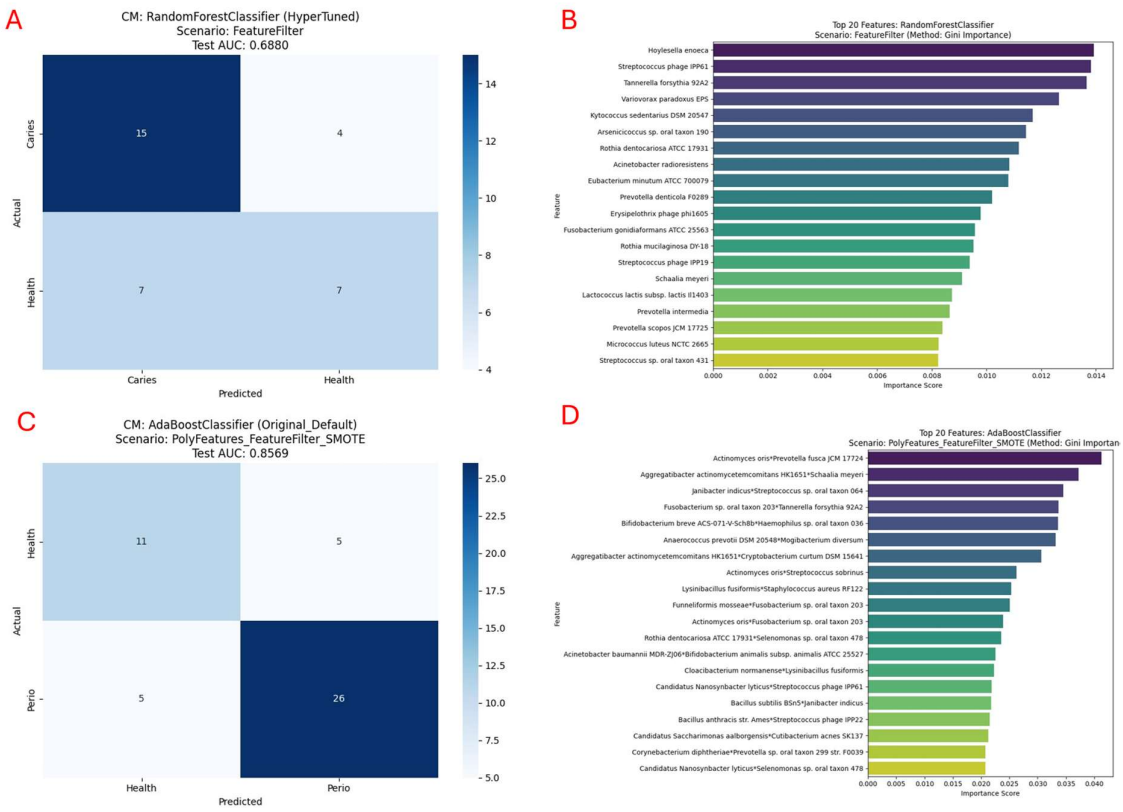


Figure 4.2. Final test set performance and feature importance for the top-performing taxonomic models. (A/B) Results for the Caries vs. Health model (hyperparameter-tuned Random Forest). (A) Confusion matrix showing the model's performance on the unseen test set (AUC-ROC = 0.69). (B) The top 20 most important predictive features (taxa) ranked by Gini importance. **(C/D)** Results for the Periodontitis vs. Health model (AdaBoost Classifier). (C) Confusion matrix showing the model's performance on the unseen test set (AUC-ROC = 0.86). (D) The top 20 most important predictive features (taxa and their interactions) ranked by importance score.

4.2.1.2.3 Performance of a Joint Three-Class Taxonomic Model

To further explore the distinguishing features of the oral microbiome, a joint analysis was performed to train a single model capable of classifying samples into one of three categories: Caries, Health, or Perio. The initial screening produced an optimised Random Forest Classifier with a feature filtering and SMOTE preprocessing pipeline was the top-performing model, achieving a mean cross-validated AUC-ROC of 0.859 (Fig. 4.3A).

Upon final evaluation on the held-out test set, this model confirmed its performance, achieving a multi-class AUC-ROC of 0.88. The confusion matrix revealed a variable performance profile (Fig. 4.3B). The model was highly effective at identifying periodontitis, correctly classifying 24 of 28 samples, reinforcing the distinctness of its microbial signature. Performance on healthy samples was also strong, with 21 of 29 correctly identified. The model's primary challenge was in distinguishing between caries and health, with it correctly identifying only 11 of 19 caries samples, with all of the misclassifications (8) being incorrectly labelled as healthy. This result further confirms the findings from the single disease models, demonstrating that while the microbial signature of periodontitis is clearly distinct from health, the signature of dental caries exhibits a much greater overlap with the healthy state, making it more challenging to classify. Feature importance analysis revealed that the model's predictions were driven by well-established anaerobic and facultative anaerobic bacteria, including *Parvimonas micra*, various *Fusobacterium* species, and *Treponema putidum*, many of which are associated with periodontitis, further highlighting periodontitis' dominant predictive signal (Fig. 4.3C).

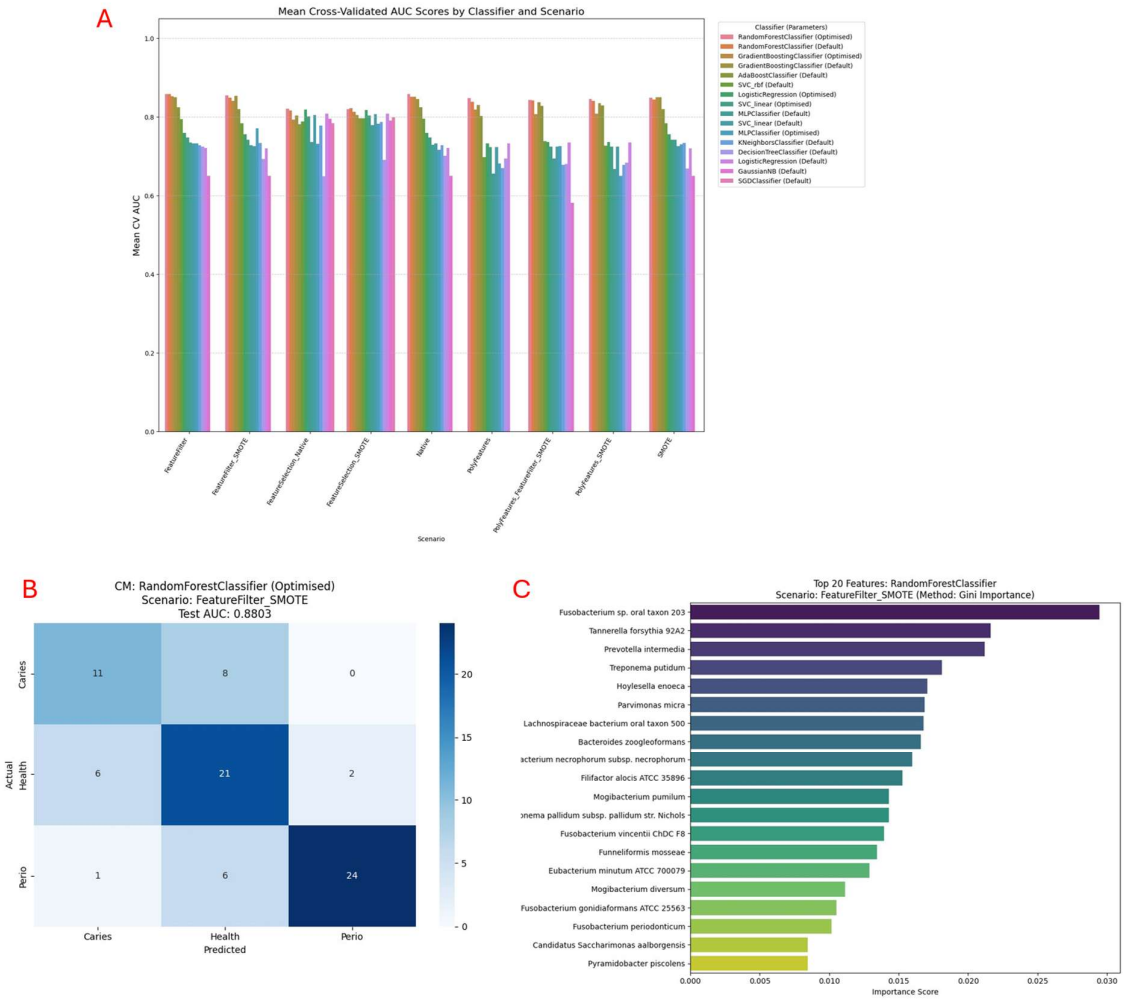


Figure 4.3. Performance of a multi-class model for classifying samples as Caries, Health, or Periodontitis. (A) Initial screening results showing the mean cross-validated AUC scores for various classifiers. The optimised Random Forest model was the top performer. **(B)** Confusion matrix for the final Random Forest model on the held-out test set, achieving a multi-class AUC-ROC of 0.88. **(C)** The top 20 most important predictive features (taxa) for the multi-class model, ranked by Gini importance.

4.2.1.3 Resistome Profiles as Predictive Features

Following the taxonomic analysis, the predictive potential of the oral resistome was investigated using the ARG profiles as features.

4.2.1.3.1 Caries-Specific Resistome Model Performance

For the dental caries cohort's resistome profiles, the models demonstrated limited predictive use, mirroring the taxonomic models. During the initial cross-validation screening, the top-performing models, including Gradient Boosting and AdaBoost classifiers, achieved only modest mean AUC-ROC scores, peaking at approximately 0.65 (Fig 4.4A). The subsequent statistical analysis of the top five models confirmed that no single algorithm or preprocessing pipeline was statistically superior to the others. The best-performing model from this phase, the default Gradient Boosting Classifier with polynomial features, was selected for final evaluation.

Upon assessment with the held-out test data, the model's performance failed to generalise, achieving a final Test AUC-ROC of 0.49 (Fig. 4.5A). This score, being less than the 0.50 baseline for random chance, indicates that the subtle predictive patterns identified during training were not clear and did not translate to unseen data, potentially suggesting overfitting to the training data. The confusion matrix further illustrated this poor performance, showing that the model struggled to correctly classify either group, misclassifying a significant portion of both caries and healthy samples (Fig. 4.5A). Although the model was unable to generalise, analysis of its feature importances revealed that it attempted to use interaction terms between various tetracycline resistance genes, such as *tet(44)* and *tet(T)*, as predictive features (Fig. 4.5B). This suggests that while these ARGs are present and variable within the dataset, their patterns are not consistent enough to build a reliable predictive model for dental caries.

4.2.1.3.2 Periodontitis-Specific Resistome Model Performance

In contrast to the caries cohort, the ARG profiles from the periodontitis cohort proved to be highly predictive of disease status, as with the taxonomic models. The initial screening phase showed good performance across several models, with Random Forest classifiers consistently achieving the highest scores. The top-performing configurations yielded mean cross-validated AUC-ROC

scores of approximately 0.83 (Fig 4.4B). Similar to previous analyses, the statistical comparison of these top models found no significant differences in their performance, indicating a consistent signal within the data. The best-performing model, a Random Forest Classifier combined with feature filtering and SMOTE, was carried forward for final assessment on the test set.

On the held-out test data, this final model showed strong performance, with a Test AUC-ROC of 0.78 (Fig. 4.5C). The confusion matrix further demonstrated the model's effectiveness, showing that it correctly identified 27 out of 30 periodontitis samples, meaning a high sensitivity for the disease state (Fig. 4.5C). While it was less precise in identifying healthy samples, correctly classifying 9 out of 16, the overall performance confirms that the periodontitis resistome contains a reliable predictive signature.

Analysis of the model's feature importances revealed the specific ARGs that were most predictive of periodontitis (Fig. 4.5D). The most influential feature was *novA*, a gene conferring resistance to novobiocin, followed by *msbA*, which is associated with multidrug resistance. The model also identified genes conferring resistance to clinically relevant antibiotic classes, including sulphonamides (*sul2*) and tetracyclines (*tetB(58)*), as highly predictive. The importance of these specific ARGs demonstrates that, beyond a simple shift in taxonomic composition, the dysbiotic state in periodontitis is characterised by the selection of a distinct and functionally relevant resistome that can be successfully used for disease prediction.

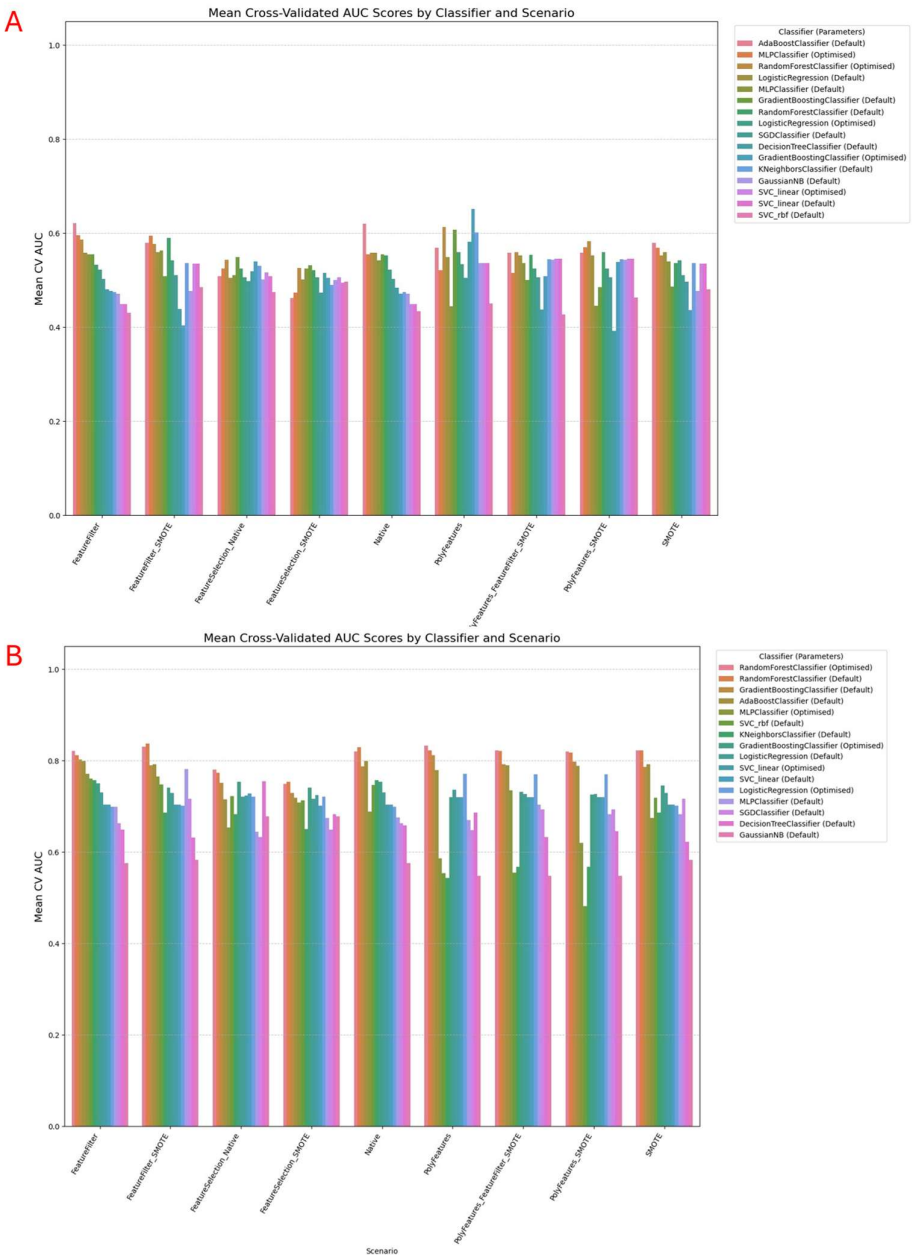


Figure 4.4. Initial screening of machine learning models for disease classification based on ARG resistome profiles. Bar charts show the mean cross-validated Area Under the Curve (AUC-ROC) scores for various classifiers and preprocessing scenarios. **(A)** Caries vs. Health cohort, showing limited predictive performance with top models achieving AUC-ROC scores of ~0.65. **(B)** Periodontitis vs. Health cohort, showing good predictive performance with top models achieving AUC-ROC scores of ~0.83.

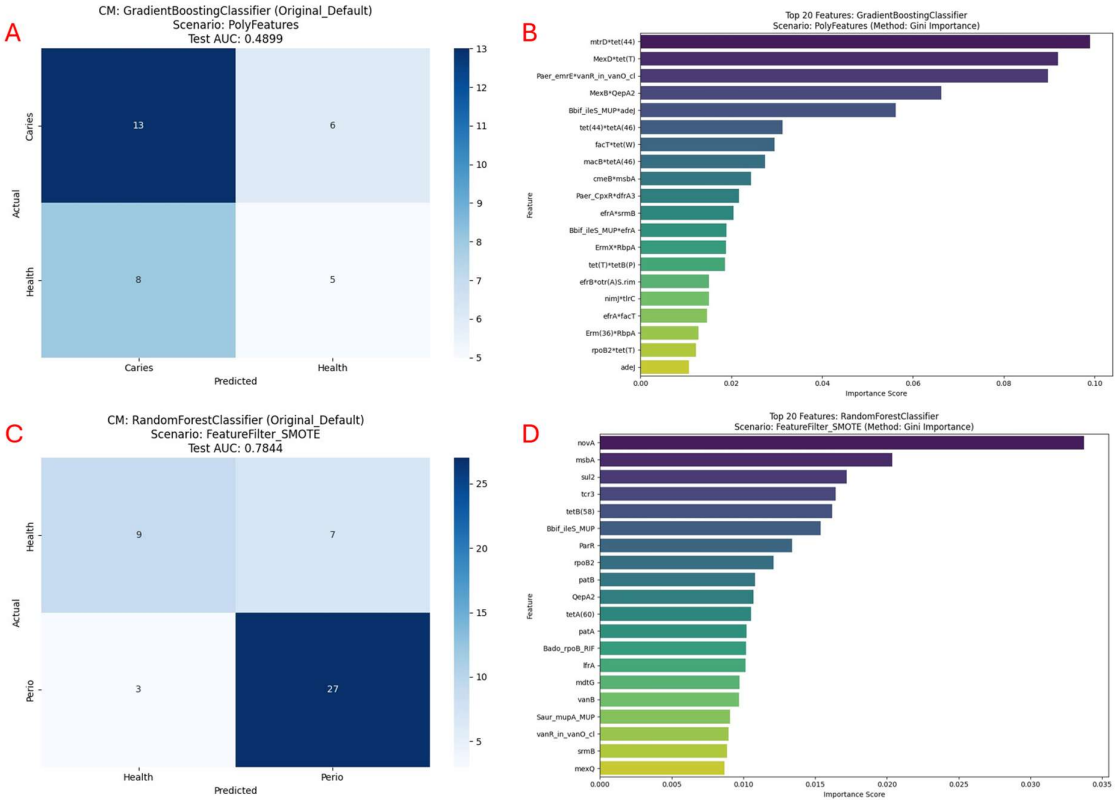


Figure 4.5. Final test set performance and feature importance for the top-performing resistome models. (A/B) Results for the Caries vs. Health model (Gradient Boosting Classifier). (A) Confusion matrix showing the model's poor performance on the unseen test set (AUC-ROC = 0.49). (B) The top 20 most important predictive features (ARGs and their interactions). **(C/D)** Results for the Periodontitis vs. Health model (Random Forest Classifier). (C) Confusion matrix showing the model's strong performance on the unseen test set (AUC-ROC = 0.78). (D) The top 20 most important predictive features (ARGs).

4.2.2 Final Statistical Comparison of Model Performances (Traditional Models)

To formally assess the predictability of each disease state and data type, final statistical comparisons were performed on the cross-validation AUC-ROC scores from the best-performing traditional model configuration for each analysis.

4.2.2.1 Cross-Project Comparison: Periodontitis vs. Dental Caries

A Mann-Whitney U test was used to compare the performance of the best models trained on periodontitis data versus those trained on caries data for each data type. The results consistently and decisively showed that periodontitis was a significantly more predictable disease state. The models trained on periodontitis data were statistically better for both Species (Mean AUC-ROC 0.91 vs 0.66, *padj* = 0.0179) and ARGs (Mean AUC-ROC 0.84 vs 0.65, *padj* = 0.0159) (Fig 4.6A/B).

4.2.2.2 Within-Project Comparison: Periodontitis Data Types

Within the periodontitis cohort, the predictive performances of models trained on different data types were compared. The analysis revealed no statistically significant difference in predictive power between the model trained on the Species profile (Mean AUC-ROC 0.91) and the model trained on the ARG profile (Mean AUC-ROC 0.84), although the taxonomic model trended towards better performance (*padj* = 0.0556) (Fig 4.6D). This suggests that while both data types contain a strong predictive signal, neither is statistically different to the other for classifying periodontitis.

4.2.2.3 Within-Project Comparison: Caries Data Types

A similar comparison for the caries cohort found no statistically significant difference in performance between models trained on Species (Mean AUC-ROC 0.66) and ARGs (Mean AUC-ROC 0.65) (*padj* = 1.0) (Fig 4.6C). This indicates that for dental caries as neither the taxonomic nor the resistome profile offers a predictive advantage, both provide a similarly moderate and limited signal for disease classification.

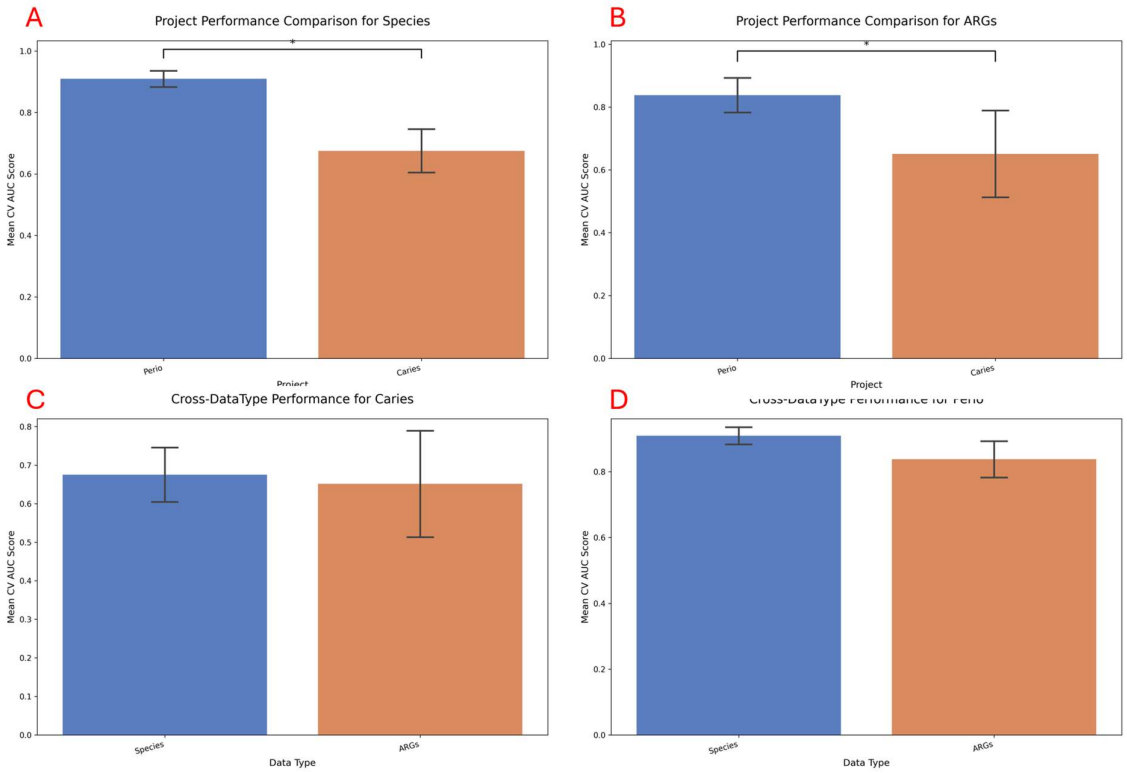


Figure 4.6. Statistical comparison of model performances based on mean cross-validated AUC scores. Bar charts show the mean AUC-ROC scores with error bars. Asterisks indicate a statistically significant difference ($p < 0.05$). **(A)** Comparison between Periodontitis and Caries models using Species data. **(B)** Comparison between Periodontitis and Caries models using ARG data. **(C)** Comparison between data types (Species vs. ARGs) for the Caries cohort. **(D)** Comparison between data types (Species vs. ARGs) for the Periodontitis cohort.

4.2.3 Results - Deep Learning Model

A deep learning approach using a neural network was employed in parallel to explore if this methodology could capture more complex, non-linear patterns within the data. A 10-fold cross-validation strategy was used to provide a better estimate of performance.

4.2.3.1 Taxonomic Profiles

The deep learning models reinforced the disease-specific nature of the taxonomic signatures. For the caries cohort, the initial screening of preprocessing pipelines showed no significant performance differences, with the best scenario (Filtered and SMOTE) yielding a mean cross-validated AUC-ROC of 0.57. The final hyperparameter-tuned model achieved a low Test AUC-ROC of 0.64 (Fig. 4.7A). The confusion matrix revealed balanced but low accuracy, correctly identifying 7 of 14 healthy samples and 14 of 19 caries samples (Fig. 4.7A). Feature importance identified *Penicillium salamii* and various bacteriophages as the most influential features (Fig. 4.7B).

In contrast, the periodontitis model performed strongly. The best initial pipeline was native and scaled, yielding a mean CV AUC-ROC of 0.77. The final tuned model achieved a Test AUC-ROC of 0.76 (Fig. 4.7C). The classification report showed high precision (0.83) and recall (0.77) for the periodontitis class (Fig. 4.7C). The most important features identified were *Pichia kudriavzevii* and *Ralstonia pickettii* (Fig. 4.7D).

The joint three-class model achieved a strong Test AUC-ROC of 0.81 (Fig. 4.7A). The confusion matrix mirrored that of the Random Forest model, it excelled at identifying periodontitis (26/31 correct) and health (18/29 correct) but struggled to separate caries from health (only 8/19 caries samples were correctly identified, with 10 misclassified as healthy) (Appendix 4.1).

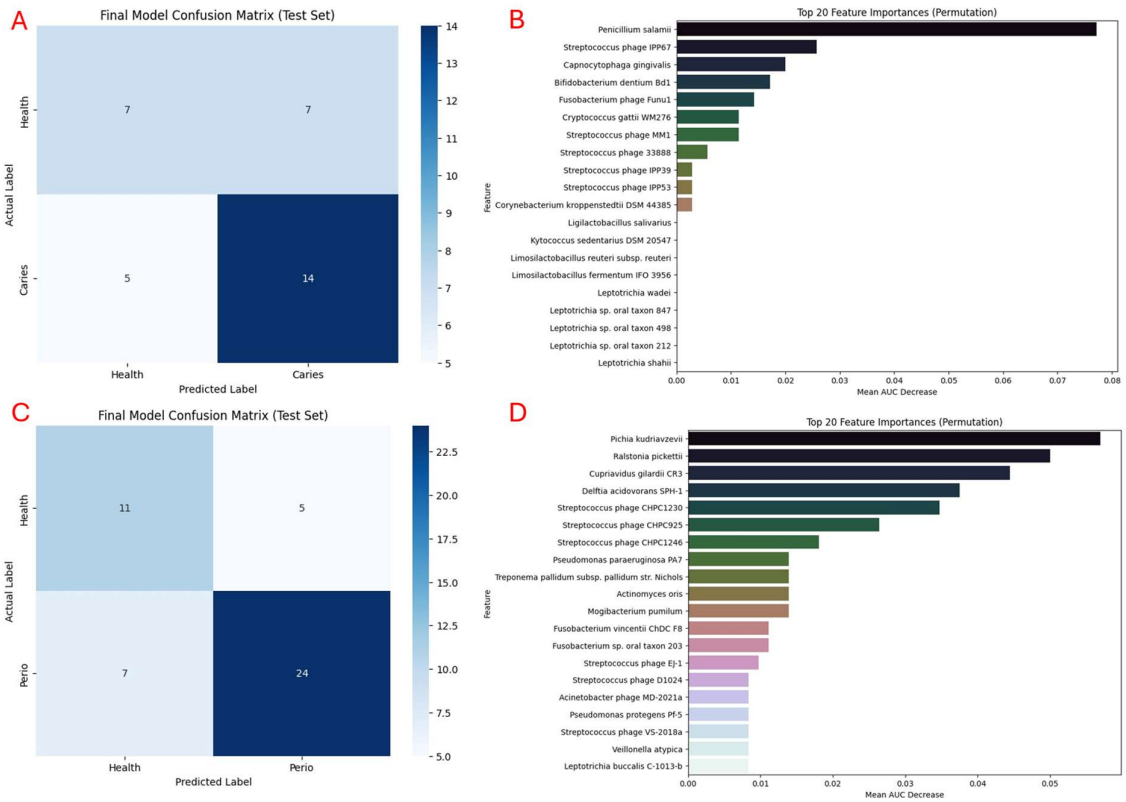


Figure 4.7. Performance and feature importance of deep learning models using taxonomic profiles. (A/B) Results for the Caries vs. Health model. (A) Confusion matrix on the held-out test set (Test AUC-ROC = 0.64). (B) The top 20 most important predictive features (taxa) determined by permutation importance. **(C/D)** Results for the Periodontitis vs. Health model. (C) Confusion matrix on the held-out test set (Test AUC-ROC = 0.76). (D) The top 20 most important predictive features (taxa).

4.2.3.2 Resistome Profiles

The functional resistome profiles showed a similar difference between diseases. For the caries cohort, the model performed poorly, achieving a Test AUC-ROC of 0.49, indicating it could not generalise from the training data (Fig. 4.8A). The initial screening showed that a combination of feature filtering and SMOTE provided a slight, non-significant benefit (Mean CV AUC-ROC: 0.53).

For the periodontitis cohort, the resistome was more predictive, but still performed weakly. The model achieved a Test AUC-ROC of 0.58, with the initial screening again favouring a Filtered and SMOTE approach (Mean CV AUC-ROC: 0.74) (Fig. 4.8B). The joint model for the resistome performed modestly, with a final Test AUC-ROC of 0.65. The confusion matrix showed some ability to identify all three classes, but with significant overlap (Appendix 4.2).

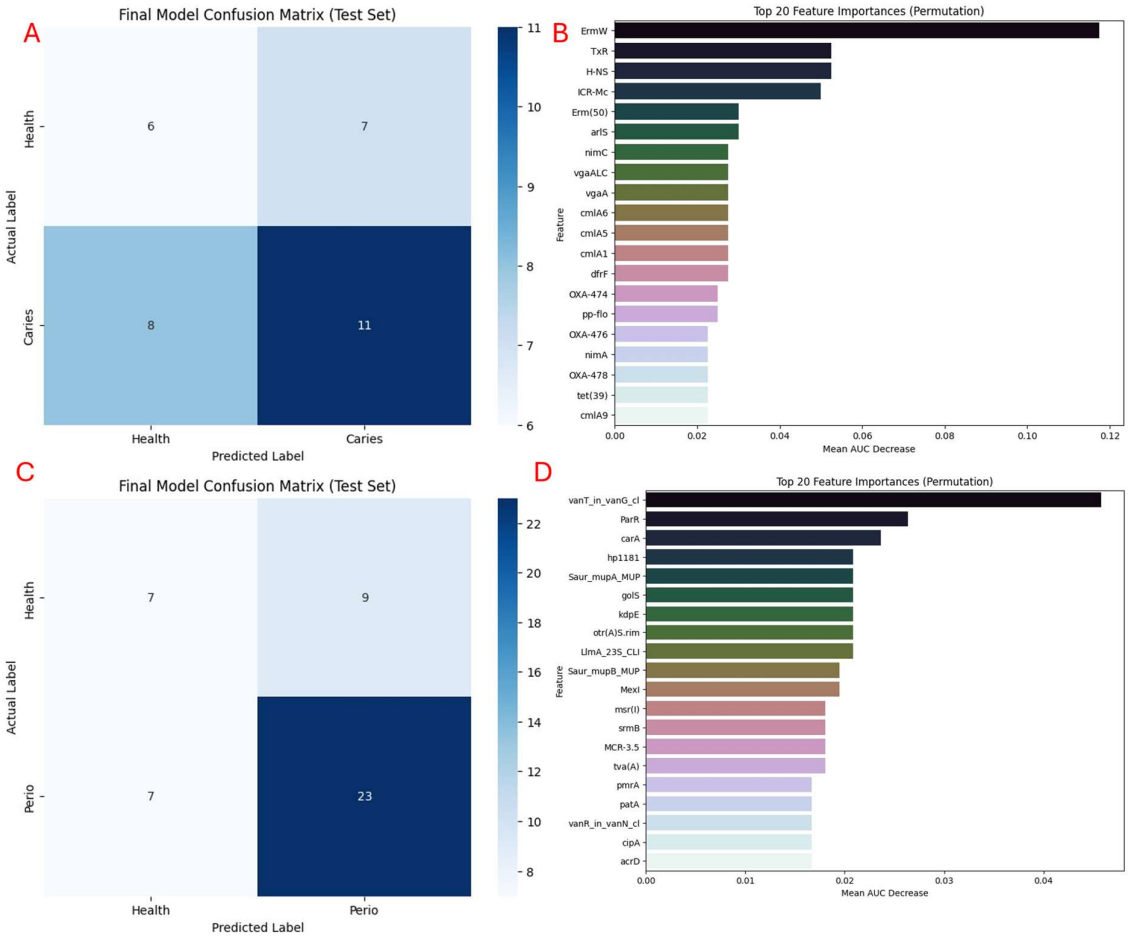


Figure 4.8. Performance and feature importance of deep learning models using resistome (ARG) profiles. (A/B) Results for the Caries vs. Health model. (A) Confusion matrix on the held-out test set (Test AUC-ROC = 0.49). (B) The top 20 most important predictive features (ARGs) determined by permutation importance. **(C/D)** Results for the Periodontitis vs. Health model. (C) Confusion matrix on the held-out test set (Test AUC-ROC = 0.58). (D) The top 20 most important predictive features (ARGs).

4.2.3.3 Functional Gene Profiles

For periodontitis, the gene profile, comprising over 53,000 features, proved to be one of the most predictive data types. The initial screening revealed that the filtered and SMOTE pipeline was clearly optimal, achieving a high mean CV AUC-ROC of 0.89. The final hyperparameter-tuned neural network achieved a Test AUC-ROC of 0.82 (Fig. 4.9A). The model demonstrated high recall (0.87) for the periodontitis class, correctly identifying 27 out of 31 periodontitis samples in the test set (Fig. 4.9A). Feature importance analysis of the validation set data identified a hypothetical protein (*HMPREF0733 11021*) as the top predictor (Fig. 4.9B).

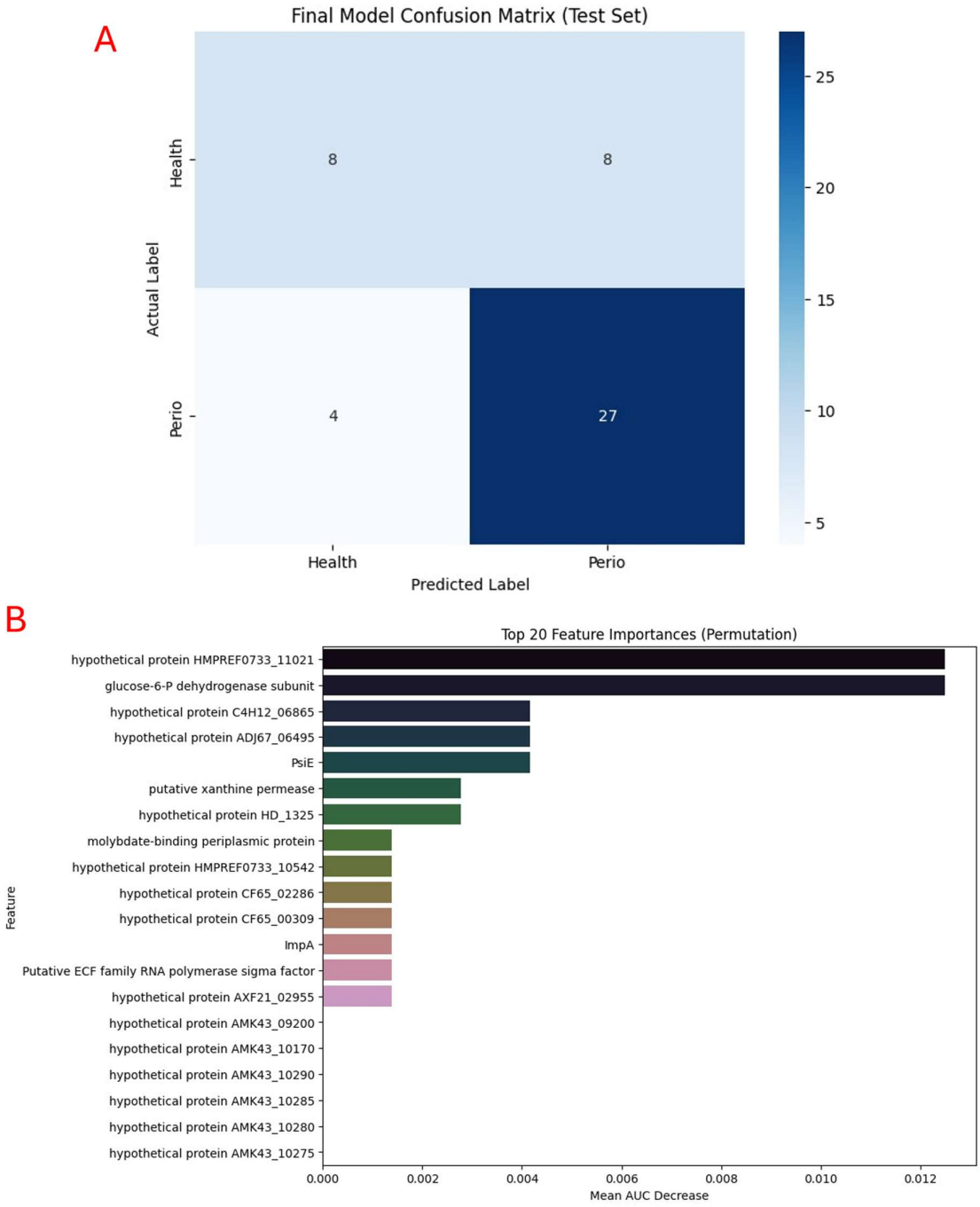


Figure 4.9. Performance and feature importance of the deep learning model using gene profiles for periodontitis classification. (A) Confusion matrix for the final model on the held-out test set, achieving a Test AUC-ROC of 0.82. **(B)** The top 20 most important predictive features (genes) determined by permutation importance, with a hypothetical protein being the most influential.

4.2.3.4 Mobilome (Plasmid) Profiles

The predictive use of the oral mobilome was weaker than that of other profiles but still demonstrated the characteristic split between the two diseases. The caries-mobilome model performed poorly, with a final Test AUC-ROC of 0.63 (Fig. 4.10A). The SMOTE pipeline performed best during initial screening (Mean CV AUC-ROC: 0.69), and the final confusion matrix showed the model correctly identified 12 of 19 caries samples but only 5 of 13 healthy samples, indicating a bias (Fig. 4.10A).

Conversely, the periodontitis-mobilome model was a standout success, achieving an excellent Test AUC-ROC of 0.84 (Fig. 4.10C). The best initial pipeline was Filtered and SMOTE (Mean CV AUC-ROC: 0.85). The confusion matrix for the final model showed high accuracy, correctly identifying 27 of 31 periodontitis samples and 11 of 16 healthy samples (Fig. 4.10). The most important feature was a plasmid from *Pseudomonas aeruginosa* (Fig. 4.10D).

The joint three-class model reinforced this narrative, achieving a modest Test AUC-ROC of 0.68 (Fig. 4.10). The confusion matrix again showed an ability to identify periodontitis (22/31 correct) but significant confusion between caries (11/19 correct) and health (9/29 correct) (Appendix 4.3).

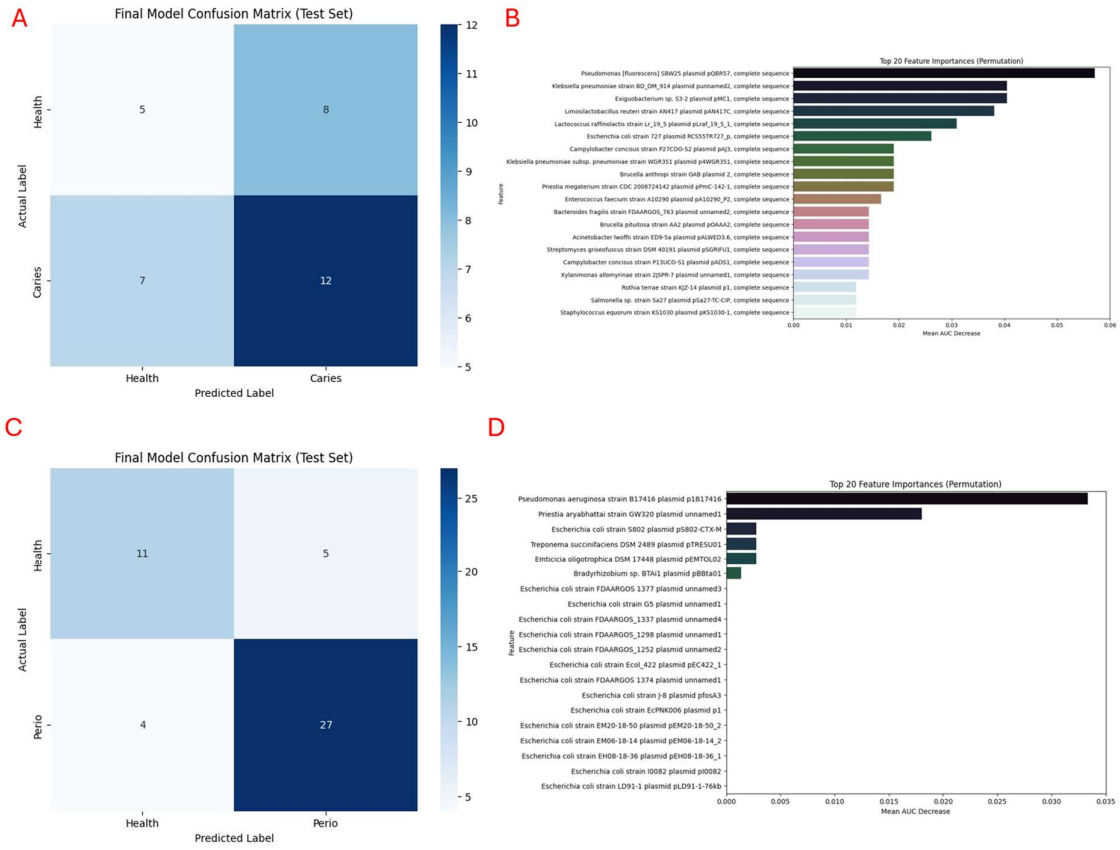


Figure 4.10. Performance and feature importance of deep learning models using mobilome (plasmid) profiles. (A/B) Results for the Caries vs. Health model. (A) Confusion matrix on the held-out test set (Test AUC-ROC = 0.63). (B) The top 20 most important predictive features (plasmids) determined by permutation importance. **(C/D)** Results for the Periodontitis vs. Health model. (C) Confusion matrix on the held-out test set (Test AUC-ROC = 0.84). (D) The top 20 most important predictive features (plasmids).

4.2.3.5 Final Statistical Comparison of Model Performances (Deep Learning Models)

To formally assess the predictability of each disease state and data type, final statistical comparisons were performed on the 10-fold cross-validation AUC-ROC scores from the best-performing deep learning model configuration for each analysis.

4.2.3.6 Cross-Project Comparison: Periodontitis vs. Dental Caries

A Wilcoxon signed-rank test was used to compare the performance of models trained on periodontitis data versus those trained on caries data for each data type. The results consistently showed that periodontitis was a significantly more predictable disease state. The periodontitis models were statistically better for Species (*padj* = 0.0098), Plasmids (*padj* = 0.0195), and ARGs (*padj* = 0.0195) (Fig 4.11C/D/E).

4.2.3.7 Within-Project Comparison: Periodontitis Data Types

Within the periodontitis cohort, model performances across different data types were compared using ANOVA, followed by a Tukey's HSD test. The analysis revealed a statistically significant difference in predictive power among the data types (ANOVA *padj* = 0.017). Post-hoc testing confirmed that the model trained on the total functional gene profile performed significantly better than the model trained on the resistome (ARG) profile alone (*padj* = 0.021). No other pairwise comparisons reached statistical significance (Fig 4.11B).

4.2.3.8 Within-Project Comparison: Caries Data Types

A similar comparison was performed for the caries cohort. In this case, no statistically significant differences were found in the predictive performance between models trained on Species, ARGs, genes or Plasmids (ANOVA *padj* = 0.213) (Fig 4.11A). All post-hoc pairwise comparisons were also non-significant, indicating that no single data type offered a superior predictive advantage for classifying dental caries.

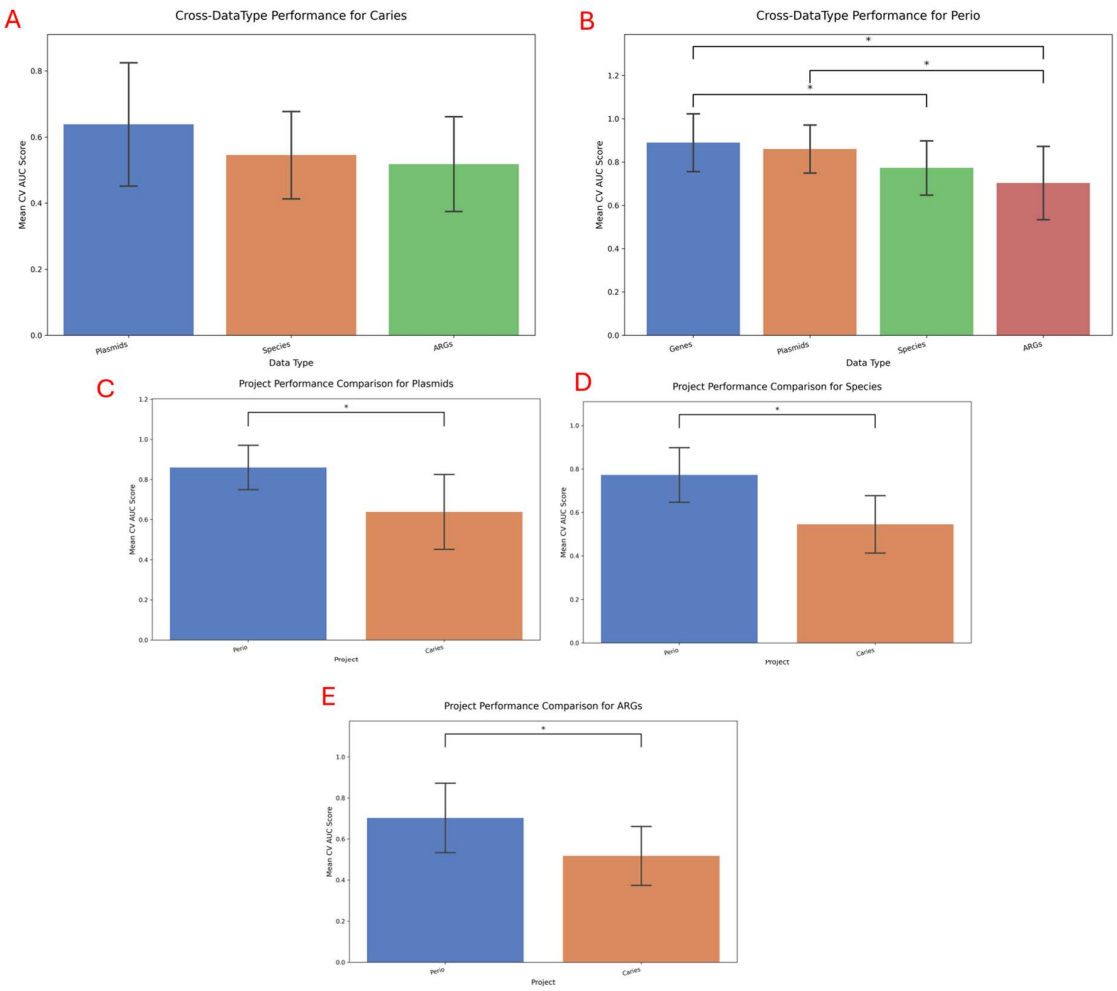


Figure 4.11. Statistical comparison of deep learning model performances based on mean 10-fold cross-validated AUC-ROC scores. Bar charts show the mean AUC-ROC scores with error bars. Asterisks indicate a statistically significant difference ($p < 0.05$). **(A)** Within-project comparison of model performance across data types for the Caries cohort. **(B)** Within-project comparison of model performance across data types for the Periodontitis cohort. **(C/D/E)** Cross-project comparisons between Periodontitis and Caries models using **(C)** Plasmids, **(D)** Species, and **(E)** ARGs as features.

4.3 Discussion

This chapter aimed to move beyond descriptive analyses of the oral microbiome, resistome, and mobilome to decipher their predictive utility for distinguishing states of health and disease. By employing a systematic method comparing both traditional machine learning and a deep learning approach, we aimed to quantify the strength of the taxonomic and genetic signatures identified in the previous chapter. The findings presented here provide quantitative evidence that, while both diseases represent a shift from oral eubiosis, their respective dysbiotic signatures yield different outcomes in a predictive modelling context.

A theme that emerged across both methodologies is the difference between the two oral diseases. Periodontitis shows as a community-wide ecological shift that is readily detectable and predictable. In contrast, dental caries presents as a more subtle imbalance that remains challenging to distinguish from health beyond the clinical observations. This is evidenced by the primary finding of this study, with the clear disparity in predictive performance between periodontitis and dental caries when using taxonomic profiles.

The high predictive accuracy for periodontitis, with traditional models achieving a Test AUC-ROC of 0.86, directly reflects the ecological shift described in the previous chapter. This performance is consistent with other machine learning investigations of periodontitis, which frequently report high AUC-ROC values based on microbial signatures (Kim *et al.*, 2020, Kageyama *et al.*, 2025). More recent large-scale studies have further validated this, with one multi-cohort analysis developing a universal periodontitis signature that achieved an external validation AUC-ROC of 0.85, and another predicting disease progression with an AUC-ROC of 0.88 by integrating clinical data with salivary biomarkers (Geng *et al.*, 2024, Furquim *et al.*, 2025).

The algorithms used in this study, particularly methods like AdaBoost and Random Forest, are well-suited for high-dimensional microbiome data. They are inherently resilient to the inclusion of many non-informative features and excel at capturing complex, non-linear relationships between taxa (Papoutsoglou et al., 2023). The success of the AdaBoost classifier, which finds

optimal decision boundaries by focusing on difficult to classify samples, suggests that the separation between the healthy and periodontitis-associated microbiomes is not as clear-cut as the results suggest, but is successfully able to separate them. The observation that pre-tuned models often performed as well as their hyperparameter-tuned versions implies that the biological signal is strong enough that even default algorithm parameters are sufficient to capture it effectively.

The feature importance analysis from the joint Random Forest model provides a biological explanation for the high predictability of periodontitis. The model's most predictive features were dominated by a consortium of established periodontal pathogens, including *Parvimonas micra*, various *Fusobacterium* species, and emerging *Treponema putidum*. These organisms are markers of the subgingival dysbiotic community and are strongly associated with inflammation and tissue destruction (Rams, Sautter and van Winkelhoff, 2020, Chen *et al.*, 2022). *P. micra*, in particular, is considered an inflammophilic pathogen, whose presence is a significant predictor of poor treatment outcomes, and which can enhance the virulence of other pathogens like *Porphyromonas gingivalis,* having been shown to stimulate expression of *P. gingivalis* gingipains (Neilands et al., 2019). Similarly, *Fusobacterium nucleatum* acts as a critical bridging organism essential for the structural development of the pathogenic biofilm by allowing the co-aggregation of early and late colonisers (Thurnheer et al., 2019). Recent work has even distinguished between its subspecies, suggesting *F. nucleatum* subsp. *animalis* is the most clinically relevant at sites of dental abscesses, in contrast to the plaque-associated *F. nucleatum* subsp. *polymorphum* (Chen et al., 2022; Krieger et al., 2024). The model essentially learned to recognise the signature of this pathogenic community. Its success further supports the biological relevance of these species as biomarkers, supporting the 'keystone pathogen' and 'polymicrobial synergy' hypotheses, which suggests specific pathogens initiate a community-wide shift towards a disease-promoting state, and the presence of multiple pathogenic species worsens disease prognosis (Hajishengallis and Lamont, 2012, Hajishengallis, Darveau and Curtis, 2012). Interestingly, in the deep learning model for periodontitis, we identified microbes not traditionally associated with periodontal disease. The model identified the yeast *Pichia kudriavzevii*, an emerging high-priority fungal pathogen, and the opportunistic nosocomial bacterium *Ralstonia pickettii* as the most important predictive

features for periodontitis (Ryan, Pembroke and Adley, 2006, Nguyen *et al.*, 2024). The model's selection is interesting, suggesting that the ML detected subtle ecological signatures that do not fit our conventional understanding of periodontitis pathology. These findings may represent novel biomarkers or indicate that the dysbiotic periodontal environment creates a niche for unusual opportunistic microbes, with further investigation required.

In contrast, the models trained on the caries-associated microbiome demonstrated only moderate and less consistent predictive performance, with a final Test AUC-ROC of 0.69. This level of performance, while better than random chance, is a common finding in caries prediction studies, which often struggle to achieve the high accuracy seen in periodontitis research (Wu et al., 2021). For example, a recent study predicting caries in adolescents using behavioural and demographic data achieved a somewhat comparable mean AUC-ROC of 0.77 (Väyrynen et al., 2025). While some studies focusing on specific cohorts, such as early childhood caries, have reported higher accuracy with AUC-ROCs of 0.91 and even a perfect 1.0 using ensemble learning, these results highlight that predictability is context-dependent and does not negate the general challenge of classifying caries (Wei *et al.*, 2025, Zhang *et al.*, 2025).

The performance of the joint three-class model provided the most direct confirmation of this principle. The model, while excellent at identifying periodontitis (AUC-ROC 0.88), had its primary difficulty distinguishing caries from healthy samples, frequently confusing the two. This quantitatively reinforces the hypothesis that the microbial signature of dental caries demonstrates a substantial overlap with the healthy state, reflecting a shift in transcription acidogenic pathways rather than a definitive higher abundance of these gene pathways (Takahashi and Nyvad, 2011, Zhang, Chu and Yu, 2022). The identification of *Hoylesella enoeca* as a top caries predictor is interesting. This species, formerly classified within the proteolytic genus *Prevotella*, is frequently found in dysbiotic biofilms associated with advanced carious lesions (Könönen *et al.*, 2021). This aligns with the 'ecological plaque hypothesis', which holds that caries results from environmental changes (frequent dietary sugar intake) that favour the growth of acid-producing bacteria already present in the healthy microbiome, rather than the invasion of exogenous pathogens (Marsh, 2003). Additionally, the identification of specific bacteriophages, such as *Streptococcus phage*

IPP61, as key predictive features in the caries model may suggest that phages are not passive bystanders in the oral ecosystem but can actively shape the bacterial community through predation whilst acting as vectors for horizontal gene transfer (HGT), mobilising virulence and resistance genes (Dion, Oechslin and Moineau, 2020, Naureen *et al.*, 2020). Despite the low accuracy, their importance in the caries predictive model suggests that the virome is an active participant in the shift towards dysbiosis, contributing to the instability that characterises the disease. Future studies should focus more on the virome rather than just on bacteria.

The predictive analysis of the oral resistome mirrored the taxonomic findings. The periodontitis resistome contained a strong, generalisable signal, with a Random Forest model achieving a Test AUC-ROC of 0.78. This is a significant finding, as it demonstrates that the functional potential encoded in the accessory genome is as predictive as the core taxonomic profile. The identification of genes like *novA* (novobiocin resistance) and *msbA* (multidrug resistance) as highly predictive features suggests that the dysbiotic state is characterised not only by pathogenic taxa but also by a functionally relevant accessory genome geared towards survival and competition in an inflamed environment (Warinner *et al.*, 2014, Sukumar *et al.*, 2024). The *msbA* gene encodes a broad-spectrum efflux pump, homologous to mammalian multidrug resistance proteins, which serves as a detoxification mechanism against host-derived antimicrobial compounds (Dawson and Locher, 2006). Similarly, the presence of *sul2* is often a marker for mobile plasmids carrying resistance to multiple antibiotics like ampicillin and trimethoprim, indicating co-selection (Perreten and Boerlin, 2003). The identification of tetracycline resistance genes is also clinically relevant, as they are markers of resistance to a commonly used class of periodontal antibiotics and are frequently spread via broad-host-range transposons (Seymour and Heasman, 1995, Roberts, 1996).

Conversely, the complete failure of the caries resistome model to generalise (Test AUC-ROC of 0.49) is a critical finding. It suggests that while specific ARGs may be statistically enriched in caries, their patterns are not consistent enough across the population to serve as a reliable predictive signature, as previously demonstrated by the lack of distinction when observing resistotyping results compared to periodontitis. This reinforces the idea that the primary driver of

caries is the upregulation of metabolic pathways, such as acidogenesis, rather than a shift in the community's functional potential (Baker and Edlund, 2019). This result demonstrates that statistical significance in differential abundance testing does not directly translate to predictive power when utilising machine learning.

The statistical analyses from both the traditional and deep learning frameworks provided insight into whether periodontitis is a significantly more computationally predictable disease state than dental caries. The traditional models, using a Mann-Whitney U test, showed this finding to be true across the tested data sets, Species ($padj$ = 0.0179) and ARGs ($padj$ = 0.0159), reinforcing that the signature of periodontitis is more distinct and consistent than that of caries. The deep learning analysis, using a Wilcoxon signed-rank test, confirmed this conclusion for Species (p = 0.0098), ARGs (p = 0 .0195), and Plasmids (p = 0.0195).

Furthermore, the within-project comparisons revealed additional differences. For periodontitis, while the taxonomic profile yielded a higher mean AUC-ROC than the resistome, the difference did not reach statistical significance in the traditional models ($padj$ = 0.0556), suggesting that both are powerful predictors. For caries, both analyses confirmed that neither the taxonomic nor the resistome profile offered a superior predictive advantage, with both performing at a similarly moderate level. This statistically substantiates the conclusion that for caries, a simple taxonomic or resistome-based biomarker is unlikely to be sufficient, whereas for periodontitis, both approaches hold significant promise.

4.3.1 The Role of Preprocessing and Model Complexity

The systematic evaluation of preprocessing pipelines revealed key insights into the nature of the data and its interaction with the learning algorithm. The application of SMOTE, for instance, proved most useful in scenarios with a class imbalance but a weak underlying signal, such as the caries-mobilome analysis (Mean CV AUC-ROC 0.69). Creating synthetic examples of the minority class allowed the model to train more effectively on a balanced dataset. However, its limited success in improving final test performance highlights that while SMOTE can help a model learn a weak signal, it cannot create a strong signal where none natively exists. This can

lead to the model learning features specific to the synthetic data that do not generalise to the unseen test set, a known risk of oversampling that can cause overfitting to the synthetic training set (Blagus and Lusa, 2013).

Conversely, for datasets with a strong intrinsic signal, such as the periodontitis mobilome, a simple scaled approach was optimal for the neural network. Here, SMOTE was unnecessary, suggesting that the normal data distribution already provided a clear enough signal for the model to learn. A flexible model, such as a neural network, can be overfit to these artificial data, learning patterns that exist only in the synthetic data. This can harm its ability to generalise to unseen test data, meaning the technique can hinder final performance (Krawczyk, 2016). In these cases, it is often more effective to use methods like feature weighting, which adjusts the model's loss function to pay more attention to the minority class without altering the data itself (Johnson and Khoshgoftaar, 2019).

Feature filtering was critical for the gene dataset, as the periodontitis gene profile contained>53,000 features. The increased performance of the Filtered and SMOTE pipeline (Mean CV AUC-ROC: 0.89) demonstrates that removing uninformative, low-variance genes is an essential step. This reduces the risk of overfitting and makes the computational task more feasible, allowing the model to focus on a smaller, more informative subset of features (Saeys et al., 2007).

Comparing the two modelling methods, the traditional machine learning models often performed on par with, or slightly better than, the deep learning model, particularly for the caries datasets. This suggests that, given the complexity and sample sizes of the datasets analysed, the advanced non-linear feature extraction provided by neural networks did not always offer a decisive advantage. The scikit-learn pipeline's ability to engineer features, such as the interaction terms in the periodontitis taxonomic model, was highly effective. This highlights a key difference with traditional models often benefit from manual feature engineering. At the same time, neural networks are designed to learn these features automatically, which can be a black box process (Torun *et al.*, 2023, Hassija *et al.*, 2024). The feature importance analysis from the scikit-learn models, the identification of interactions between known periodontal pathogens, and the neural

network's identification of individual pathogenic markers all validate the biological relevance of results from differential abundance testing, albeit through different methods.

4.3.2 Limitations and Future Directions

While this study successfully demonstrates the potential of machine learning for classifying oral disease states, it is essential to acknowledge the limitations. A primary drawback is the 'black box' nature of many high-performing algorithms, particularly neural networks. Although feature importance analysis provides valuable insight, it does not fully explain the complex relationships the model has learned internally (Belle and Papantonis, 2020). Future work could use more advanced explainable AI (XAI) techniques to gain a deeper, sample-specific understanding of model predictions (Belle and Papantonis, 2021, Saarela and Podgorelec, 2024). The application of XAI methods, such as SHAP is becoming increasingly utilised for enhancing the transparency and trustworthiness of clinical AI models, allowing clinicians to understand the reasoning behind individual predictions (Novielli et al., 2024).

Furthermore, while the use of a held-out test set provides a measure of generalisability, actual robustness can only be confirmed through validation on external, prospectively collected cohorts, which may be possible through the collection of new publicly available sequencing data (Peng et al., 2025). Uncontrolled metadata, such as diet and host genetics, were not uniformly available in the metadata and therefore could not be accounted for.

Additionally, these models identify predictive correlations, not causal relationships, and may not have deeper biological meaning when further explored. An important limitation of using ML as a potential diagnostic, even with high accuracy, is the inherent errors it produces. False positive errors, where a health sample is classified as a disease, may have adverse outcomes such as patient anxiety, as well as unnecessary economic burden both on the patient and the wider healthcare system. However, these are often more pronounced in life-threatening diseases such as cancer (Brodersen and Siersma, 2013, 'Artificial Intelligence in Skin Cancer Diagnosis: A Reality Check', 2024). Conversely, false negatives, where a disease sample is classified as healthy, may lead to delayed treatment, ultimately meaning worse prognoses, which is detrimental in

progressive diseases like caries and periodontitis. Optimising models to minimise false positives or negatives can often lead to more false classifications in the other direction, meaning risk must be determined for the specific disease and scenario as to which is deemed more acceptable (Maxim et al., 2014).

Furthermore, an interesting finding from the high-resolution gene-level analysis was the identification of a hypothetical protein as the most predictive feature for periodontitis. The consistent appearance of this and other genes of unknown function in the top feature lists highlights a significant gap in our understanding of the oral functional metagenome. While these uncharacterised proteins serve as biomarkers for the model, their biological roles in the pathogenesis of periodontitis remain unknown, showing the need for further research to uncover their function and potential as therapeutic targets (Galperin and Koonin, 2004).

A final issue arose in the classification of gene and plasmid profiles: both locally and HPC running algorithms failed to find optimised classifiers for periodontitis and caries in the traditional models and for caries in the neural network deep learning models, highlighting the complexity of this work. This was due to limitations in the available resources locally, and the time limit/priority restrictions set on the HPC. Future work may utilise dedicated nodes for successful completion of these models.

The findings and limitations of this work open several paths for future research. A clear next step is the integration of multi-omics data. While this study analysed taxonomic, resistome, and mobilome profiles separately, future models should combine these feature sets to investigate whether an integrated profile provides synergistic predictive power. The moderate success in predicting caries suggests that taxonomic data alone may be insufficient, though incorporating metatranscriptomic or metabolomic data could provide the functional information needed to capture the metabolic activity that is associated with this caries (Costalonga and Herzberg, 2014, Duran-Pinedo, 2021). Furthermore, longitudinal studies are essential to move beyond the cross-sectional snapshots analysed here to develop models capable of predicting disease onset and progression.

4.3.3 Clinical Implications and Conclusion

From a clinical perspective, the ultimate goal is the translation of these ML models into diagnostic tools. The identification of a small subset of highly predictive features could form the basis for the development of simpler, targeted diagnostics. For periodontitis, for instance, a qPCR panel targeting the top 10-15 taxonomic, resistome, and gene features identified by the models could provide a rapid, non-invasive, and cost-effective method for assessing disease status or risk. This represents a potential pathway from high-dimensional data analysis to evidence-based dentistry.

In conclusion, this chapter successfully leveraged two distinct machine learning paradigms to quantify the predictive potential of oral metagenomic signatures. The results from both traditional models and a deep learning approach demonstrated the same conclusion that periodontitis is defined by a highly predictive ecological and functional shift, whereas dental caries represents a more subtle dysbiosis that is significantly harder to distinguish from health. The failure of the caries resistome and mobilome to provide generalizable predictions, in contrast to their more apparent success in periodontitis, further evidences the fundamental differences in the microbiology of these two oral diseases. This work highlights the potential of machine learning as a powerful tool for developing non-invasive diagnostics, whilst providing a potential avenue into deepening our understanding of the complex microbial dynamics in oral health and disease.

Chapter 5: Analysis of Metagenomic and Metatranscriptomic Changes across a Complex Oral Co-Culture Model

5.1 Introduction

The previous chapters have demonstrated the use of bioinformatic and machine-learning approaches to understand the complexity of the oral microbiome using publicly available metagenomic data. Through this analysis, distinct taxonomic, resistome, and mobilome signatures were identified that can effectively classify dysbiotic states in dental caries and periodontitis. However, they are less effective at distinguishing health from caries samples. However, a fundamental limitation of in silico cross-sectional studies is that they provide a single snapshot of the disease state, with samples collected only from patients once they are clinically diseased. While they can identify predictive correlations, they are unable to fully explain the longitudinal processes that drive the transition from eubiosis to dysbiosis and therefore cannot detect the intricate host-microbe interactions that both result from and worsen this shift (Kodikara et al., 2022). The reliance on microbial sequencing data alone, without the context of host response, leaves a critical gap in our understanding of disease progression (Mountcastle et al., 2020).

To bridge this gap and transition from correlation to mechanistic understanding, there is a clear need for advanced models that can replicate the key features of the oral environment in a controlled laboratory setting. While traditional *in vitro* models have been invaluable, their oversimplification, often using a limited consortium of microbes, restricts their ability to reflect the complex polymicrobial and host-interactive nature of oral diseases (Sánchez *et al.*, 2011, Do, Devine and Marsh, 2013). To understand the development of dysbiosis, a model must integrate greater microbial complexity derived from human oral microbiome samples, incorporate the host components that may affect the microbial environment, and facilitate the longitudinal study of a dysbiotic shift over time (Naginyte *et al.*, 2019, Vila *et al.*, 2019).

This chapter, therefore, describes the use of an advanced co-culture model, designed specifically to address the limitations of cross-sectional analyses by building on established protocols with key novel adaptations. This system utilises a complex, multi-species biofilm cultivated from a

pooled inoculum of healthy human donors using specific health- and dysbiosis-associated media (Naginyte et al., 2019). However, whereas previous use of this biofilm model cultivated healthy and dysbiotic biofilms in isolated experiments, this study introduces a novel longitudinal design. A mature, representative healthy biofilm was first established over 14 days using health-associated media, after which dysbiotic shift was induced by introduction of periodontitis-associated media. Once mature, these biofilms were co-cultured together with a 3D stratified oral epithelial tissue model, adapted from an established collagen gel protocol (Hewitt et al., 2022). To more accurately reflect the specific oral environment, the protocol was adapted by incorporating hTERT gingival fibroblasts and dysplastic oral keratinocytes. By collecting samples at multiple time points across this induced transition and subsequently co-culturing the biofilm with stratified epithelial layers, this novel setup provided a unique platform to perform longitudinal analysis of the associated dynamic taxonomic and functional changes as well as the associated host response.

The primary aims of this chapter are multifaceted. We seek to characterise the metagenomic and metatranscriptomic shifts within the biofilm's microbiome, resistome, and mobilome as it transitions from a state of health to dysbiosis, and to investigate the corresponding transcriptomic response of the host epithelial tissue to this evolving microbial community.

5.2 Results

5.2.1 Sequencing Data Overview

Shotgun metagenomics of the co-culture samples achieved a mean depth of 35.5 million reads (±0.6 million) for DNA sequencing, whilst RNA sequencing had a mean depth of 35.5 million reads, though with a greater standard deviation of ±1.3 million. This high throughput data provided the basis for the subsequent multi-omics characterisation of biofilm and host response through disease progression.

5.2.2 Metagenomic Analysis

5.2.2.1 Alpha and Beta Diversity

To characterise the development of dysbiosis in the host-biofilm co-culture model, the microbial community dynamics were assessed at the DNA level over the 14-day experimental period. Alpha and beta diversity metrics were calculated from the species-level profiles to evaluate changes in community richness, evenness, and overall structure following the induction of a dysbiosis at Day 0. The alpha diversity of the microbial community, which reflects the diversity within each sample, displayed a dynamic trend throughout the time course (Fig 5.1). Species richness, as measured by the Observed species index, showed an initial decrease from the inoculum to Day 0. This suggests that not all species present in the planktonic material were successful in the initial attachment and colonisation of biofilm formation. Following this, richness remained relatively stable before declining again in the later stages of the experiment. A similar pattern was observed for community evenness, with the Shannon, Simpson, and Inverse Simpson indices all showing an increase in diversity from the inoculum to Day 2, followed by a gradual decline, particularly by Day 14. This decline in evenness suggests the dominance of fewer successful species in the mature dysbiotic biofilm. However, despite these observable trends, pairwise statistical comparisons between other time points and Day 0 revealed that none of these shifts in alpha diversity were statistically significant ($p > 0.05$ for all comparisons). The lack of statistical significance suggests that while the identity of dominant species may be changing, the overall richness and evenness of the community at the DNA level did not collapse or expand during the 14-day period.

In contrast, the beta diversity, which measures the differences in community composition between samples, revealed a statistically significant shift over the 14-day period ($p = 0.001$) (Fig 5.1). This indicates that while the number of species remained relatively constant, the specific species present changed. This was visualised using a PCoA of Bray-Curtis dissimilarities, which illustrated a clear shift. The inoculum community was clearly separated from all subsequent biofilm time points, highlighting the difference between the planktonic and biofilm states. Following the establishment of the biofilm, the community structure at Day 0 and Day 2 clustered

tightly together, representing an early-stage, stable community. A small shift was observed from Day 5 onwards, forming a second, distinct cluster that separated from the earlier time points along the primary PCoA axis. This clear separation of the community structure over time was highly significant, with a PERMANOVA analysis confirming that the time point accounted for over 75% of the observed variance in community composition ($R^2 = 0.761$). This result provides strong evidence that the dysbiosis media induced at Day 0 successfully drove change, leading to the development of a distinct, late-stage dysbiotic microbial community in the *in vitro* model.

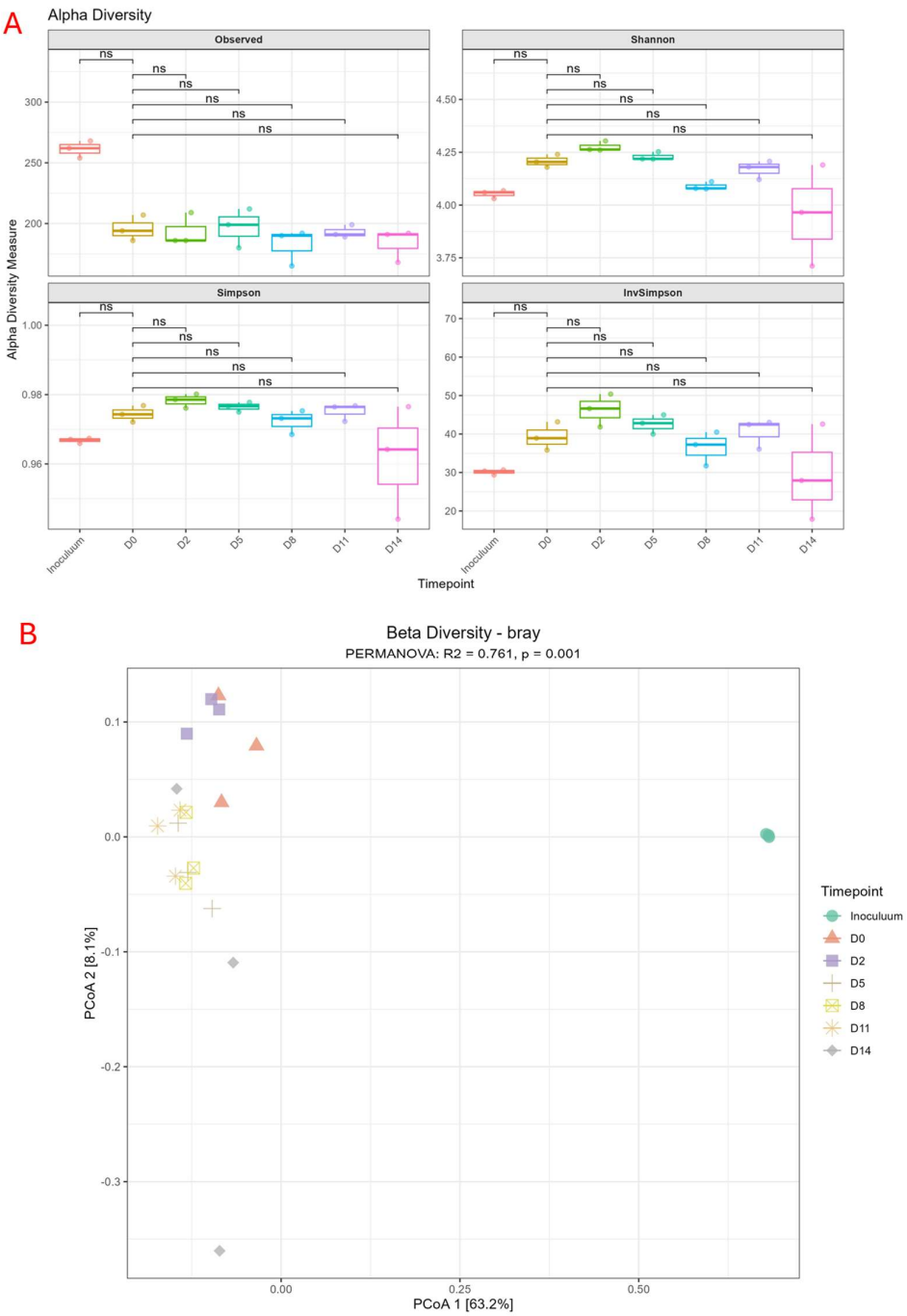


Figure 5.1. Microbial community diversity dynamics over a 14-day in vitro dysbiosis model. (A) Alpha diversity metrics (Observed, Shannon, Simpson, and Inverse Simpson) over time. Boxplots show the distribution of values for each timepoint. Brackets labelled 'ns' indicate no significant difference in pairwise comparisons to Day 0 ($p > 0.05$). **(B)** Beta diversity shown as a Principal Coordinate Analysis (PCoA) plot of Bray-Curtis dissimilarities. Each point represents a sample, coloured and shaped by timepoint. The plot illustrates a significant shift in community structure over time (PERMANOVA, $R^2 = 0.761$, $p = 0.001$).

5.2.2.2 Overview of the Biofilm Metagenome

To visualise the changes of the most abundant taxa, the relative abundances of the top 15 species from the metagenomic data were plotted over the 14-day experiment (Appendix 5.1). The planktonic inoculum was dominated by early colonisers like *Streptococcus sp. oral*. Upon biofilm formation at Day 0, there was a clear increase of *Actinomyces sp. oral* and *Prevotella sp. oral*. Notably, the relative abundance of known periodontal pathogens such as *Parvimonas micra* and *Fusobacterium vincentii* increased from D0, becoming established members of the community by Day 14 (Costalonga and Herzberg, 2014). This visual representation clearly shows a transition from a community dominated by facultative anaerobes to more obligate anaerobes known for their proteolytic functions.

5.2.2.3 Longitudinal Analysis Reveals Key Taxonomic Shifts Over Time

To identify the key species driving the progression of dysbiosis, a Likelihood Ratio Test (LRT) was conducted on the metagenomic data to identify taxa with significant changes in abundance across the three key stages, the planktonic Inoculum, the "healthy" Day 0 biofilm, and the mature dysbiotic Day 14 biofilm (Fig 5.2). The expression patterns of all significantly changing species revealed a clear and structured succession of the microbial community. The analysis identified a distinct cluster of species that were highly abundant in the Inoculum but were significantly depleted upon biofilm formation at Day 0 and remained at low levels at Day 14. This group was dominated by species commonly associated with a healthy oral microbiome, including numerous *Streptococcus* species (*S. parasanguinis, S. mitis, S. oralis*), *Prevotella* species (*P. melaninogenica, P. jejuni*), and *Haemophilus parainfluenzae* (Deo and Deshmukh, 2019). This pattern indicates that these largely commensal, facultative anaerobic species thrive in the planktonic state but are outcompeted for survival in the developing anaerobic biofilm. Conversely, a second major cluster identified the key species that successfully colonised and proliferated through the dysbiotic shift. These taxa, including *Eubacterium nodatum, Peptostreptococcus anaerobius, Bulleidia extructa*, and *Parvimonas micra* and *Mogibacterium timidum*, showed low abundance in the initial inoculum but became highly abundant by Day 14. The latter two of these are known periodontal pathogens (Neilands *et al.*, 2019, Rams, Sautter and van Winkelhoff,

2020). This shift demonstrates a clear succession towards a community enriched in anaerobic species, which is a characteristic signature of a developing periodontitis-associated biofilm (Nemoto and Ohara Nemoto, 2021).

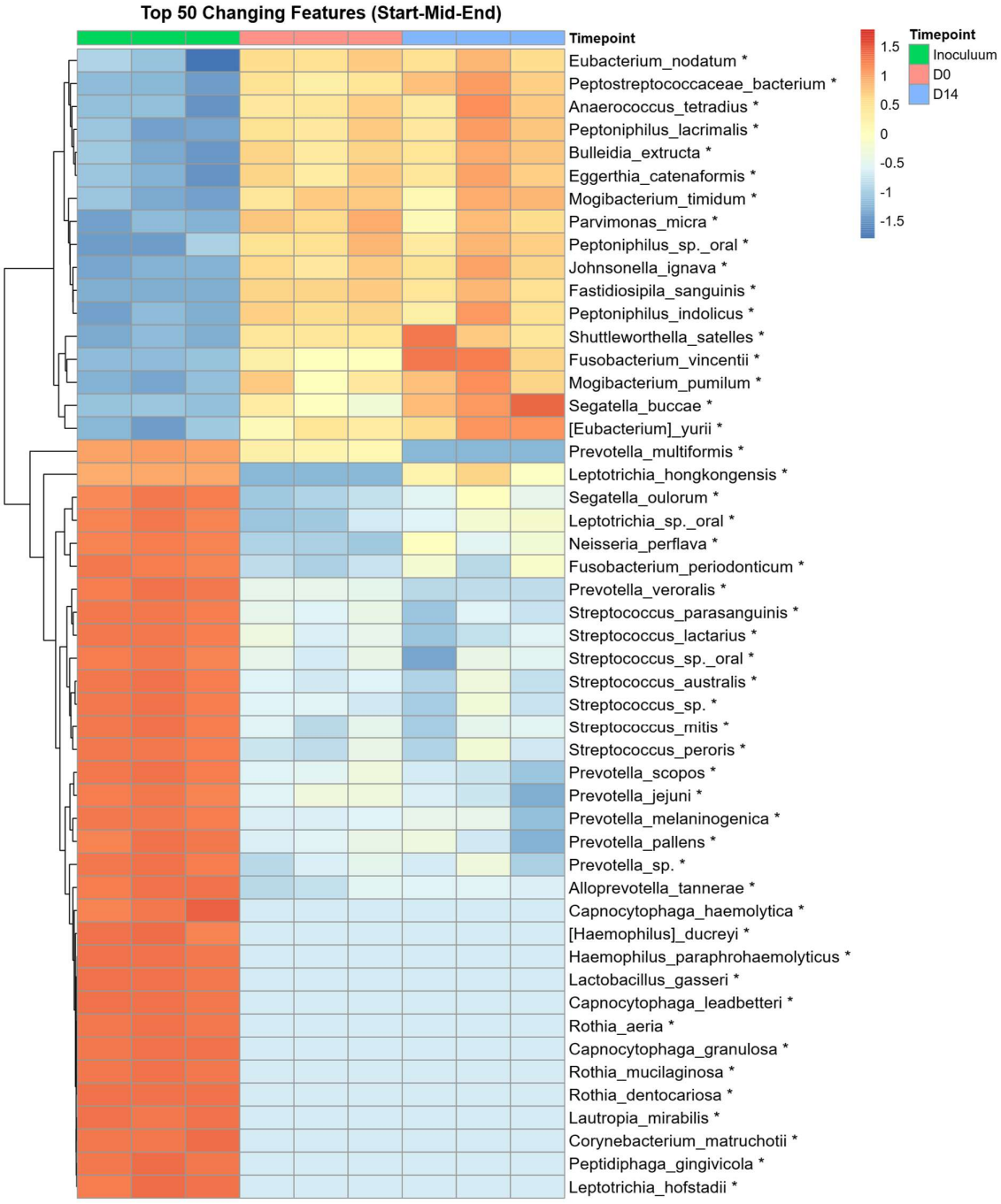


Figure 5.2. Heatmap of the significantly changing microbial species across key timepoints. The heatmap displays the scaled and centred relative abundance of the species identified by a Likelihood Ratio Test. Columns represent individual samples, and rows represent species, clustered by their expression patterns. Red indicates high abundance, and blue indicates low abundance relative to the mean. The timepoints shown are Inoculum (green), Day 0 (red), and Day 14 (blue).

5.2.2.4 Metagenomic Changes Throughout Biofilm Maturation

LRT analysis of the metagenome across the entire biofilm maturation process (Day 0 to Day 14) provided a more detailed view of the dynamics of dysbiosis (Fig 5.3). The heatmap of all significantly changing species shows a complex restructuring of the community with distinct niches. Some species were most active during the initial stages, for example, *Pseudoramibacter alactolyticus* and *Abiotrophia defectiva* peaked in abundance during the earlier time points before declining. Another group of anaerobes, including *Peptoniphilus sp. oral* and *Anaerococcus prevotii*, started at low abundance and became more dominant during the intermediate stages of maturation, before being depleted again in the final Day 14 community. Finally, a key cluster of late colonisers emerged as the dominant members of the mature dysbiotic community. These species, including the keystone pathogen *Porphyromonas gingivalis*, as well as *Segatella oulorum* and a *Peptostreptococcaceae bacterium*, were present at low abundance in the early stages but progressively enriched to reach peak abundance by Day 14. This detailed longitudinal view reveals a highly dynamic succession, with different bacterial groups thriving at different stages of the dysbiotic process.

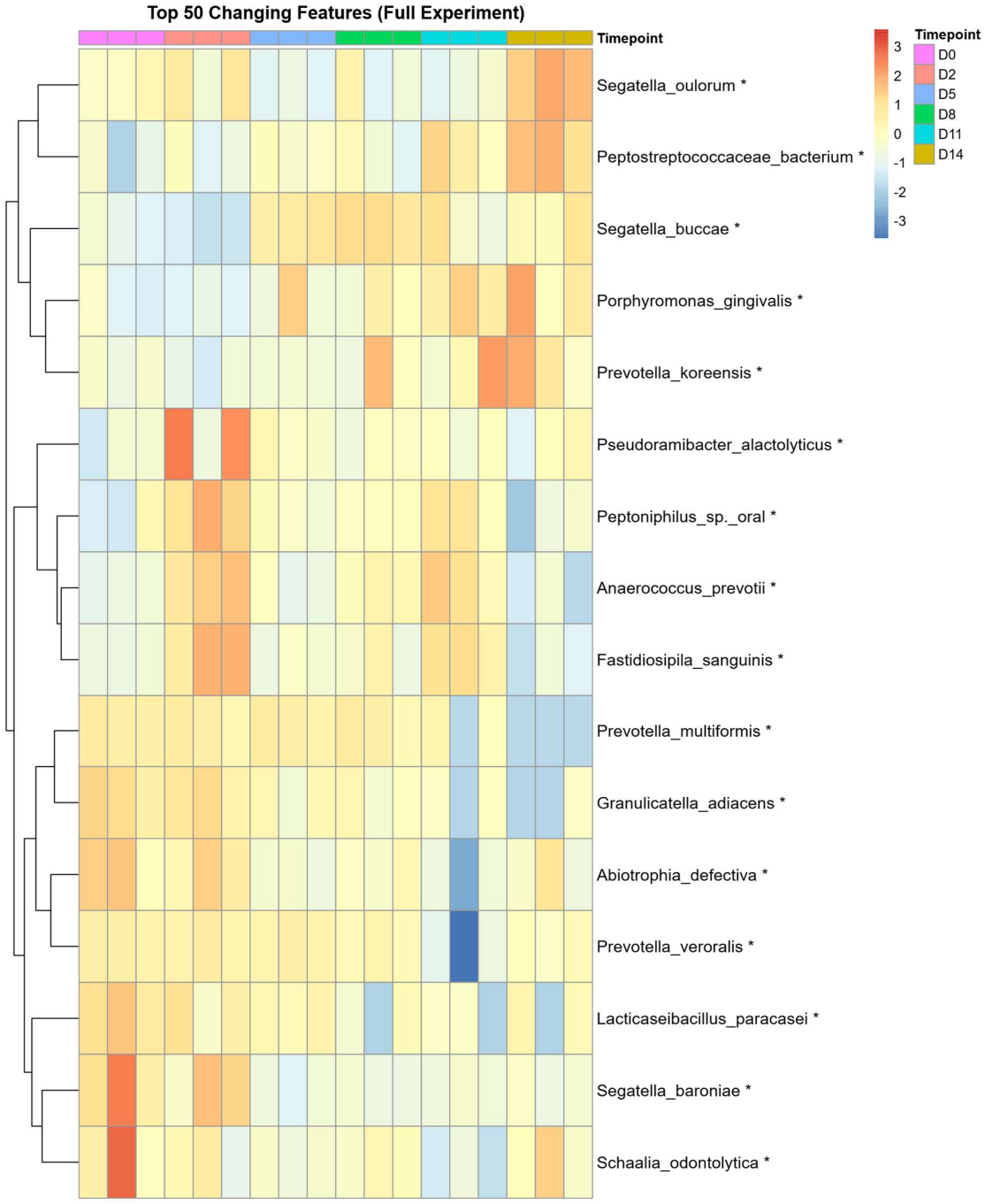


Figure 5.3. Longitudinal analysis of significantly changing microbial species during biofilm maturation. The heatmap displays the scaled and centred relative abundance of all species identified as significant by a Likelihood Ratio Test (LRT) from Day 0 to Day 14. Rows represent species, clustered by their abundance patterns, and columns represent samples, annotated by timepoint at the top. Yellow/Red indicates high abundance, and blue indicates low abundance relative to the mean for each species.

5.2.2.5 Differential Abundance of Species

To identify the specific taxa responsible for the initial colonisation of the biofilm, a pairwise differential abundance analysis was performed comparing the planktonic inoculum to the established Day 0 biofilm. This analysis revealed significant restructuring of the microbial community (Fig 5.4A). Numerous species highly abundant in the inoculum were significantly depleted upon biofilm formation. The most significantly depleted species were *Neisseria perflava*, followed by various *Streptococcus* species, *Prevotella melaninogenica*, and *Rothia mucilaginosa*. This indicates that these species, while prevalent in the initial pooled oral samples, were less successful at adhering and establishing themselves in the early biofilm environment. Conversely, a distinct set of species was significantly enriched at Day 0, indicating they were successful primary colonisers. These included known periodontal genera such as *Selenomonas infelix*, *Shuttleworthella satelles*, *Peptoniphilus indolicus*, *Catonella morbi*, and *Parvimonas sp. oral* (Rams, Sautter and van Winkelhoff, 2020, McDaniel *et al.*, 2021). The strong enrichment of these organisms demonstrates a clear selection process during the transition from a planktonic to a biofilm state, favouring species known to thrive in developing plaque communities.

5.2.2.6 Identifying the Mid-point Shift

The beta diversity analysis indicated a major structural shift in the community occurred between Day 2 and Day 8. To pinpoint the specific taxonomic changes that define this transitional stage, a pairwise differential abundance analysis was performed comparing the Day 8 biofilm to the Day 0 baseline (Fig 5.4B). This analysis revealed a significant restructuring of the community, driven by both the loss of early colonisers and the emergence of species associated with a more mature, dysbiotic state. A key finding was significant enrichment of several *Neisseria* species, including *N. meningitidis*, *N. lactamica*, and *N. gonorrhoeae*. There was also a significant increase in several *Bifidobacterium* species, such as *B. dentium* and *B. breve*. Conversely, a broad range of species that were abundant in the early biofilm were significantly depleted by Day 8. This included *Segatella baroniae*, *Abiotrophia defectiva*, and *Granulicatella adiacens*. This transition, marked by the significant loss of early colonisers and the increase of *Neisseria* and other species, defines

Day 8 onwards as the point where the biofilm community undergoes a fundamental shift towards its final dysbiotic structure.

5.2.2.7 Pairwise Comparison of Metagenome Reveals Dysbiotic Signature

A direct pairwise comparison of the metagenome between the Day 14 dysbiotic biofilm and the initial Day 0 healthy community further elucidated the specific taxonomic shifts underlying the development of a mature dysbiotic state. This analysis revealed a significant shift in the abundance of key species, many of which are associated with periodontal disease (Fig 5.4C). A notable finding was the significant enrichment of several species from the genus *Porphyromonas*, a hallmark of periodontal dysbiosis, including a significant increase in *Porphyromonas gingivalis*. This was complemented by the significant enrichment of other anaerobic taxa linked to periodontitis, such as *Cryptobacterium curtum* and *Leptotrichia hongkongensis* (Nakazawa *et al.*, 1999, Eribe and Olsen, 2008). Interestingly, after being depleted during initial biofilm formation, several *Neisseria* species, including *Neisseria sp. oral* and other pathogenic members of the genus, were significantly enriched at Day 14 compared to Day 0. This suggests a secondary colonisation or adaptation by these organisms in the mature, dysbiotic environment. In contrast, several species associated with the early biofilm were significantly depleted by Day 14. The most significantly depleted species were *Prevotella multiformis*, followed by *Segatella baroniae* and *Granulicatella adiacens*. This pattern of taxonomic shift, with the displacement of early colonisers and enrichment of proteolytic, anaerobic species, strongly indicates that the *in vitro* model successfully captured a shift towards a periodontitis-like microbial community at the DNA level.

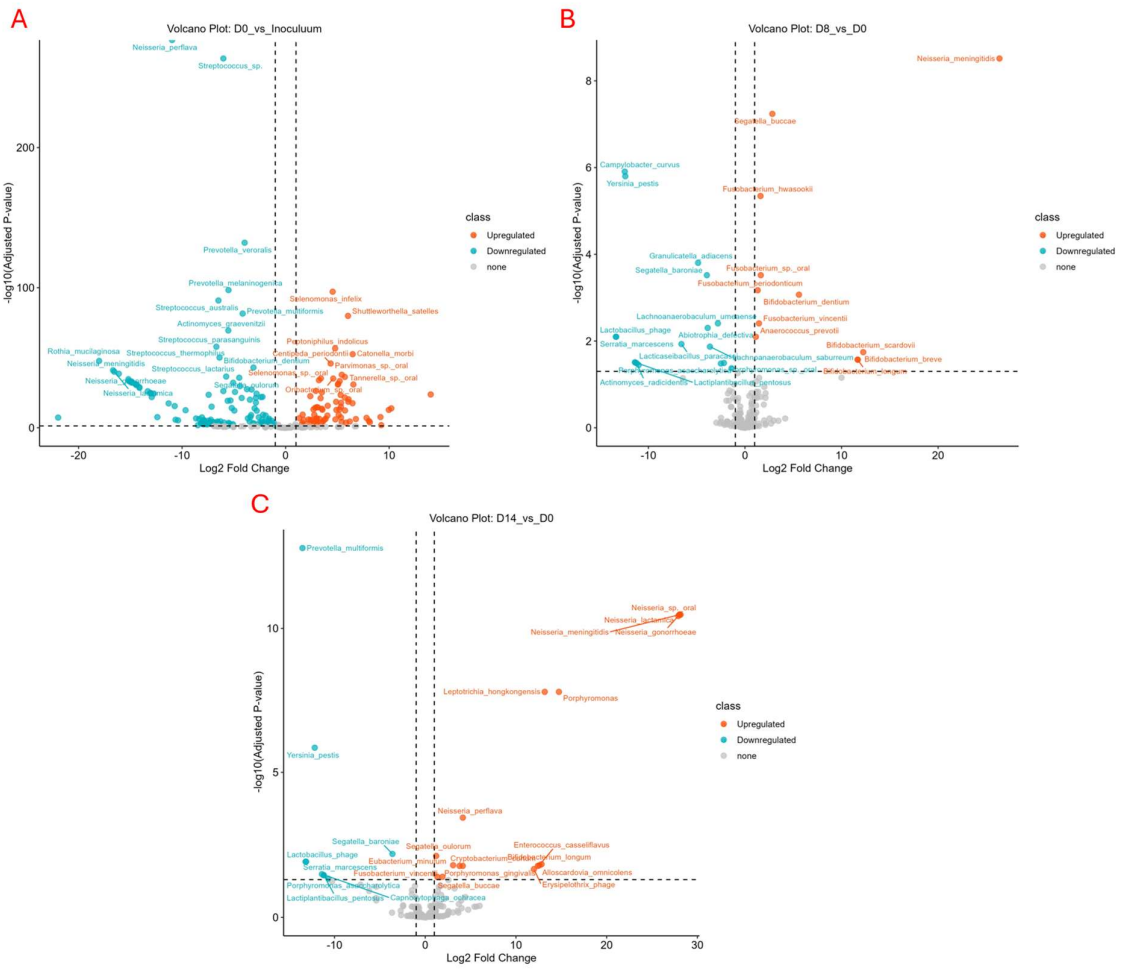


Figure 5.4. Pairwise differential abundance of microbial species at key transitional stages of biofilm development. Volcano plots show Log_2 Fold Change versus -log_{10} adjusted p-value. Red points represent species significantly upregulated (enriched), and teal points represent species significantly downregulated (depleted) in the first condition of each comparison. **(A)** Day 0 biofilm vs. planktonic Inoculum. **(B)** Day 8 biofilm vs. Day 0 biofilm. **(C)** Day 14 biofilm vs. Day 0 biofilm.

5.2.2.8 Functional Metagenomics: Gene-level Insights (DNA)

5.2.2.8.1 Functional Potential of the Biofilm Metagenome

To understand the functional potential within the community's collective genome, the gene pool diversity was analysed from metagenomic data. The alpha diversity of the functional metagenome showed a dynamic pattern over the 14-day period (Fig 5.5). The number of observed genes was highest in the initial inoculum, decreased upon establishment of the Day 0 biofilm, and remained relatively stable thereafter. A more distinct trend was evident in the diversity metrics: the Shannon, Simpson, and Inverse Simpson indices all showed an initial peak in functional diversity at Day 0, followed by a notable drop at Day 2 and Day 5, before gradually recovering in the later stages of the experiment. However, as with the taxonomic diversity, these changes in functional alpha diversity did not reach statistical significance ($p > 0.05$ for all comparisons).

While the overall diversity of functions remained relatively stable, the composition of the functional metagenome changed significantly over time (Fig 5.5). The PCoA of Bray-Curtis dissimilarities revealed a clear and progressive shift in the profile of encoded genes. The inoculum's functional profile was distinct, and subsequent time points showed clear separation, particularly between the early (D0-D5) and late (D8-D14) stages of biofilm development. This structural shift in the gene pool was highly significant, with the PERMANOVA analysis confirming that time point accounted for approximately 62% of the variance in the functional metagenome. This finding indicates a substantial programming of the community's metabolic and functional genetic priorities as it adapted to the dysbiotic conditions over the 14-day period.

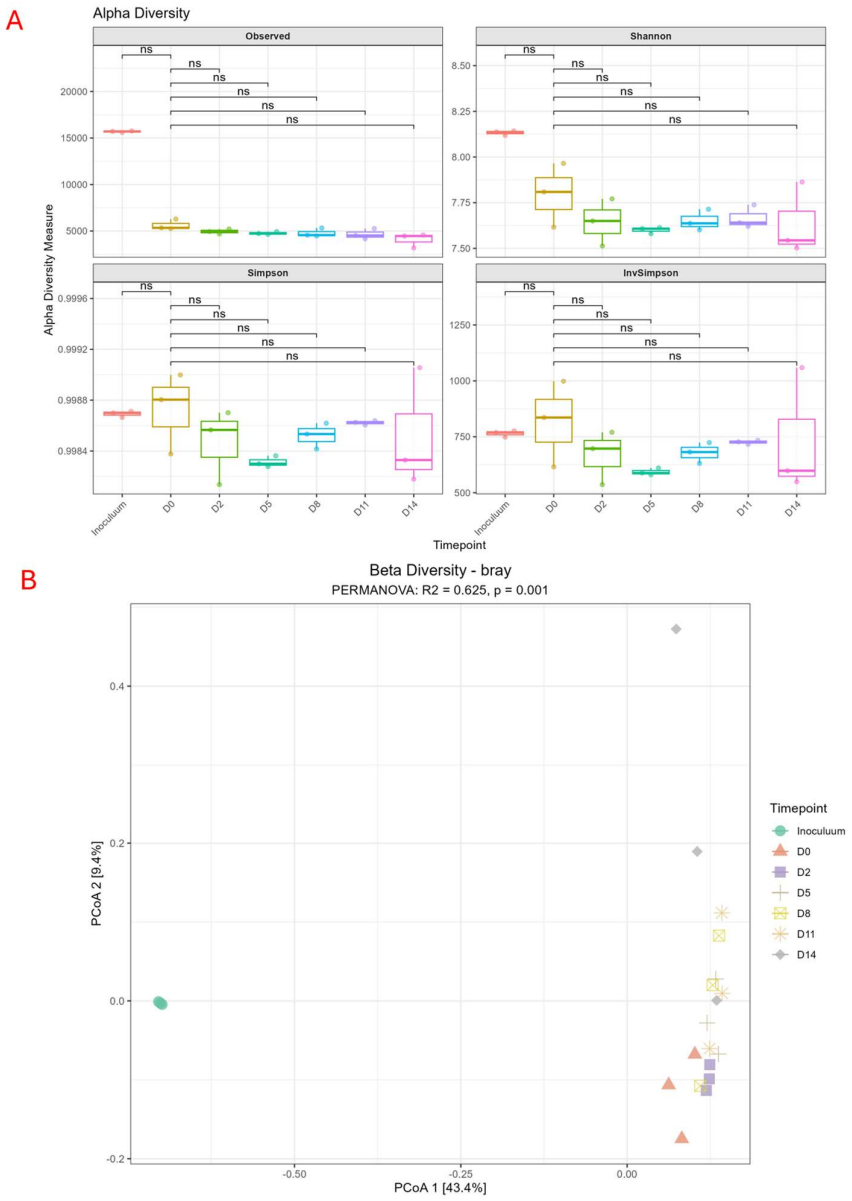


Figure 5.5. Functional diversity of the metagenome over a 14-day in vitro dysbiosis model. (A) Alpha diversity metrics (Observed, Shannon, Simpson, and Inverse Simpson) for the gene pool over time. Boxplots show the distribution of values for each timepoint. Brackets labelled 'ns' indicate no significant difference in pairwise comparisons to Day 0 ($p > 0.05$). **(B)** Beta diversity of the functional metagenome, shown as a Principal Coordinate Analysis (PCoA) plot of Bray-Curtis dissimilarities. Each point represents a sample, coloured and shaped by timepoint. The plot illustrates a significant shift in the gene pool over time (PERMANOVA, $R^2 = 0.625$, $p = 0.001$).

5.2.2.8.2 Differential Abundance of Genes Reveals a Shift Towards Pathogenic Potential

To identify the specific genetic markers of the dysbiotic shift, a differential abundance analysis was performed on the metagenomic gene data. LRT was used to identify genes that showed significant changes in abundance across the key stages of biofilm development: the initial inoculum, the established 'healthy' biofilm at Day 0, and the late-stage dysbiotic biofilm at Day 14. This analysis revealed that a substantial portion of the microbial metagenome was significantly altered, confirming a major genetic reprogramming of the community.

This reprogramming was defined by two opposing trends. Firstly, there was a significant enrichment of genes associated with virulence and genetic mobility in the mature dysbiotic state (Table 5.1). Among the most enriched genes were those encoding a conjugal transfer protein *(TraE)* and a Type IV secretion system protein *(VirD4)*, both of which are critical for HGT and host-pathogen interactions (Grohmann et al., 2018). The abundance of a collagen-binding domain protein gene was also significantly increased, suggesting an enhanced genetic capacity for the biofilm to interact with and potentially degrade the collagen-rich periodontal ligament (Lallier et al., 2001).

Conversely, the analysis revealed a depletion of genes essential for a healthy, proliferative community. Genes central to core metabolism (S-adenosylmethionine synthetase), protein synthesis (30S ribosomal protein S10), transcription (RNA polymerase sigma factor RpoD), DNA replication (DNA polymerase III subunit delta), and carbohydrate metabolism (PTS system) were all significantly less abundant in the mature dysbiotic biofilm. Collectively, the data show a clear functional trade-off, where the genetic potential for core metabolic processes is reduced in favour of genes that facilitate virulence, host tissue interaction, and genetic exchange.

Table 5.1- LRT significant genes of DNA metagenomes

Gene	log2 Fold Change	Adjusted (padj)	p-value
conjugal transfer protein TraE	14.85	0.0290	
type IV secretion system protein VirD4	14.28	0.0290	
collagen binding domain protein	11.75	0.0290	
serine kinase	4.55	0.0290	
S-adenosylmethionine synthetase	-14.08	0.0290	
30S ribosomal protein S10	-13.37	0.0290	
RNA polymerase sigma factor RpoD	-11.68	0.0290	
DNA polymerase III subunit delta	-11.15	0.0290	
PTS system, cellobiose-specific IIC component	-9.41	0.0290	

5.2.2.8.3 Pairwise Comparison Highlights Key Genetic Changes Between Healthy and Dysbiotic States

A direct pairwise comparison between the final dysbiotic state (Day 14) and the initial healthy state (Day 0) provided a clearer picture of the dysbiotic shift's functional endpoint. This analysis identified numerous genes that were significantly up- or down-regulated, highlighting a metabolic shift towards pathogenicity and stress response (Fig 5.6). Among the most significantly enriched genes were those associated with core metabolic functions and transport, such as aspartate-ammonia ligase and serine kinase, alongside an ATP-binding protein of an ABC transporter system. In contrast, the analysis revealed a depletion of genes central to the metabolism of the healthy biofilm. Key depleted genes included the PTS system, cellobiose-specific IIC component, a primary transporter for carbohydrate in commensal species, as well as S-adenosylmethionine synthetase and homoserine O-acetyltransferase, which are involved in essential amino acid biosynthesis. Furthermore, a significant depletion was observed for the virulence-associated protein E. Finally, a depletion of the Transposase and IS5 family protein indicates a potential change in the abundance of mobile genetic elements within the mature dysbiotic biofilm.

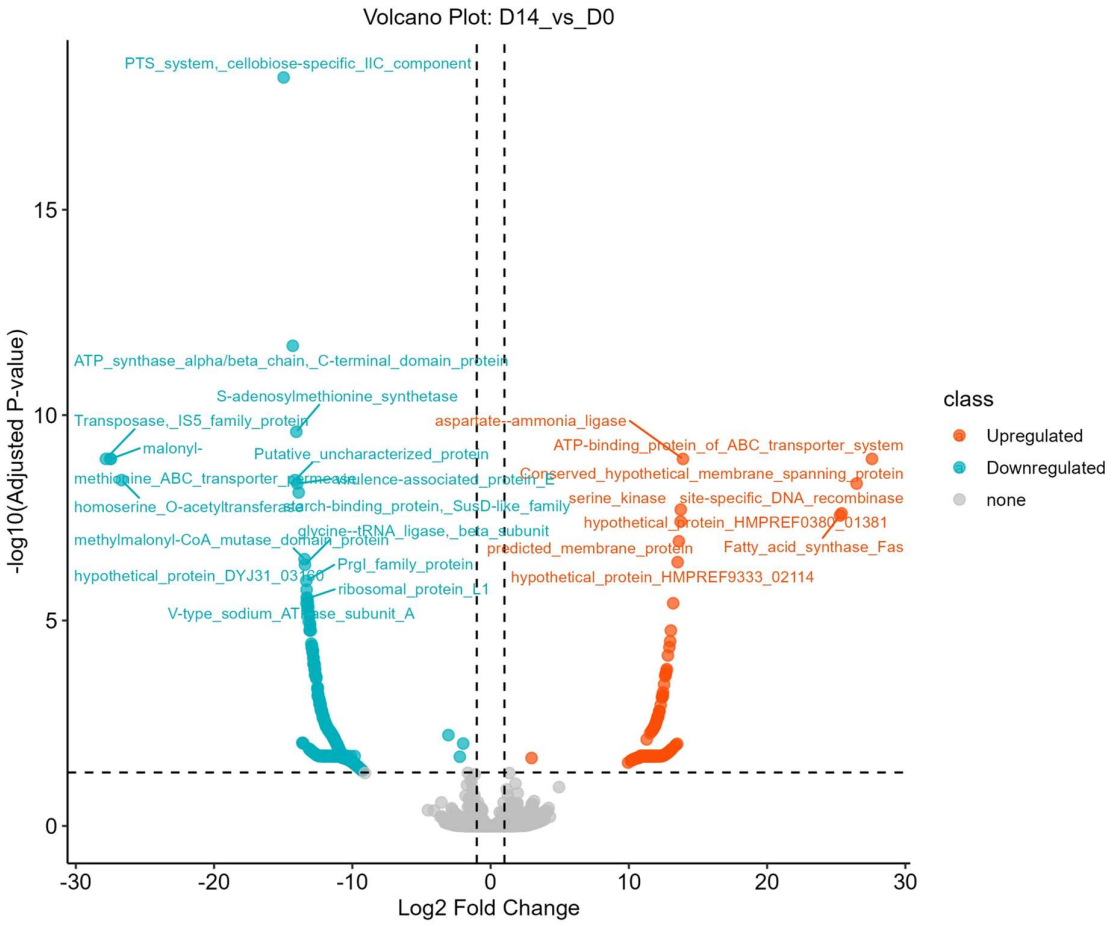


Figure 5.6. Pairwise differential abundance of genes between the mature dysbiotic biofilm (Day 14) and the initial healthy biofilm (Day 0). The volcano plot shows Log_2 Fold Change versus $-log_{10}$ adjusted p-value for individual genes. Red points represent genes significantly upregulated (enriched) at Day 14, and teal points represent genes significantly downregulated (depleted) at Day 14 compared to Day 0.

5.2.3 Metatranscriptomic Analysis: Community Activity (RNA)

5.2.3.1 Active Community Dynamics: Alpha and Beta Diversity

To complement the metagenomic data and provide insight into the functionally active members of the community, a parallel analysis of the metatranscriptome was conducted. The alpha diversity of the active microbial community, as revealed by RNA data, showed different trends than that of the total community DNA. While species richness (Observed) remained relatively stable, the diversity indices reflecting evenness (Shannon, Simpson, and Inverse Simpson) displayed a more pronounced V-shape, with a distinct drop in diversity at Day 2 followed by a gradual recovery and increase in the later time points (Fig 5.7). Despite these observable trends, the pairwise comparisons between time points did not reach statistical significance ($p > 0.05$).

The overall structure of the active community, however, mirrored the significant shift observed in the DNA analysis. The PCoA of Bray-Curtis dissimilarities based on the RNA profiles showed a clear and significant development (Fig 5.7). The active community composition shifted progressively over the 14-day period, with the later time points (Day 11, and 14) forming cluster moving away from the early time points (Day 0, 2). This shift was highly significant, with the PERMANOVA analysis demonstrating that time point explained over 80% of the variance in the active community structure. This confirms that the dysbiotic shift was not only a change in species presence but also a significant alteration in the transcriptionally active members of the microbial community.

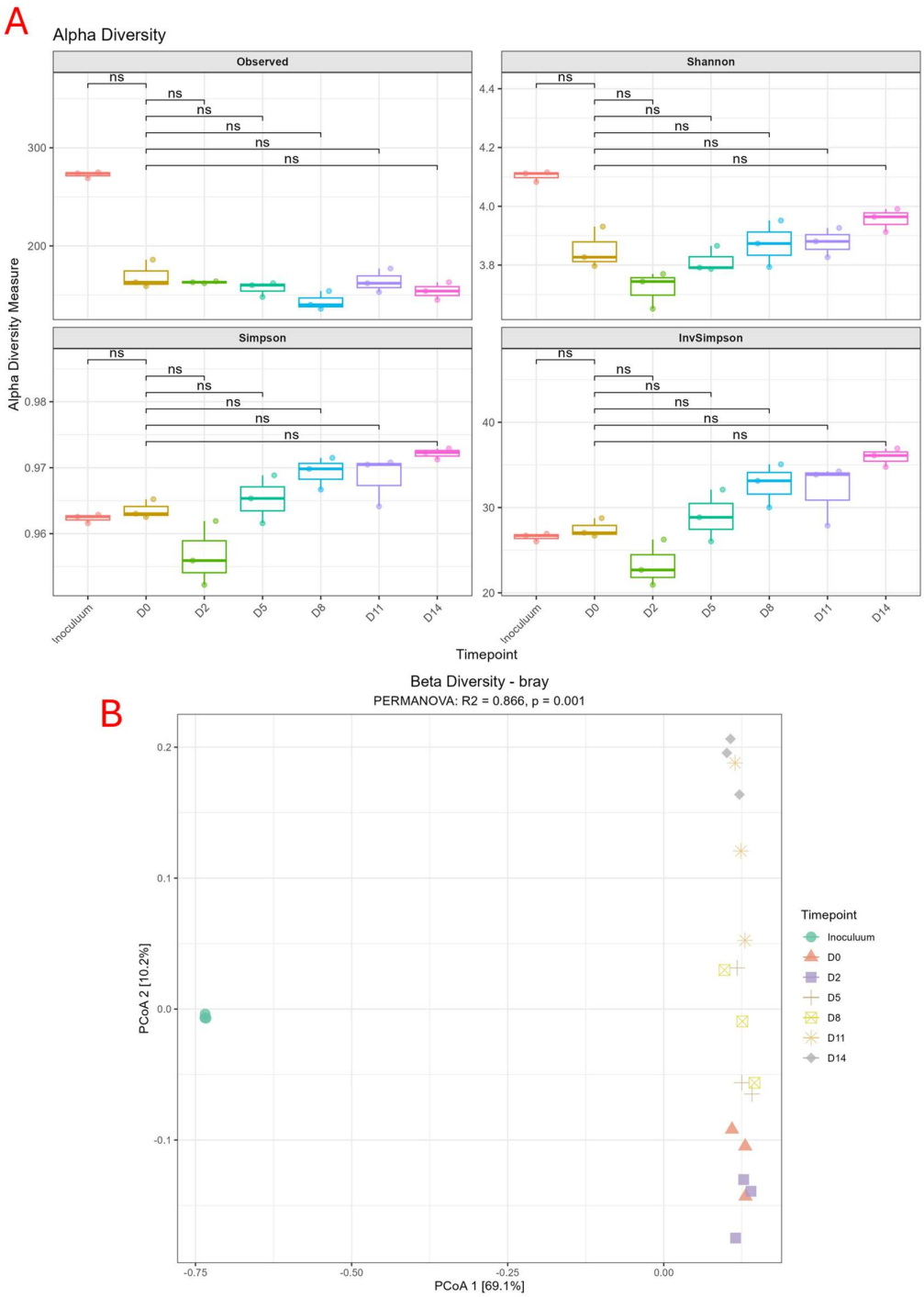


Figure 5.7. Diversity of the active microbial community (metatranscriptome) over a 14-day in vitro dysbiosis model. (A) Alpha diversity metrics (Observed, Shannon, Simpson, and Inverse Simpson) for the active community over time. Boxplots show the distribution of values for each timepoint. Brackets labelled 'ns' indicate no significant difference in pairwise comparisons to Day 0 ($p > 0.05$). **(B)** Beta diversity of the metatranscriptome, shown as a Principal Coordinate Analysis (PCoA) plot of Bray-Curtis dissimilarities. Each point represents a sample, coloured and shaped by timepoint. The plot illustrates a significant shift in the active community over time (PERMANOVA, $R^2 = 0.866$, $p = 0.001$).

5.2.3.2 Compositional Overview of the Active Biofilm

To provide a high-level overview of the active community, the relative abundance of the top 15 most transcriptionally active species and gene functions was visualised over time (Appendix 5.2). At the species level, the composition of the active community shifted. The inoculum was dominated by *Streptococcus sp. oral*. Upon biofilm formation at Day 0, there was a significant increase in the activity of *Streptococcus anginosus* and *Prevotella sp. oral*, which together accounted for a substantial portion of the community's transcriptional output. As the biofilm matured into a dysbiotic state, these early active members were gradually replaced by late-stage colonisers. Notably, by Day 14, *Tannerella forsythia* and *Treponema maltophilum* showed a marked increase in their relative transcriptional activity, aligning with their role as key players in periodontal dysbiosis (Haffajee and Socransky, 2005).

At the functional level, the metatranscriptome was dominated by core housekeeping genes throughout the experiment. Functions such as translation elongation factor Tu, the chaperone protein DnaK, and the DNA-directed RNA polymerase beta subunit consistently accounted for a major fraction of the transcriptional output. However, a clear shift was visible after the initial establishment phase. The relative abundance of transcripts for the ABC transporter ATP-binding protein increased after inoculation and remained among the most highly expressed gene functions for the remainder of the experiment, suggesting that a sustained high level of transport activity is a key feature of the biofilm state. This compositional analysis demonstrates that, while core metabolic functions remain active, the specific species driving this activity and the relative priority of other functions shift significantly during dysbiosis development.

5.2.3.3 Active Species Driving the Dysbiotic Shift: D14 vs D0

A pairwise comparison of the metatranscriptome between Day 14 and Day 0 identified the most transcriptionally active species that defined the mature dysbiotic state (Fig 5.8). The analysis revealed a clear shift from a community dominated by early colonisers to one driven by the activity of late-stage, pathogenic species. The most significantly upregulated active species at Day 14 was *Prevotella denticola*, with a log2 fold change of over 10.5x. This was followed by a

significant increase in the activity of *Capnocytophaga leadbetteri*. In contrast, several species associated with oral health and early colonisation were among the most significantly downregulated. *Schaalia odontolytica* and *Schaalia meyeri* showed the most decrease in transcriptional activity, followed by *Treponema medium*. The significant decrease in the activity of *Segatella baroniae* further confirmed the displacement of the early community. Interestingly, the keystone pathogen *Porphyromonas gingivalis* showed a significant increase in activity, confirming that its increased abundance at the DNA level translated to a significant functional role in the late-stage biofilm. This transcriptional data provides a clear picture of the functional shifts, highlighting the specific species that were not only present but actively driving the metabolic and pathogenic activities of the mature dysbiotic community.

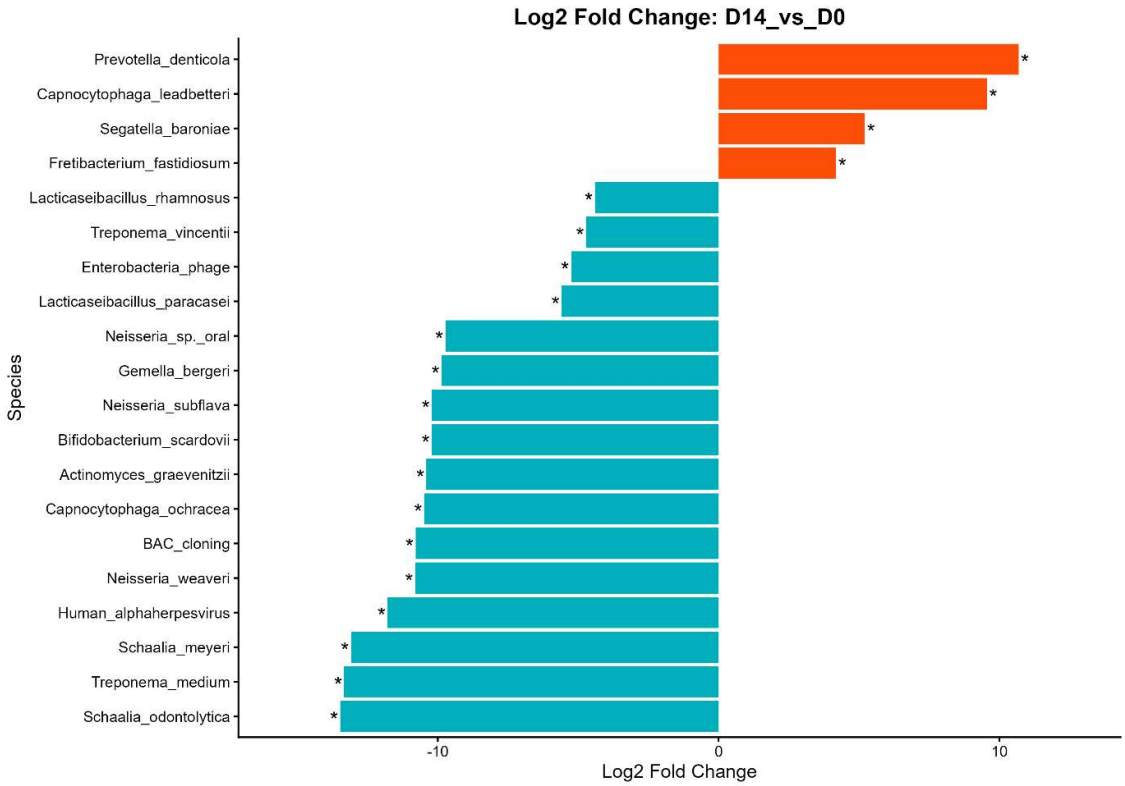


Figure 5.8. Differential transcriptional activity of microbial species between Day 14 and Day 0. The bar chart displays the Log_2 Fold Change in transcriptional activity for significantly altered species. Red bars indicate species with higher activity at Day 14 (upregulated), and teal bars indicate species with lower activity at Day 14 (downregulated) compared to Day 0. Asterisks denote statistical significance.

5.2.3.5 Longitudinal Transcriptional Shifts Across the Full Experiment

An LRT analysis across the full experimental time course (Day 0 to Day 14) provided a more granular view of the transcriptional evolution of the dysbiotic biofilm. The heatmap of all significantly changing genes reveals several distinct expression patterns (Fig 5.9). A cluster of genes, including the T3/T7-like RNA polymerase, showed high expression during the initial healthy state (Day 0 and Day 2), followed by a progressive decline as the biofilm transitioned to dysbiosis. This pattern suggests a shutdown of core transcriptional activities associated with the early commensal community.

In contrast, a distinct late-stage cluster was strongly upregulated at Day 11 and Day 14. This group included genes for key pathogenic functions, such as the starch-binding protein (*SusD-like family*) and glycoside hydrolase, as well as genes indicating genetic exchange like conjugal transfer protein TraE. This shows an activation of both nutrient acquisition and HGT machinery in the mature dysbiotic biofilm.

Other genes displayed more complex, transient expression patterns. For example, an RNA-binding transcriptional accessory protein showed peak expression during the early-to-mid stages of the experiment, followed by a drop-off. Core replication machinery such as ATP-dependent DNA helicase maintained expression throughout development before showing a significant depletion specifically in the late-stage Day 14 community. These fluctuating patterns suggest a highly dynamic and regulated transcriptional response as the community adapts during the development of dysbiosis.

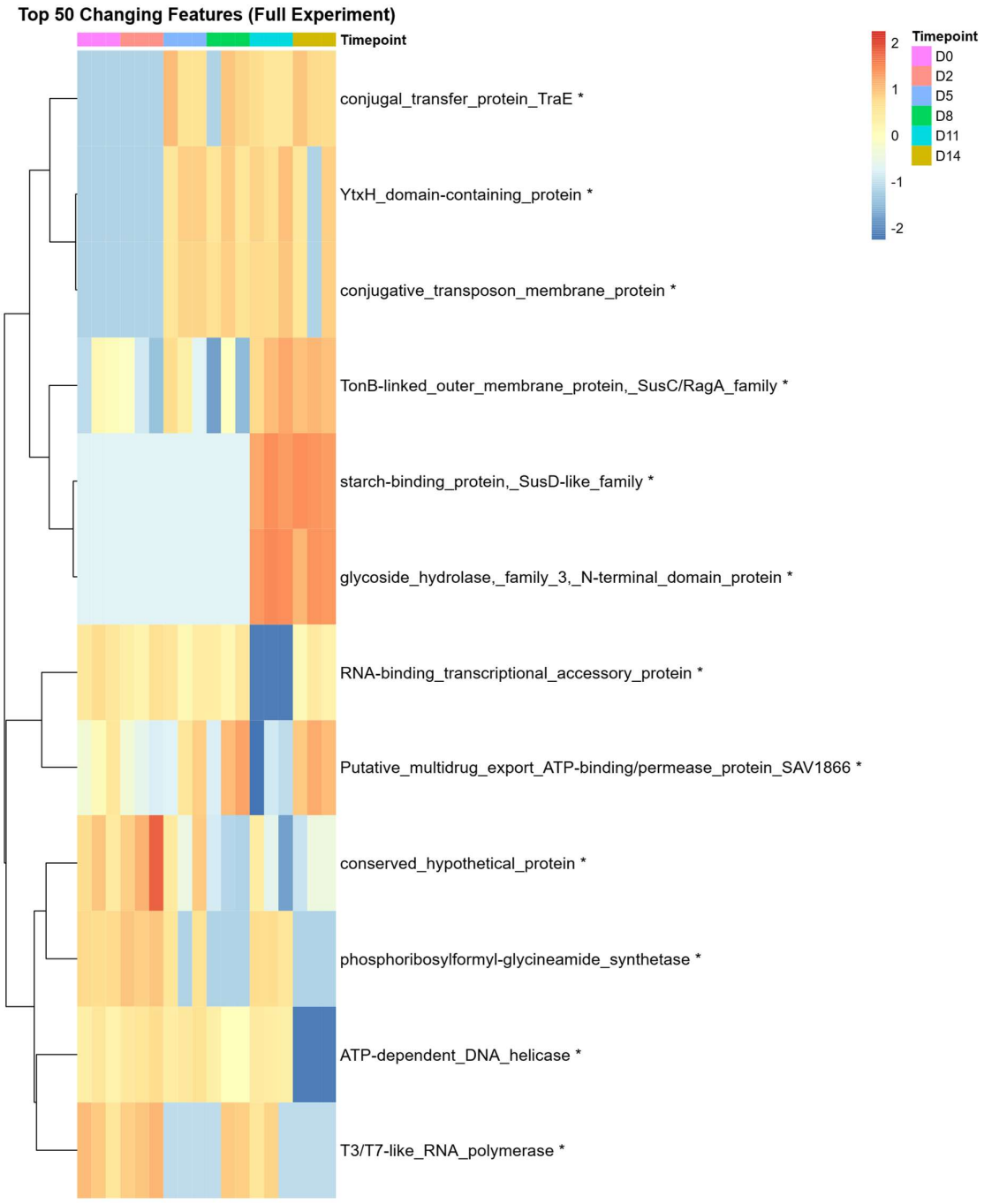


Figure 5.9. Heatmap of significantly changing gene transcripts over the full biofilm maturation process. The heatmap displays the scaled and centred relative abundance of all genes identified as significant by a Likelihood Ratio Test (LRT) from Day 0 to Day 14. Rows represent genes, clustered by their expression patterns, and columns represent samples, annotated by timepoint at the top. Yellow/Red indicates high expression, and blue indicates low expression relative to the mean for each gene.

5.2.3.6 Gene Expression Dynamics Start, Middle, End

To understand the transcriptional profiles that define the transition from a healthy to a dysbiotic state, an LRT analysis was performed on the RNA gene expression data, comparing the initial inoculum, the established Day 0 biofilm, and the final Day 14 dysbiotic community. The expression patterns of all significantly changing genes, as determined by the LRT, were visualised in a heatmap (Fig 5.10). This analysis revealed distinct clusters of co-regulated genes, highlighting a clear shift in the biofilm's functional activity.

A large cluster of genes, including those for 3-oxoacyl-ACP synthase and numerous ribosomal proteins (e.g., 30S ribosomal protein S14, 50S ribosomal protein L14), showed high expression in the initial inoculum but was strongly repressed upon biofilm formation at Day 0 and Day 14. This suggests these functions are critical for the planktonic phase but are not maintained in the biofilm.

Conversely, another major cluster demonstrated high transcriptional activity during the early stages of biofilm formation. Genes in this group, which included functions like 1-acyl-sn-glycerol-3-phosphate acyltransferase, were highly expressed in both the inoculum and the Day 0 biofilm. Still, their expression significantly decreased by Day 14. This indicates an initial burst of metabolic and protein synthesis activity that subsides as the biofilm matures into a dysbiotic state.

Finally, a smaller cluster of genes, including 4-hydroxybutyrate dehydrogenase, exhibited a pattern characteristic of late-stage dysbiosis, showing low expression in the early biofilm stages before becoming significantly upregulated at Day 14. This analysis of gene expression provides clear evidence of a multi-stage transcriptional program that drives the development and maintenance of the dysbiotic state *in vitro*.

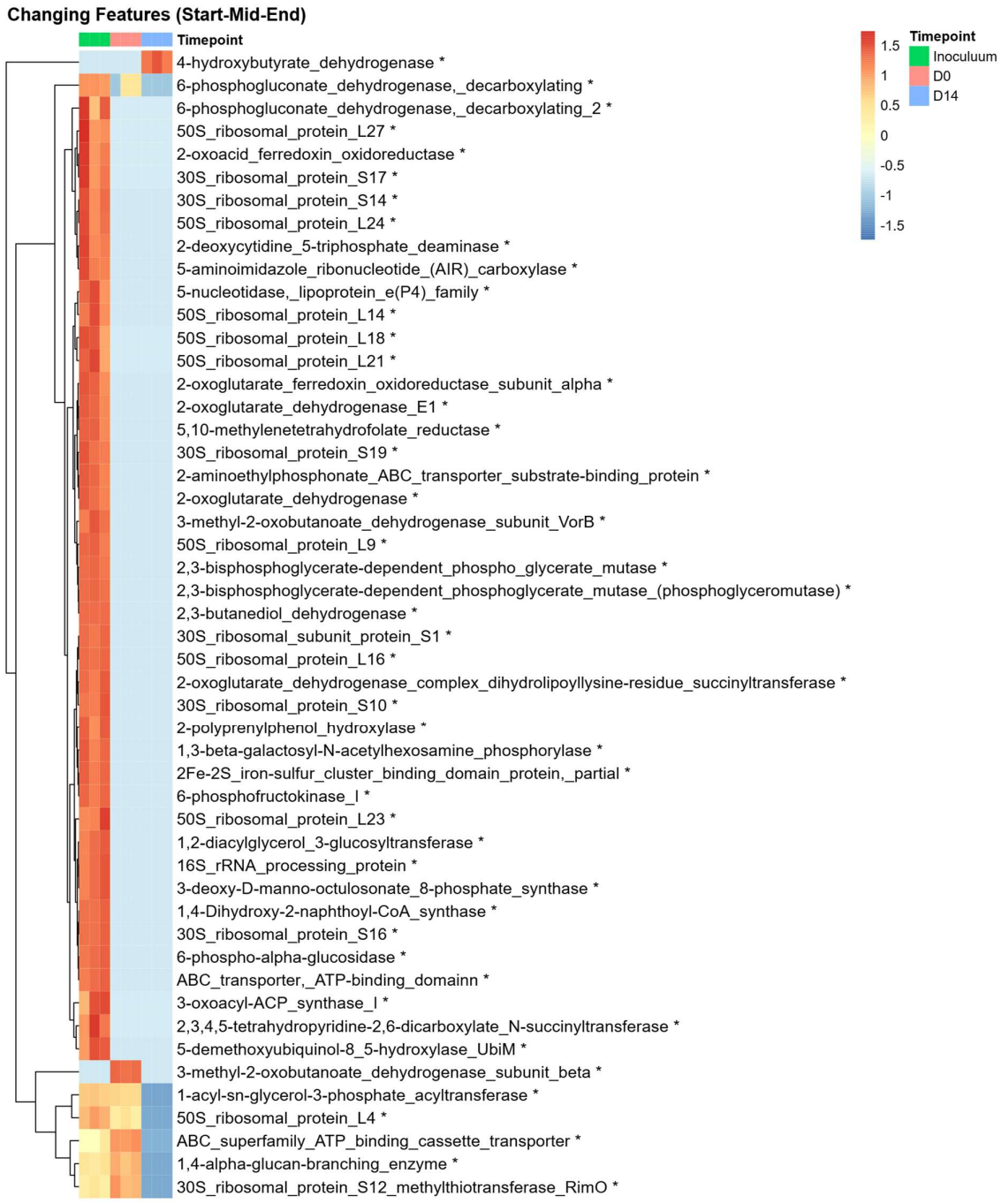


Figure 5.10. Heatmap of significantly changing gene transcripts across key stages of biofilm development. The heatmap displays the scaled and centred relative abundance of all genes identified as statistically significant by a Likelihood Ratio Test (LRT). Columns represent individual samples, and rows represent genes, clustered by their expression patterns. Red/yellow indicates high expression, and blue indicates low expression relative to the mean. The timepoints shown are Inoculum (green), Day 0 (red), and Day 14 (blue).

5.2.3.7 Pairwise Comparison of the Transcriptome

A direct pairwise comparison of the metatranscriptome between the Day 14 dysbiotic biofilm and the Day 0 healthy biofilm highlighted the key functional activities driving the disease state. This analysis revealed a strong transcriptional signature of an active, pathogenic community (Fig 5.11A). The most significantly upregulated gene at Day 14 was a CfxA family class A broad-spectrum beta-lactamase, with a log2 fold change greater than 15, indicating an increase in the expression of a key antibiotic resistance mechanism. This aligns with the metagenomic findings from the clinical periodontitis cohort (Chapter 3) and suggests the *in vitro* dysbiotic state is changed for resistance. Furthermore, several genes associated with virulence and colonisation were highly upregulated, showing a community geared towards nutrient acquisition and host interaction. Transcripts for a starch-binding protein, SusD-like family, were the second most upregulated feature, suggesting a strong focus on carbohydrate metabolism, a key function for pathogens like *Porphyromonas* in the nutrient-limited subgingival environment (Moradali and Davey, 2021). This was complemented by the high expression of genes involved in protein degradation and adhesion, such as PP-loop family protein and a Cna protein B-type domain protein, which facilitates collagen binding. The upregulation of these functions is shifted towards a more proteolytic and adhesive phenotype, characteristic of pathogenic biofilms. In contrast, the analysis identified a significant downregulation of core machinery required for replication and transcription. Key downregulated transcripts included RNA polymerase, beta subunit, ATP-dependent DNA helicase, and various other genes involved in basic cellular processes like iron-sulfur cluster assembly (iron-sulfur cluster-binding protein). This suggests a potential metabolic shift in the mature dysbiotic state, where the community conserves energy by downregulating housekeeping functions like rapid proliferation and instead towards expressing specific virulence, resistance, and nutrient acquisition functions. This transcriptional profile provides evidence that the *in vitro* model successfully recapitulated key functional activities characteristic of a periodontitis-associated dysbiotic biofilm.

5.2.3.4 Identifying the Transcriptional Shift at Day 8: D8 vs D0

To pinpoint the transcriptional events that occurred during the major structural shift at Day 8, a pairwise comparison of gene expression was performed between Day 8 and Day 0. This analysis revealed a significant shift in the functional activity of the biofilm (Fig 5.11B). The most striking change was the massive upregulation of genes associated with mobile genetic elements and stress responses. Genes such as the Transposon Tn1000 resolvase and the resolvase domain protein were among the most significantly upregulated, suggesting increased genetic mobility. This was accompanied by a strong induction of chaperone proteins, including dnaK protein and DnaK chaperone protein, indicating a community-wide stress response (Tomoyasu et al., 2012). Conversely, several core metabolic and structural genes were significantly downregulated by Day 8. Notably, transcripts for several ribosomal proteins, such as 50S ribosomal protein L15 and 30S ribosomal protein S7, were strongly depleted. Genes involved in fundamental processes, including IclR family transcriptional regulator and a conserved hypothetical integral membrane protein, were also among the most significantly downregulated. This transcriptional shift away from core biosynthetic processes towards stress response and genetic mobility signifies a reprogramming of the biofilm as it enters the late dysbiotic stage.

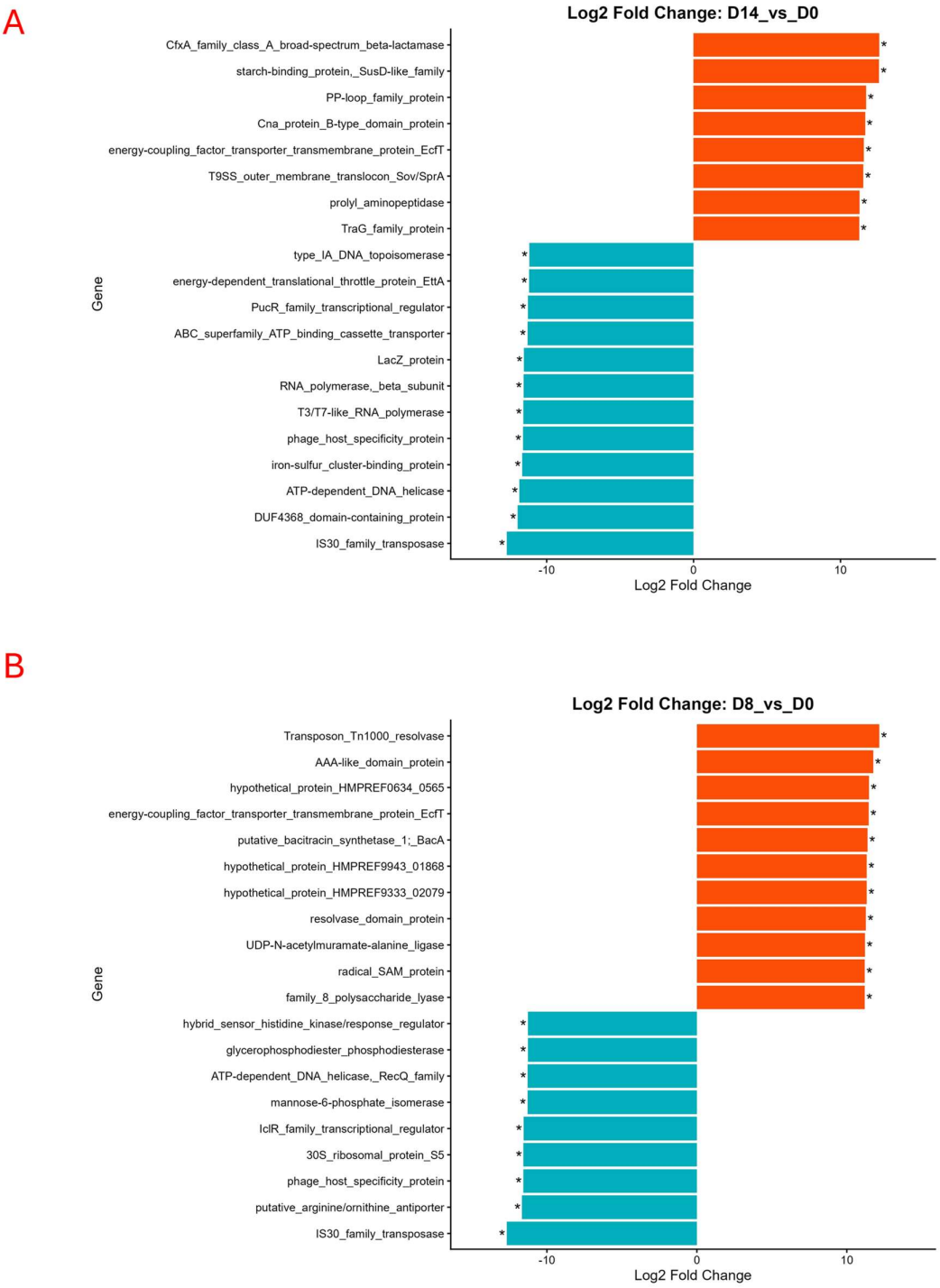


Figure 5.11. Differential transcriptional activity of genes at key stages of biofilm maturation. Bar charts display the Log_2 Fold Change in transcriptional activity for significantly altered genes. Red bars indicate genes with higher activity (upregulated), and teal bars indicate genes with lower activity (downregulated) in the first condition of each comparison. Asterisks denote statistical significance. **(A)** Day 14 biofilm vs. Day 0 biofilm. **(B)** Day 8 biofilm vs. Day 0 biofilm.

5.2.4 Resistome Analysis

5.2.4.1 Metagenomic Resistome (DNA)

To assess the genetic potential for antibiotic resistance within the developing biofilm, the metagenomic data were aligned to the Comprehensive Antibiotic Resistance Database (CARD) (Section 2.4).

5.2.4.2 Longitudinal Changes in ARG Abundance

An LRT analysis of the metagenomic data identified the ARGs with the most significant changes in abundance across the key time points of the experiment (Inoculum, Day 0, and Day 14). The heatmap of all significantly changing ARGs revealed distinct patterns of depletion in the mature biofilm. One large cluster of ARGs, including the macrolide resistance gene *ErmX*, was highly abundant in the initial inoculum but was significantly depleted upon biofilm formation at Day 0 and remained at low levels at Day 14. Conversely, other genes, such as the vancomycin resistance regulator *vanR*, were abundant in both the inoculum and the early Day 0 biofilm before showing a marked decrease in abundance in the late-stage dysbiotic biofilm at Day 14.

5.2.4.3 Pairwise Comparison of the Resistome

A pairwise comparison between the established Day 0 biofilm and the initial inoculum revealed a significant shift in the ARG profile. Genes conferring resistance to macrolides, such as *erm(45), ErmB*, and *ErmY*, were all significantly upregulated at Day 0. In contrast, several efflux pump components, including *mtrC* and *MexW*, were significantly downregulated (Fig 5.12A).

Comparing the final dysbiotic state at Day 14 to the healthy Day 0 biofilm, a different pattern emerged (Fig 5.12B). The most significantly upregulated genes at Day 14 were all components of multidrug efflux systems, including *mtrC, macA*, and *smeF*. Conversely, a number of genes, including vancomycin resistance regulators (*vanR in vanF cluster*) and the multidrug efflux component *OmpA*, were significantly downregulated. This indicates a shift from a diverse resistome in the early biofilm to one dominated by the genetic potential for multidrug efflux in the mature dysbiotic state.

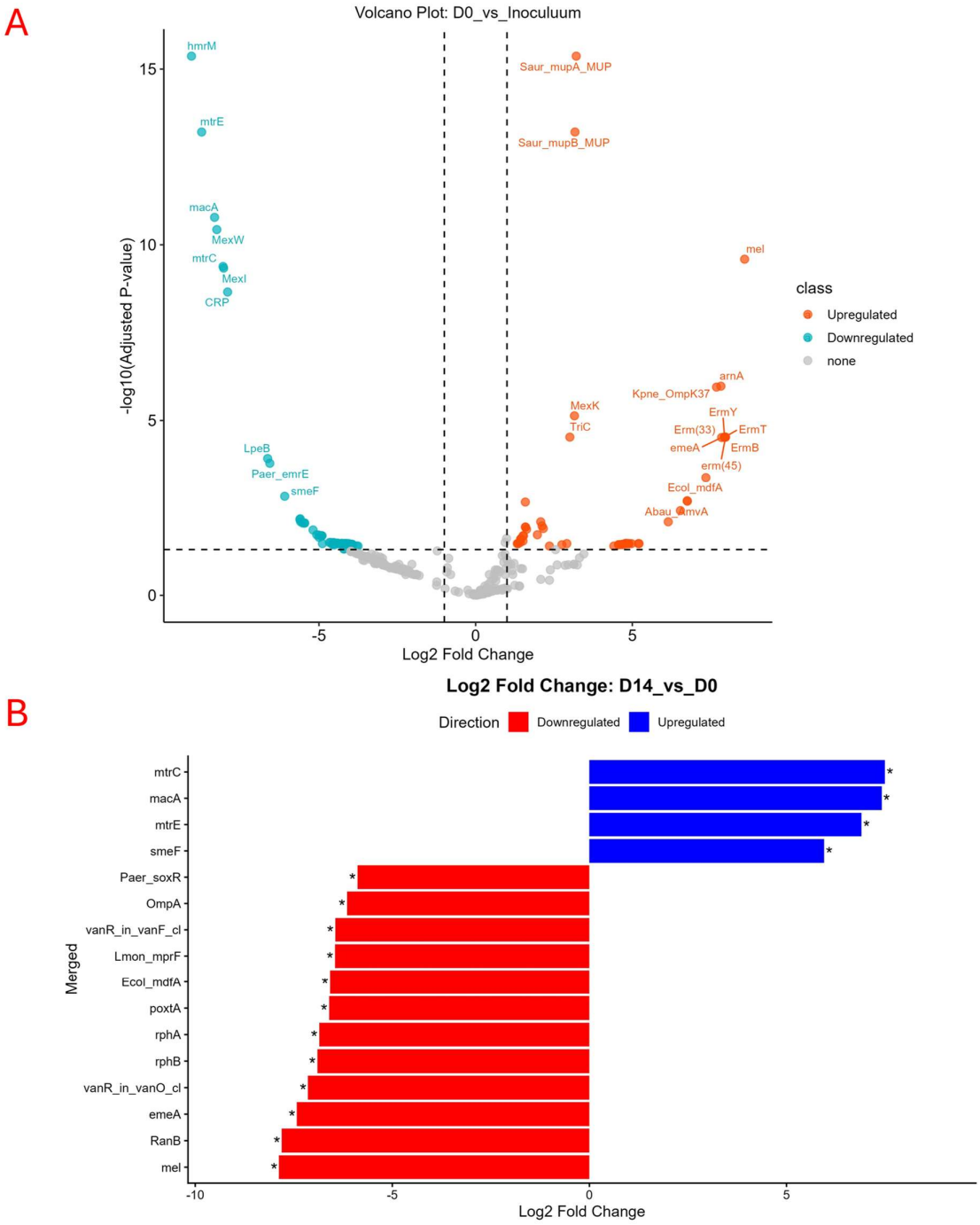


Figure 5.12. Differential abundance of Antibiotic Resistance Genes (ARGs) at key stages of biofilm development. (A) Volcano plot comparing the Day 0 biofilm to the planktonic Inoculum. Red points represent ARGs significantly enriched (upregulated) at Day 0, while teal points are depleted (downregulated). **(B)** Bar chart comparing the Day 14 biofilm to the Day 0 biofilm. Blue bars indicate ARGs enriched at Day 14, while red bars indicate ARGs depleted at Day 14. Asterisks denote statistical significance.

5.2.4.4 Evolution of Potential Resistance Mechanisms

To provide a higher-level functional narrative of the resistome's genetic potential, the abundance of ARGs was summarised by their mechanism of resistance. This analysis revealed a clear evolution in the community's potential defence strategies over the 14-day period. The LRT analysis across the full time course showed a dynamic pattern of potential resistance mechanisms (Fig 5.13). The genetic potential for antibiotic target alteration was high in the early biofilm (Day 0) before progressively declining through the later stages. In contrast, the potential for antibiotic target protection was low initially but became increasingly prominent during the mid-to-late stages of dysbiosis, reaching its peak abundance by Day 14. This indicates that the community's collective resistance is highly dynamic, with different potential becoming more prominent at different stages of the dysbiotic succession.

A more apparent change was evident when comparing the key transitional stages (Fig 5.13). The genetic potential of the established Day 0 biofilm was rich in genes related to antibiotic target alteration, which were at baseline levels in the initial planktonic inoculum. However, in the mature dysbiotic state at Day 14, the resistome's genetic potential refocused. The potential for target alteration was reduced, while there was a corresponding increase in the abundance of genes for antibiotic target protection. This result demonstrates that the dysbiotic shift involves a fundamental change in the community's encoded resistance strategy, moving from a genetic potential based on modifying antibiotic targets to one based on protecting them from antibiotic action.

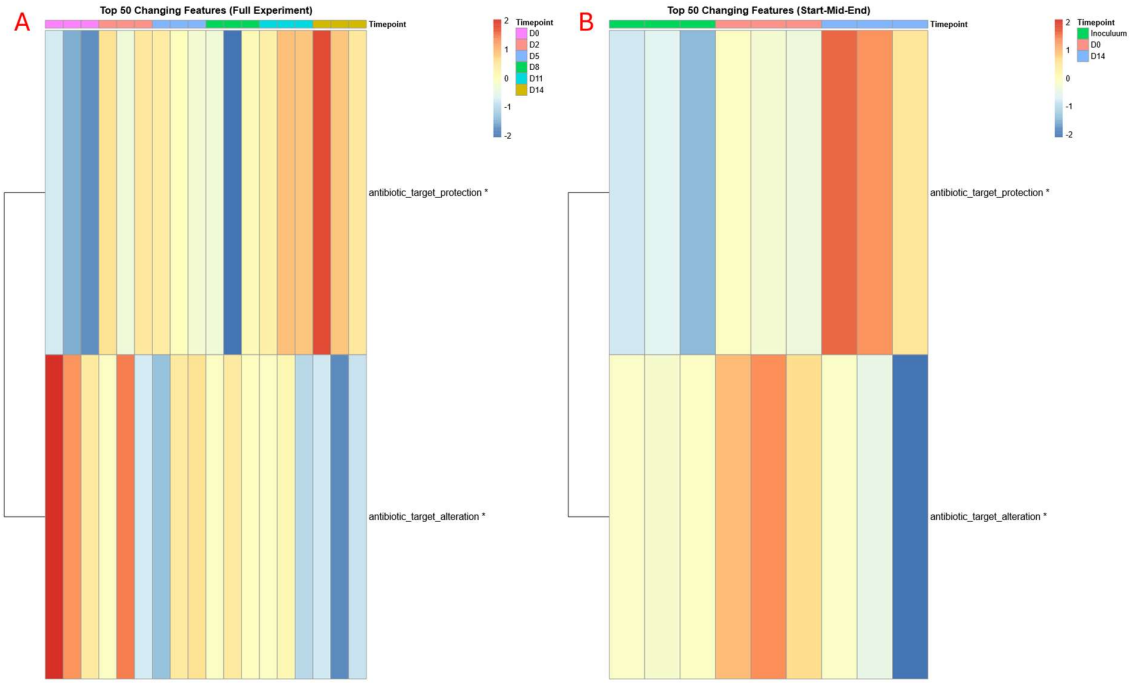


Figure 5.13. Changes in the abundance of ARG resistance mechanisms over time. Heatmaps display the scaled and centred relative abundance of ARG mechanisms identified as significant by a Likelihood Ratio Test (LRT). Rows represent the two main mechanisms, and columns represent individual samples. Red/yellow indicates high abundance, and blue indicates low abundance relative to the mean. **(A)** Analysis of the full experiment from Day 0 to Day 14. **(B)** Analysis of the key transitional stages: Inoculum, Day 0, and Day 14.

5.2.4.5 Transcriptional Dynamics of the Resistome (RNA)

To investigate the active expression of antibiotic resistance mechanisms during the development of dysbiosis, the metatranscriptomic data were aligned to the CARD.

5.2.4.6 Diversity of the Active Resistome (RNA)

The diversity of the active resistome was assessed using the metatranscriptomic data. The alpha diversity of the expressed ARGs showed a notable increase from the inoculum to the early biofilm stages (D0 and D2), followed by a decline in the mid-to-late stages of dysbiosis. While these trends were clear, as with previous alpha diversity analysis, none of the pairwise comparisons reached statistical significance.

The beta diversity of the active resistome, however, showed a highly significant shift over time. The PCoA plot illustrates a clear separation between the inoculum and the biofilm time points, as well as a distinct trajectory from the early-stage biofilms (D0, D2, D5) to the later, mature dysbiotic communities (D8, D11, D14) (PERMANOVA, $R^2 = 0.707$, $p = 0.001$) (Appendix 5.4). This indicates that the types of ARGs being actively expressed changed fundamentally as the biofilm developed.

5.2.4.7 Longitudinal Changes in ARG Expression

The expression patterns of all significantly changing ARG classes over the full experimental time course revealed a dynamic and structured response (5.14A). A notable cluster of ARGs, including those conferring resistance to diaminopyrimidine and sulphonamide antibiotics, showed peak expression in the early stages of biofilm formation (Day 0 and Day 2), followed by a decline in the later stages. This suggests an early activation of certain resistance pathways. Conversely, a large and distinct cluster of ARGs demonstrated a clear pattern of late-stage activation. Gene classes with resistance to tetracycline, glycopeptide, macrolide, and rifamycin antibiotics were all strongly upregulated at Day 11 and Day 14. This upregulation of multiple resistance mechanisms in the mature dysbiotic biofilm points towards the development of a multi-drug resistant phenotype. This evolution corresponded to a shift in the community's defence mechanisms, as detailed in section 5.3.2.4.

5.2.4.8 Key Stages of Resistome Activity

An LRT analysis focusing on the key time points of Inoculum, Day 0, and Day 14 provided further insight into the major transitions in the active resistome (Fig 5.14B). This analysis highlighted a shift in the types of resistance genes being expressed. The inoculum was characterised by high expression of genes conferring resistance to antibacterial free fatty acids and fluoroquinolones. However, as the biofilm matured, these were replaced by different ARGs. By Day 14, the community showed an upregulation of genes for resistance to peptide and lincosamide antibiotics, among others. This clear succession in the active resistome demonstrates a functional reprogramming of the community's defence mechanisms in response to the changing environmental conditions of the dysbiotic model. This succession was accompanied by changes in the dominant resistance mechanisms employed by the community.

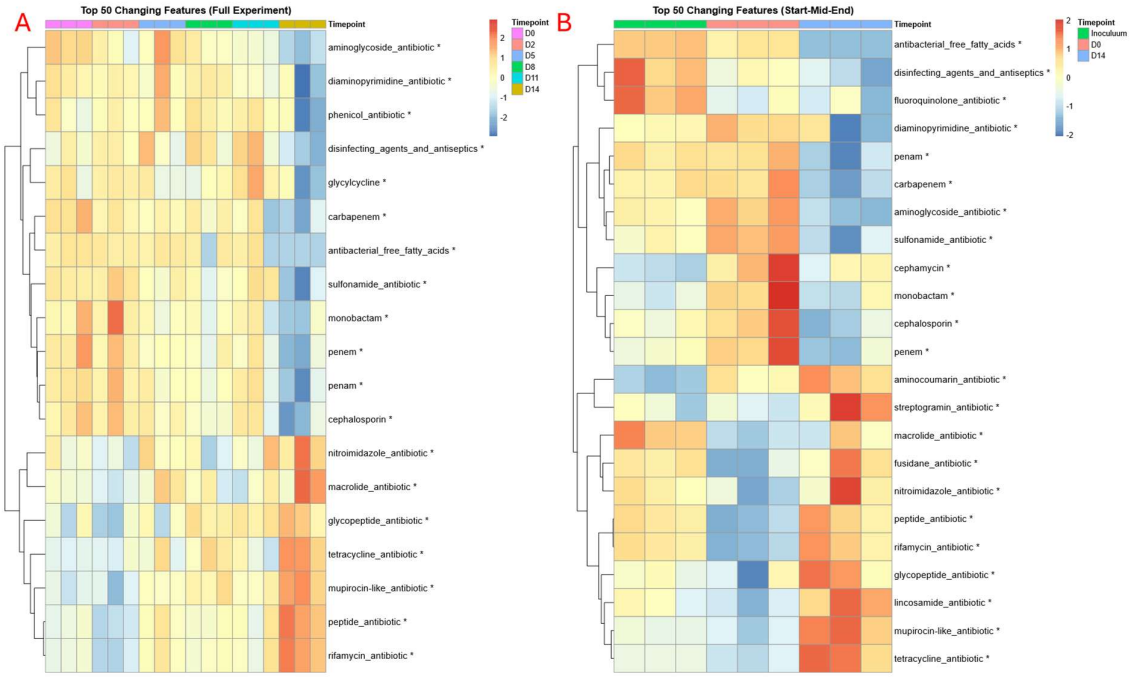


Figure 5.14. Changes in the transcriptional activity of ARG classes over time. Heatmaps display the scaled and centred relative abundance of ARG classes identified as significant by a Likelihood Ratio Test (LRT). Rows represent ARG classes, clustered by their expression patterns, and columns represent individual samples. Red/yellow indicates high expression, and blue indicates low expression relative to the mean. **(A)** Analysis of the full experiment from Day 0 to Day 14. **(B)** Analysis of the key transitional stages: Inoculum, Day 0, and Day 14.

5.2.4.9 Evolution of Active Resistance Mechanisms

To provide a higher-level functional narrative of the active resistome, the expression of ARGs was summarised by their mechanism of resistance. This analysis revealed a clear evolution in the community's primary defence strategies over the course of the dysbiotic shift. The LRT analysis across the full biofilm maturation process (Day 0 to Day 14) demonstrated a sequential activation of different resistance strategies (Fig 5.15A). In the early stages of biofilm formation (Day 0 and Day 2), the community's transcriptional activity was dominated by genes for antibiotic inactivation and antibiotic efflux. This suggests that the initial biofilm relies on strategies that either degrade antimicrobials or actively pump them out of the cell. As the biofilm matured, a distinct shift was observed, with a peak in the expression of genes associated with antibiotic target alteration in the mid-to-late stages (Day 8 and Day 11). In the final, late-stage dysbiotic community, the dominant active mechanism was antibiotic target replacement. This succession indicates a sophisticated, multi-stage reprogramming of the resistance strategy, moving from broad, generalised defences to more specific mechanisms of target modification within the mature pathogenic community.

The analysis of the key transitional stages (Inoculum, Day 0, Day 14) further clarified this (Fig 5.15B). The planktonic inoculum was characterised by a high expression of genes for antibiotic target alteration and antibiotic target replacement. Upon establishment of the biofilm at Day 0, the dominant active strategy shifted towards antibiotic efflux and antibiotic inactivation. However, by the final dysbiotic state at Day 14, these early mechanisms were replaced by the upregulation of genes for antibiotic target alteration and antibiotic target protection. This confirms that the development of the mature dysbiotic state is defined not just by a change in specific ARGs, but by a fundamental shift in the primary functional strategies the community uses to resist antimicrobial pressures.

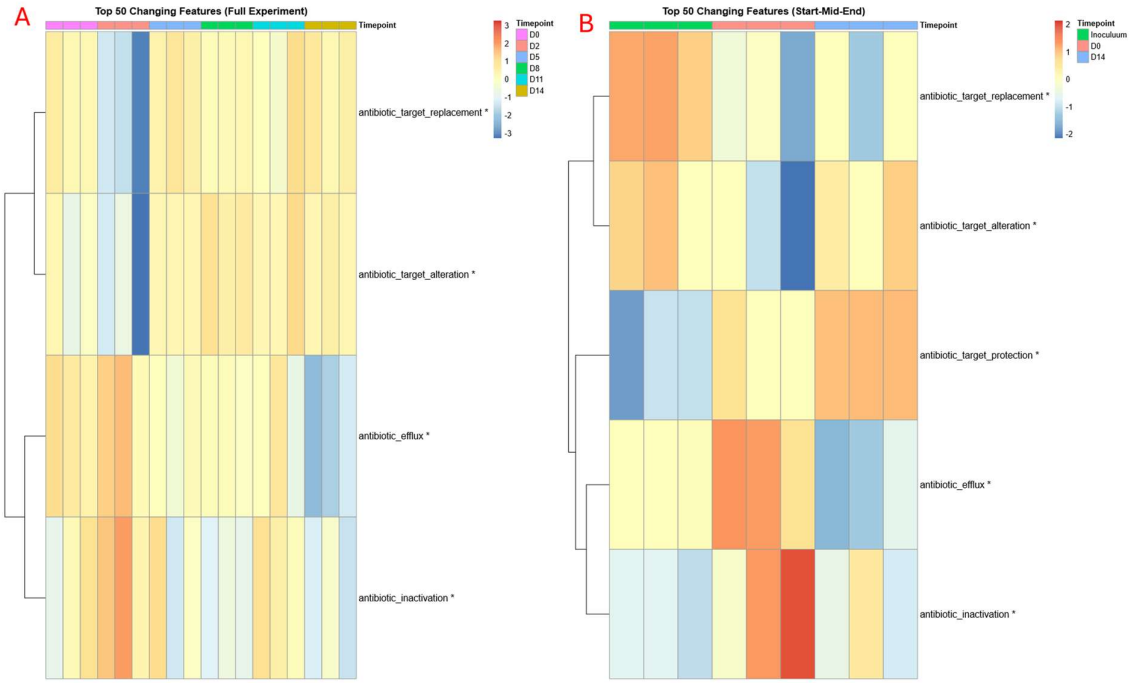


Figure 5.15. Changes in the transcriptional activity of ARG resistance mechanisms over time. Heatmaps display the scaled and centred relative abundance of ARG mechanisms identified as significant by a Likelihood Ratio Test (LRT). Rows represent the resistance mechanisms, clustered by their expression patterns, and columns represent individual samples. Yellow indicates high expression, and blue indicates low expression relative to the mean. **(A)** Analysis of the full experiment from Day 0 to Day 14. **(B)** Analysis of the key transitional stages: Inoculum, Day 0, and Day 14.

5.2.4.10 Pairwise Comparison of the Active Resistome

To better understand the active resistome, pairwise comparisons were performed across key time points. Comparing the Day 0 biofilm to the initial Inoculum revealed a significant shift in the expression of ARGs upon biofilm formation (Fig 5.16A). Several efflux pump components, such as *mdtO* and *eptA*, were significantly upregulated, suggesting that an active efflux-based resistance is an early and important feature of the established biofilm. Conversely, genes conferring resistance to tetracyclines (*tetA(60)*) and other antibiotic classes were significantly downregulated, indicating a selective pressure that alters the resistome profile during the transition from a planktonic to a biofilm lifestyle.

A pairwise comparison between the final Day 14 dysbiotic state and the Day 0 healthy biofilm highlighted the transcriptional shifts of the resistome's evolution. This analysis revealed downregulation of numerous ARG transcripts at Day 14 (Fig 5.16B). Key downregulated genes included efflux pump components like *farB* and *mdtF*, as well as genes conferring resistance to trimethoprim (*dfrEre*) and tetracycline (*tcr3*). Only a few genes, such as *novA,* which is associated with novobiocin resistance, were significantly upregulated. This suggests that while certain resistance mechanisms are highly activated at specific stages (as shown by the LRT analysis), the overall transcriptional focus of the mature dysbiotic community may shift away from a broad-spectrum resistance profile towards more specific functions.

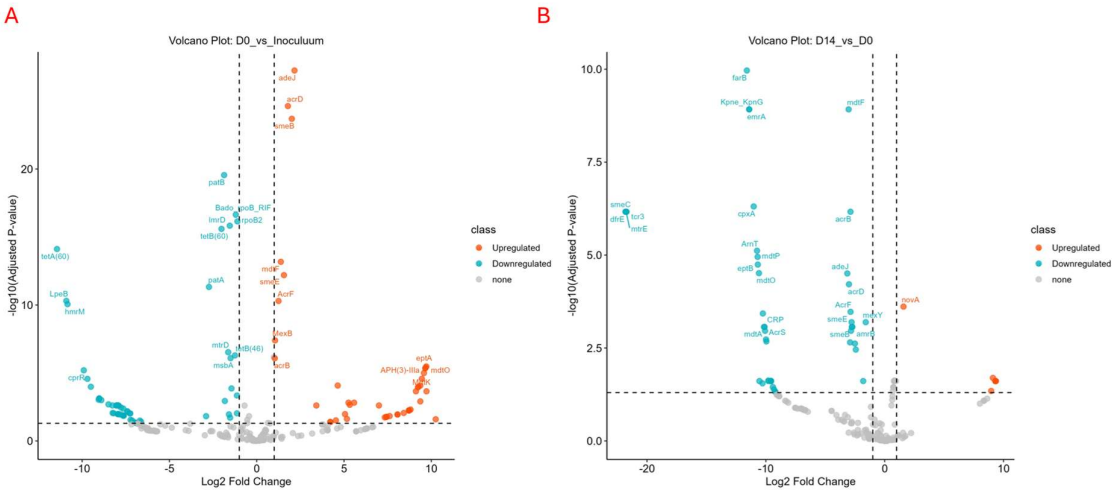


Figure 5.16. Pairwise differential transcriptional activity of Antibiotic Resistance Genes (ARGs) at key biofilm stages. Volcano plots show Log_2 Fold Change versus $-log_{10}$ adjusted p-value. Orange points represent ARGs with significantly higher expression (upregulated), and teal points represent ARGs with significantly lower expression (downregulated) in the first condition of each comparison. **(A)** Day 0 biofilm vs. planktonic Inoculum. **(B)** Day 14 biofilm vs. Day 0 biofilm.

5.2.4.11 The Resistome Structural Shift

The pairwise comparison of ARG expression between Day 8 and Day 0 reveals the transcriptional shift that accompanies the major restructuring of the microbial community (Fig 5.17). At this mid-experiment transitional stage, the active resistome underwent a significant transformation. The analysis highlighted a strong upregulation of genes conferring resistance to vancomycin and tetracycline. Several vancomycin resistance genes (*vanD, vanI, vanP*) were among the most significantly upregulated transcripts at Day 8. This was accompanied by the increased expression of multiple tetracycline resistance genes, including *tet(32)* and *tet(W)*. This suggests that as the community shifted towards a more pathogenic profile, the active expression of resistance mechanisms against clinically relevant antibiotics became a priority. In contrast, several efflux-related genes were significantly downregulated at Day 8 compared to Day 0. These included components of multidrug resistance efflux pumps such as *AcrS, APH(3)-IIIa*, and *mdtP*. This suggests a shift in the primary mode of antibiotic resistance, moving from a general efflux-based strategy in the early biofilm to the expression of more specific resistance determinants as the community matures towards a dysbiotic state.

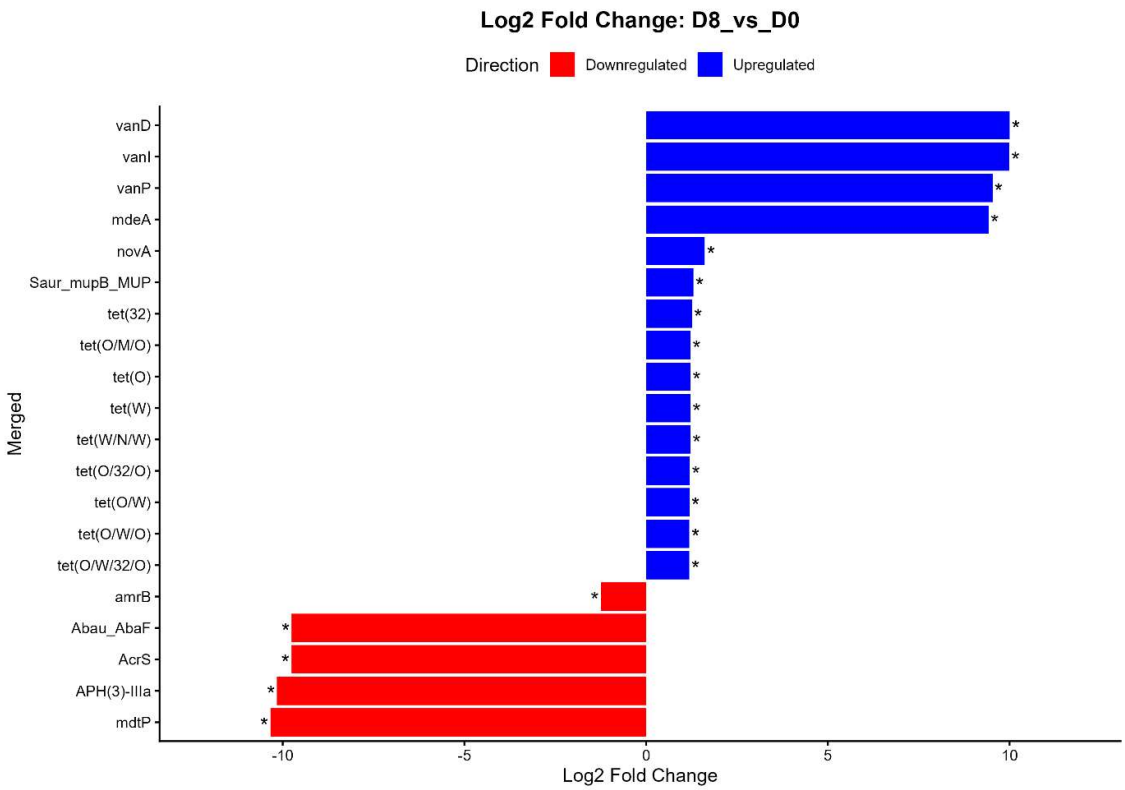


Figure 5.17. Differential transcriptional activity of ARGs between Day 8 and Day 0. The bar chart displays the Log_2 Fold Change in transcriptional activity for significantly altered ARGs. Blue bars indicate genes with higher activity at Day 8 (upregulated), and red bars indicate genes with lower activity at Day 8 (downregulated) compared to Day 0. Asterisks denote statistical significance.

5.2.5 Mobilome Analysis

To understand the dynamics of mobile genetic elements (MGEs) during the dysbiotic shift, the mobilome was characterised from both metagenomic (DNA) and metatranscriptomic (RNA) data, with plasmid populations (plasmidome) profiled using PLASMe (Section 2.5).

5.2.5.1 Plasmidome Dynamics (PLASMe)

5.2.5.2 Diversity of the Potential and Active Plasmidome

The diversity of the plasmidome, representing the total pool of plasmids, was assessed at both the DNA and RNA levels.

At the DNA level, mirroring the findings for the broader taxonomic and functional profiles, the alpha diversity of the plasmidome remained stable throughout the experiment. The richness (Observed) and evenness (Shannon, Simpson) of the plasmid pool showed some fluctuation, but no statistically significant changes were observed between any time points ($p > 0.05$). This suggests that the overall diversity of potential plasmids within the community was maintained even as the host bacterial composition changed. In contrast, the beta diversity of the potential plasmidome revealed a compositional shift over the 14-day period. The PCoA of Bray-Curtis dissimilarities showed a significant separation, with the early time points (D0-D5) distinct from the late-stage dysbiotic samples (D8-D14). This restructuring was highly significant, with PERMANOVA ($p = 0.001$, $R^2 = 0.764$) confirming that time point was a major driver of the variance (Appendix 5.5A).

A similar pattern was observed for the active plasmidome (RNA). The alpha diversity of expressed plasmids also showed no statistically significant changes across the 14-day time course, although a declining trend was visible in the later stages. However, the beta diversity of the active plasmidome showed a clear, significant compositional shift over time. The trajectory was highly significant (PERMANOVA:$p = 0.001$, $R^2 = 0.678$), confirming that the types of plasmids being actively expressed changed fundamentally as the biofilm developed (Appendix 5.5B).

Taken together, these results indicate that while the overall diversity of both potential and active plasmids remains relatively constant, the specific MGEs present and active within the community are replaced as the biofilm transitions towards a dysbiotic state.

5.2.5.3 Differential Abundance of the Potential Plasmidome

Analysis of the potential plasmidome at the DNA level revealed that the most significant changes occurred during the initial phase of biofilm establishment. A pairwise comparison between the established Day 0 biofilm and the planktonic inoculum showed highly considerable enrichment of numerous plasmids (Fig 5.20A). This enrichment was driven by plasmids from diverse bacterial genera, including multiple *Lacticaseibacillus* and *Klebsiella* species, as well as *Enterococcus faecalis*. In contrast, only a single plasmid, originating from the oral commensal *Streptococcus salivarius*, was significantly depleted upon biofilm formation.

The longitudinal LRT analysis across the key stages of development (Inoculum, Day 0, Day 14) provided a broader view of the plasmid succession (Fig 5.18B). This revealed distinct clusters of plasmids. The first cluster, which included plasmids from various *Streptococcus* species such as *S. salivarius* and *S. oralis*, was highly abundant in the inoculum but was depleted in the Day 0 and Day 14 biofilms.

A second, larger cluster showed the opposite trend, with low abundance in the inoculum followed by a significant increase in the biofilm state. This group, however, displayed different dynamics. A distinct subset, dominated by plasmids from *Lacticaseibacillus* species, peaked in abundance in the early Day 0 biofilm before declining again by Day 14. Another group, including plasmids from known oral inhabitants like *Fusobacterium* and *Selenomonas*, showed a more sustained pattern of high abundance throughout the biofilm's maturation from Day 0 to Day 14.

Interestingly, direct pairwise comparisons of the later biofilm stages (D8, D11, and D14) against the Day 0 baseline revealed a surprising lack of significant changes. Only a single plasmid, originating from *Lacticaseibacillus paracasei*, was found to be significantly downregulated at Day 11 compared to Day 0 (Fig 5.18C). The absence of widespread, significant changes in the later stages suggests that the core structure of the potential plasmidome is established early and

remains relatively stable, even as the bacterial community continues to evolve towards a dysbiotic state. This evolution may be due to the promiscuity of plasmids to change hosts and the closed nature of the model.

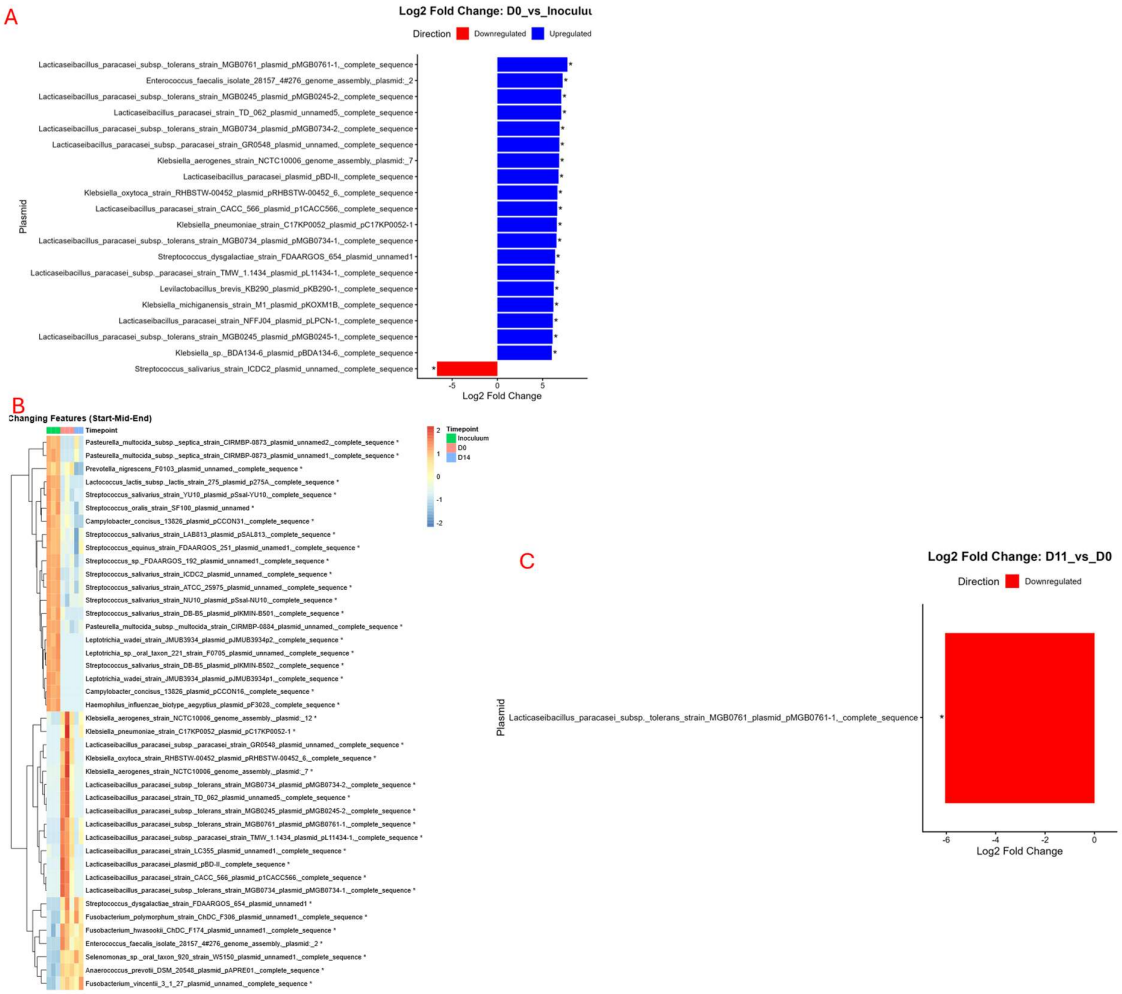


Figure 5.18. Changes in the potential plasmidome (DNA level) over time. (A) Bar chart showing the Log_2 Fold Change of plasmids between the Day 0 biofilm and the planktonic Inoculum. Blue bars indicate enrichment at Day 0, and red indicates depletion. **(B)** Heatmap showing the abundance of significantly changing plasmids across the Inoculum, Day 0, and Day 14. Yellow/orange indicates high abundance, blue indicates low abundance. **(C)** Bar chart showing the single significantly downregulated plasmid between Day 11 and Day 0.

5.2.5.4 Differential Activity of the Plasmidome

The transcriptional activity of the plasmidome revealed succession over the course of the experiment. An LRT analysis comparing the Inoculum, Day 0, and Day 14 time points highlighted a major change in plasmid expression upon biofilm formation (Fig 5.19A). A large cluster of plasmids, predominantly originating from commensal oral species such as *Streptococcus salivarius*, *Streptococcus oralis*, *Haemophilus influenzae*, and *Campylobacter concisus*, was highly active in the planktonic inoculum, but their expression was strongly repressed in both the Day 0 and Day 14 biofilms. This suggests that the functions encoded by these plasmids may contain critical functions for planktonic states but are actively shut down once the community undergoes biofilm formation. Conversely, a smaller cluster of plasmids, including those from *Anaerococcus prevotii*, showed increased activity specifically in the final Day 14 biofilm stage.

The LRT analysis across the full 14-day experiment revealed several distinct expression patterns (Fig 5.19B). Some plasmids, like those from *Fusobacterium polymorphum* and *Klebsiella aerogenes*, exhibited transient peaks of activity in the early-to-mid stages of biofilm maturation before declining. In contrast, a clear late-stage activation was observed for plasmids originating from species associated with a more dysbiotic or pathogenic state. Notably, plasmids from *Fusobacterium vincentii*, *Prevotella nigrescens*, and *Anaerococcus prevotii* all showed an increase in expression at Day 11 and Day 14. This indicates that as the biofilm transitions towards dysbiosis, there is a corresponding transcriptional activation of specific plasmids likely carrying functions that support survival and virulence in the mature pathogenic community.

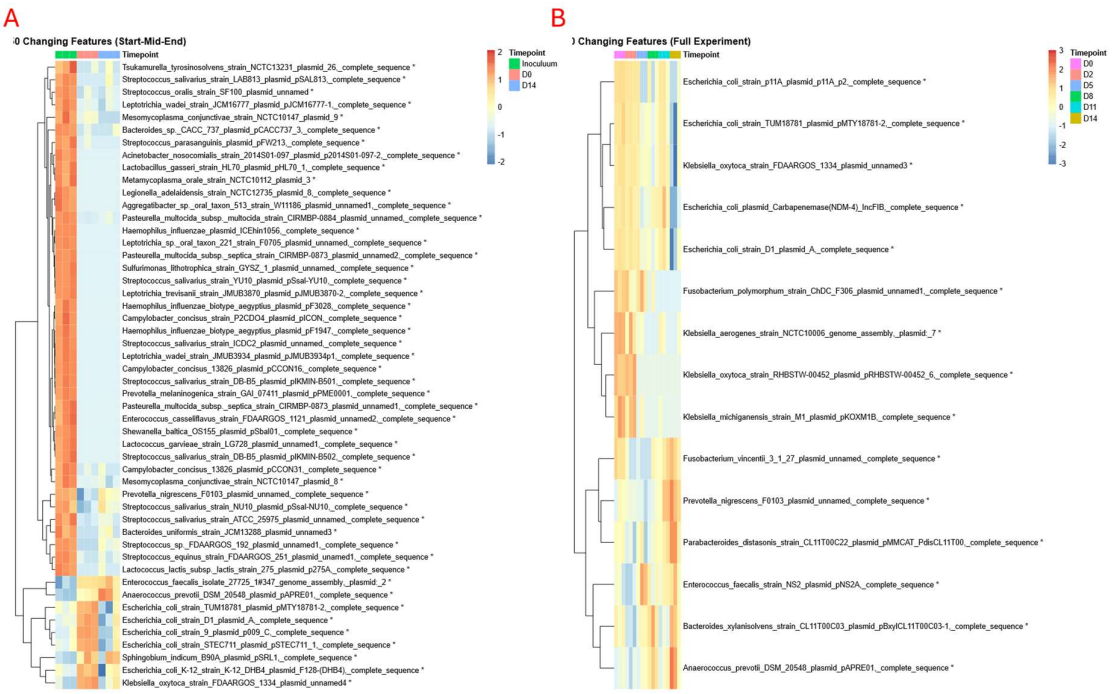


Figure 5.19. Changes in the transcriptional activity of plasmids over time. Heatmaps display the scaled and centred relative abundance of plasmids identified as significant by a Likelihood Ratio Test (LRT). Rows represent plasmids, clustered by their expression patterns, and columns represent individual samples. Yellow/orange indicates high expression, and blue indicates low expression relative to the mean. **(A)** Analysis of the key transitional stages: Inoculum, Day 0, and Day 14. **(B)** Analysis of the full experiment from Day 0 to Day 14.

5.2.5.5 Pairwise Analysis of the Active Plasmidome

Pairwise differential abundance analysis of the active plasmidome confirmed that the most profound transcriptional changes occurred during the initial transition from a planktonic to a biofilm state. When comparing the Day 0 biofilm to the inoculum, dozens of plasmids were significantly differentially expressed (Fig 5.20). The vast majority of these were significantly downregulated, including plasmids from a wide range of commensal and planktonic-associated species such as *Prevotella nigrescens*, *Streptococcus oralis*, and *Pasteurella multocida*. This indicates a community-wide transcriptional shutdown of plasmids associated with a planktonic lifestyle upon attachment and biofilm formation. In contrast, only a small number of plasmids, primarily from *Escherichia coli*, were significantly upregulated in the early biofilm.

Confirmation of the stability observed in the potential plasmidome, pairwise comparisons of the later dysbiotic time points (Day 8, Day 11, and Day 14) against the Day 0 baseline revealed no significantly differentially active plasmids. This lack of significant change in the later stages, despite the ongoing taxonomic and functional shifts in the community, suggests that the core transcriptional profile of the plasmidome is set early in biofilm development and is not substantially altered during the subsequent progression towards a mature dysbiotic state.

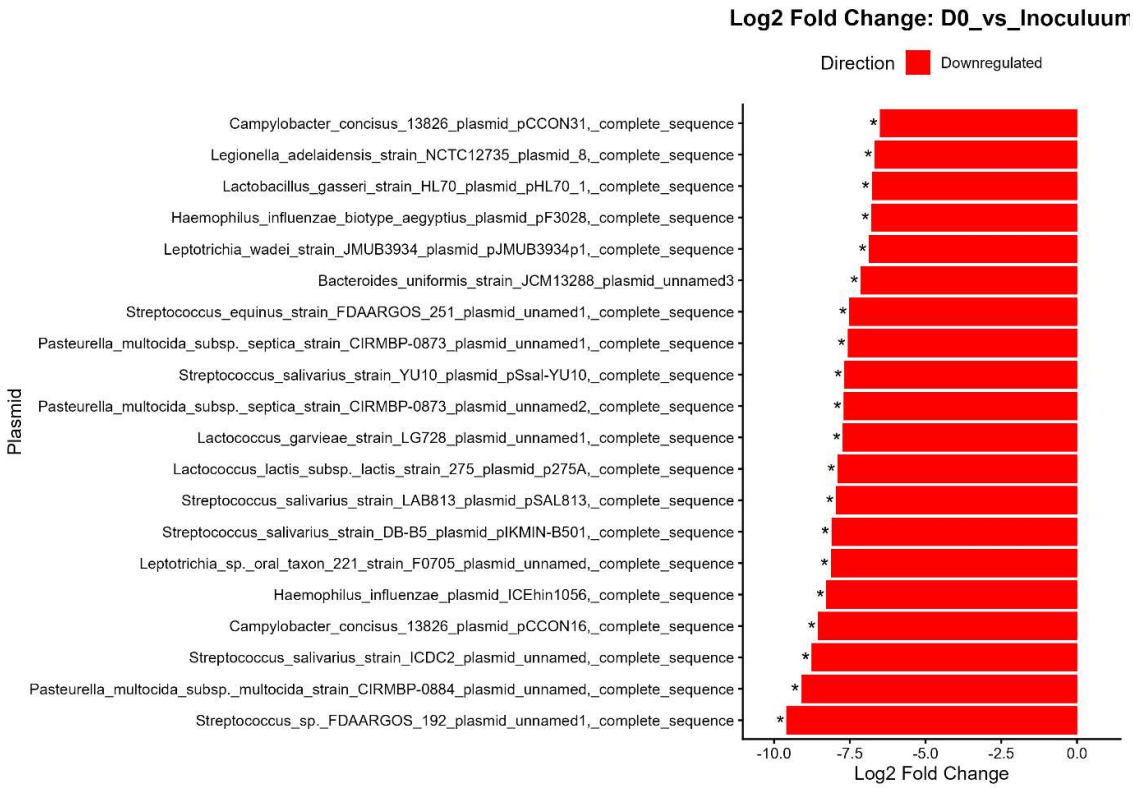


Figure 5.20. Differential transcriptional activity of plasmids between the Day 0 biofilm and the planktonic Inoculum. The bar chart displays the Log_2 Fold Change for significantly downregulated plasmids at Day 0 compared to the Inoculum. Red bars indicate lower transcriptional activity at Day 0. Asterisks denote statistical significance.

5.2.5.6 Functional Proteins of the Potential Plasmidome

To understand the functional implications of the shifting plasmidome, the genetic potential of plasmid-encoded proteins was investigated at the DNA level. An LRT analysis of the key transitional stages (Inoculum, Day 0, and Day 14) revealed a clear shift between the planktonic and biofilm states (Fig 5.21A). A distinct cluster of plasmid-borne genes, including those for core replication and transcriptional machinery such as N-6 DNA methylase and DNA topoisomerase, was at low abundance in the inoculum but became significantly enriched in the established biofilms at Day 0 and Day 14. In contrast, a large and diverse cluster of genes was highly abundant in the inoculum but became depleted and remained at low levels in the biofilm. This group was dominated by a vast number of Domains of Unknown Function (DUFs), alongside regulatory proteins and genes involved in DNA modification, such as 16S rRNA dimethyltransferase. This suggests a shift in the plasmidome's genetic potential, moving from providing a diverse and largely uncharacterised set of accessory functions for the planktonic cells to encoding core cellular processes in the biofilm state.

Longitudinal analysis across the full 14-day experiment highlighted a succession of different functional potentials encoded on the plastidome (Fig 5.21B). Genes for PTS sugar transporter systems were highly abundant in the early biofilm stages (Day 0-D5) before declining. This indicates that the early biofilm community is made for the efficient uptake of various carbohydrates. As the biofilm matured, the potential for other functions became more prominent. Genes for enzymes like alpha/beta hydrolase and those involved in specialised metabolism, such as class III lanthionine synthetase (involved in producing antimicrobial peptides), showed peaks in abundance in the later stages. The dynamic presence of various genes related to mobile elements, including multiple transposases (*IS5, IS30*) and plasmid mobilisation proteins (*MobC family*) with high abundance in the early stages, indicates that the plasmidome's genes are not static, but shaped by ongoing HGT events within the developing biofilm.

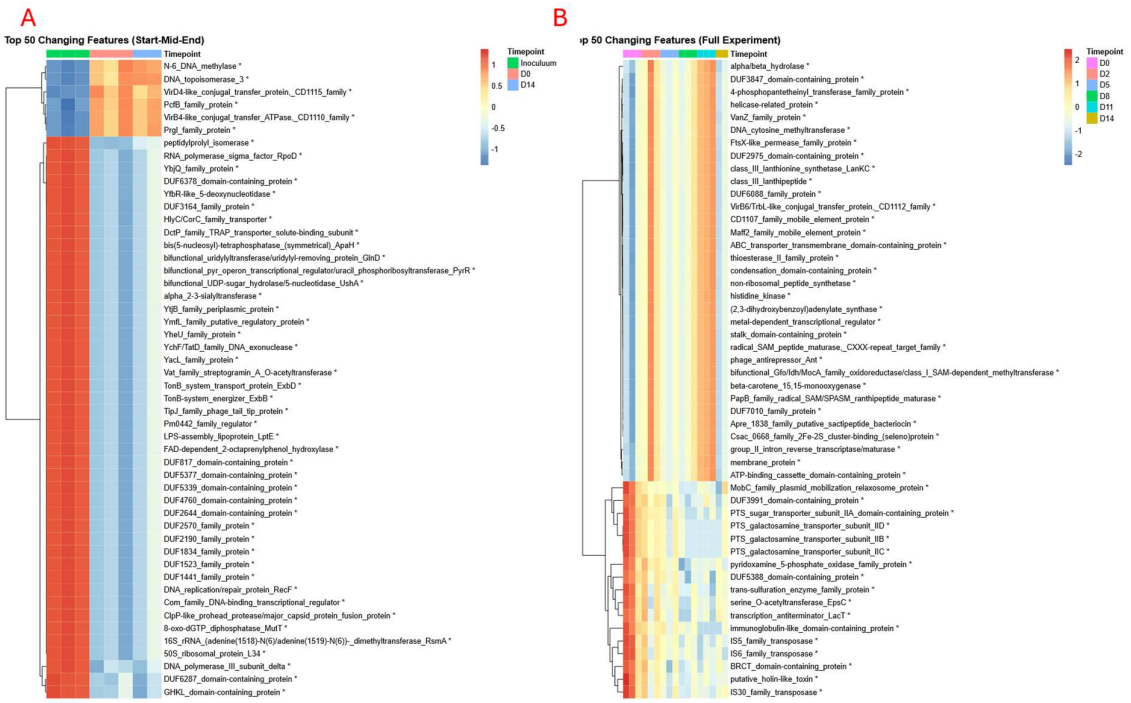


Figure 5.21. Changes in the genetic potential of plasmid-encoded proteins (DNA level) over time. Heatmaps display the scaled and centred relative abundance of plasmid-encoded proteins identified as significant by a Likelihood Ratio Test (LRT). Rows represent proteins, clustered by their abundance patterns, and columns represent individual samples. Yellow/orange indicates high abundance, and blue indicates low abundance relative to the mean. **(A)** Analysis of the key transitional stages: Inoculum, Day 0, and Day 14. **(B)** Analysis of the full experiment from Day 0 to Day 14.

5.2.5.7 Functional Genes of the Active Plasmidome

Analysis of the active functional cargo of the plasmidome, based on the metatranscriptomic data, revealed two stages of development. First, a comparison of the key transitional stages (Inoculum, Day 0, and Day 14) showed a distinction between planktonic and biofilm-associated functions (Fig 5.22A). A large cluster of plasmid-borne genes was highly expressed in the inoculum but shut down upon biofilm formation. This cluster displayed many enriched genes related to core biosynthetic and metabolic processes, including ribosomal proteins (e.g., 30S ribosomal protein S21), ATP-dependent chaperones (*CpaB*), DNA repair (*RecF*), and various transport systems (DctP family TRAP transporter). This indicates that plasmids in the planktonic state actively contribute to the fundamental cellular machinery required for growth and survival. Upon attachment, the expression of these core functions is repressed.

The longitudinal analysis across the full 14-day experiment revealed the second stage of these changes, highlighting activation of different functional gene sets as the biofilm matured (Fig 5.22B). While a wide range of genes showed dynamic expression, a distinct cluster of plasmid-borne genes was significantly and progressively upregulated in the late dysbiotic stages (Day 11 and Day 14). This late-stage functional signature was dominated by genes directly involved in HGT, including multiple conjugal transfer proteins (*TraK, TraG, MobC, TraJ, MobB*). This was accompanied by the increased expression of toxin-antitoxin systems, various methyltransferases, and a large number of DUF proteins. In contrast, other HGT-related genes, such as the *TraX* family transposon, were highly expressed in the early stages before declining. The upregulation of specific HGT machinery in the late stages strongly suggests that, as the biofilm matures into a dysbiotic state, the active role of the plasmidome shifts towards facilitating dissemination and competitive interactions within the pathogenic community.

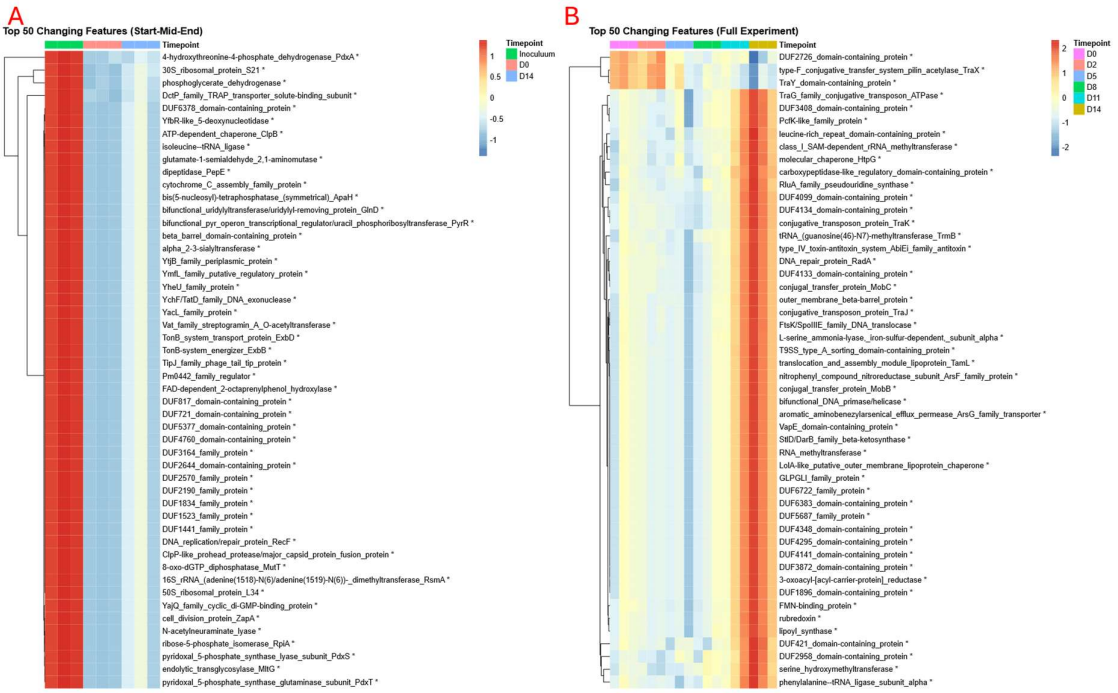


Figure 5.22. Changes in the transcriptional activity of plasmid-encoded proteins over time. Heatmaps display the scaled and centred relative abundance of plasmid-encoded proteins identified as significant by a Likelihood Ratio Test (LRT). Rows represent proteins, clustered by their expression patterns, and columns represent individual samples. Yellow/red indicates high expression, and blue indicates low expression relative to the mean. **(A)** Analysis of the key transitional stages: Inoculum, Day 0, and Day 14. **(B)** Analysis of the full experiment from Day 0 to Day 14.

5.2.5.8 Pairwise Analysis of the Active Plasmidome's Functional Genes

Pairwise analysis of the active plasmid-borne gene expression confirmed the transcriptional shutdown observed during the transition from the planktonic to the biofilm state. The comparison of the Day 0 biofilm to the inoculum revealed significant downregulation of hundreds of plasmid-encoded genes. Among the most repressed were genes central to core cellular processes, including replication (e.g., *replicative DNA helicase*, *DNA polymerase III subunit delta*), plasmid partitioning (*ParA family protein*), and protein synthesis (*50S ribosomal proteins L28 and L33*). This widespread repression of the plasmid-borne functional profile indicates a shift in community priorities away from replication, HGT, and protein synthesis upon biofilm formation. In contrast, a small number of genes were significantly upregulated, primarily related to conjugation and stress response. These included several components of the F-type conjugative transfer system (e.g., *TraF*, *TraN*) and a translation error-prone DNA polymerase.

Contrary to a stable profile, the transition into dysbiosis triggered a second, distinct wave of functional reprogramming. The comparison between Day 8 and Day 0 revealed a significant upregulation of genes involved in metabolism and biosynthesis, such as *shikimate dehydrogenase*, an *RNA polymerase sigma factor*, and *non-ribosomal peptide synthetase*. Interestingly, at this time point, two genes related to F-type conjugation (*TraY domain-containing protein* and a pilin acetylase) were downregulated. By Day 14, the profile shifted again, with a strong upregulation of genes associated with transport and HGT, including an *outer membrane beta-barrel protein*, a *T9SS type a sorting domain-containing protein*, and conjugal transfer proteins (*TraJ*, *MobB*) (Fig 5.23B). This demonstrates that rather than remaining static, the active plasmidome undergoes dynamic functional shifts as the biofilm matures, first boosting metabolic activity and later upregulating functions related to transport and conjugation in the mature dysbiotic state.

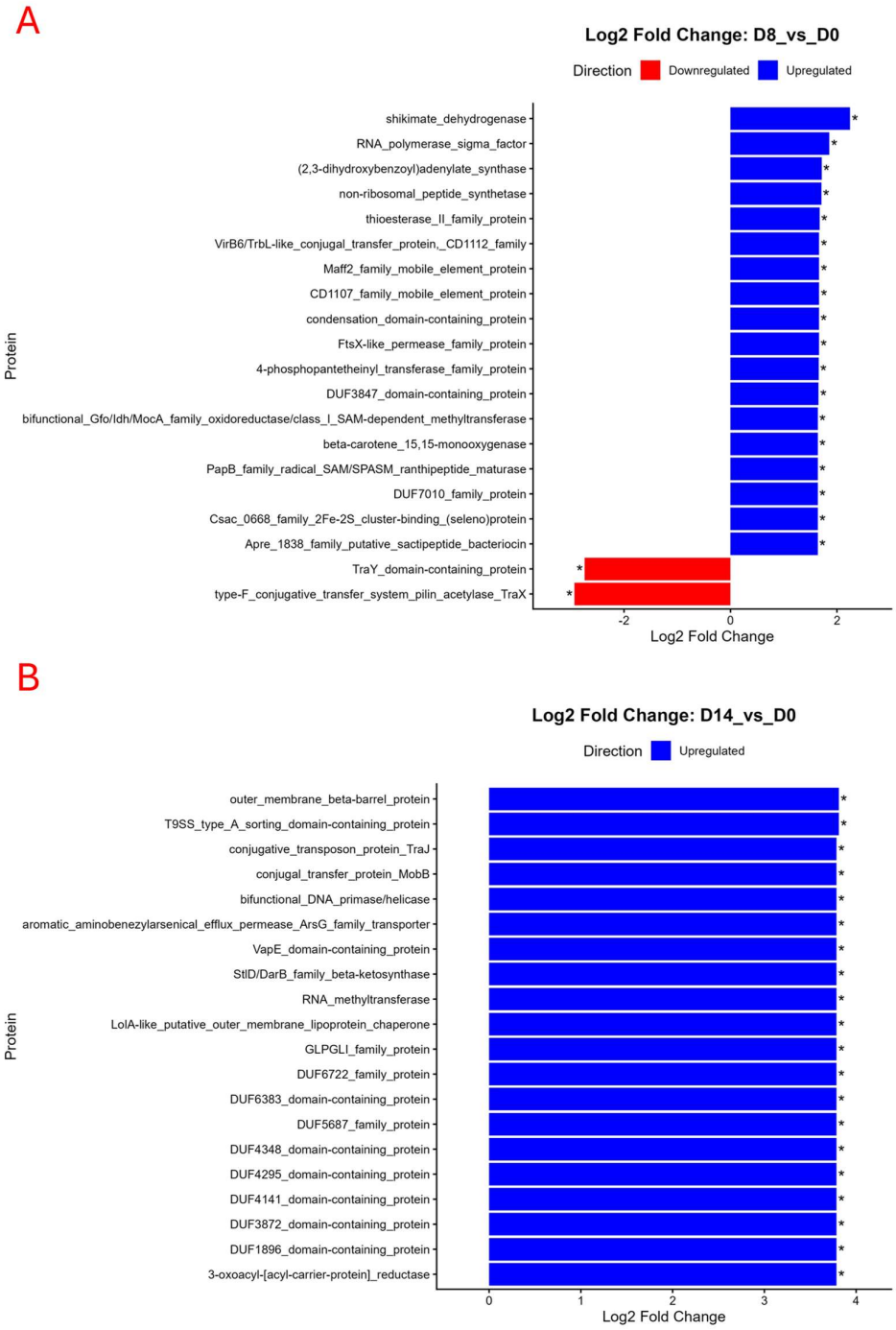


Figure 5.23. Pairwise differential transcriptional activity of plasmid-encoded proteins during biofilm maturation. Bar charts display the Log_2 Fold Change for significantly altered plasmid-borne proteins. Blue bars indicate higher expression (upregulated), and red bars indicate lower expression (downregulated) in the first condition of each comparison. Asterisks denote statistical significance. **(A)** Day 8 biofilm vs. Day 0 biofilm. **(B)** Day 14 biofilm vs. Day 0 biofilm.

5.2.6 Host Epithelial Transcriptomic Response

To determine the host's response to the developing dysbiotic biofilm, RNA was extracted from the 3D stratified epithelial tissue model at each time point and analysed by RNA-seq.

5.2.6.1 Compositional Overview of the Host Transcriptome

To provide an overview of the host response, the most transcriptionally active genes and pathways were visualised over the experimental time course. At the gene level, the host transcriptome was consistently dominated by the expression of core housekeeping genes essential for maintaining cellular structure and function (Appendix 5.3A). Genes such as actin beta (*ACTB*), multiple ribosomal proteins (*RPS3, RPS27A, RPS14* etc.), and annexin A2 (*ANXA2*) together constituted a major fraction of the total transcriptional output across all time points.

This focus on fundamental cellular processes was reflected at the pathway level. Gene Ontology (GO) term analysis showed that functions such as protein binding, cytosolic ribosome, and translation were among the most highly represented throughout the experiment (Appendix 5.3B). Similarly, the reactome pathway analysis highlighted the consistent high-level expression of pathways including Metabolism of proteins, Gene Expression (Transcription), and Immune System (Appendix 5.3C). While the overall abundance of these core pathways remained high, subtle shifts in their relative contributions were observed, particularly a notable increase in the proportion of transcripts related to the Innate Immune System in the later, dysbiotic time points. This compositional overview demonstrates that while the host tissue maintains a high level of core metabolic and structural activity, a clear immunological response signature emerges as the dysbiotic biofilm matures.

5.2.6.2 Global Shifts in Host Gene Expression

The overall host transcriptomic profile underwent a shift in response to the maturing biofilm. A PCA of the host gene expression data revealed a clear trajectory. The transcriptomes from early co-culture time points (Day 0, Day 2, Day 5, Day 8) clustered together, representing a baseline or early inflammatory response. A significant divergence was observed from Day 11 onwards, with the Day 11 and Day 14 samples forming a distinct cluster, indicating a profound change in the

host cellular state. This shift was highly significant (PERMANOVA, $R^2 = 0.849$, $p = 0.001$) confirming that the transition to a dysbiotic microbial community elicited a fundamentally different and sustained response from the host epithelial tissue (Appendix 5.6).

5.2.6.3 Longitudinal Analysis of Host Response

5.2.6.3.1 Host Gene Expression

Analysis of the host transcriptome with LRT revealed several distinct patterns of gene expression across the full 14-day experiment, highlighting a multi-stage response to the developing biofilm (Fig 5.24). An initial cluster of genes, including ribonuclease A family member 7 and keratinocyte proline-rich protein, showed high expression in the early biofilm stages (Day 0 and Day 2) before declining, suggesting an immediate, early defence response. As the biofilm matured, different expression was observed, with genes such as integrin subunit beta 1 peaking during the mid-stage of dysbiosis (Day 5 and Day 8). This may represent a transitional phase as the host attempts to modulate the initial inflammatory reaction. Finally, a large cluster of genes demonstrated a clear pattern of late-stage activation, with their expression progressively increasing to a maximum at Day 14. This group was defined by genes associated with a sustained inflammatory and tissue-remodelling response, including myocardin-related transcription factor B, FYN binding protein 1, and various pleckstrin homology domain-containing proteins. This domain is found in proteins often associated with intracellular signalling, though is found in over 11% of human proteins (Powis et al., 2023).

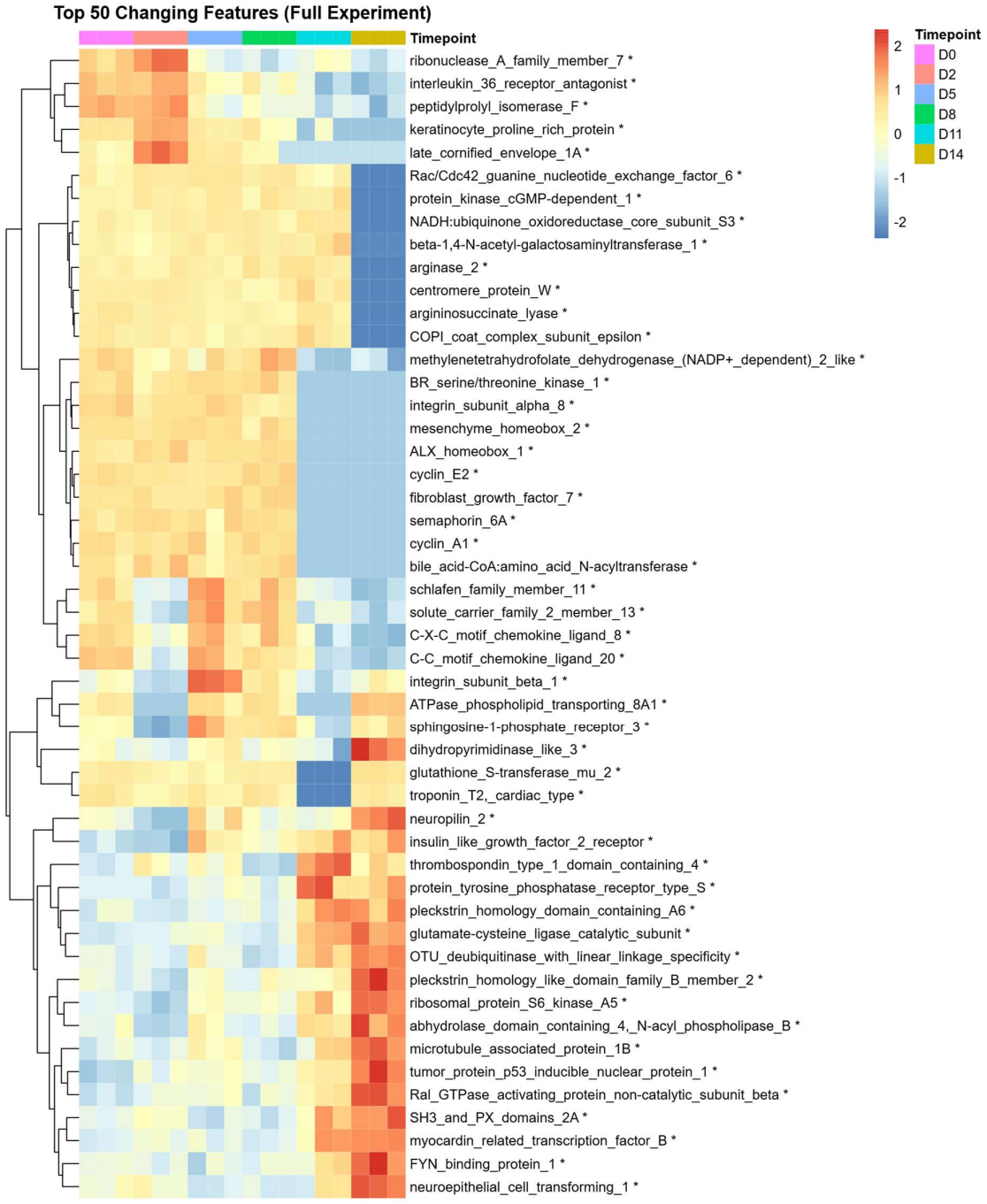


Figure 5.24. Changes in the host transcriptional activity over full experiment. Heatmaps display the scaled and centred relative abundance of plasmid-encoded proteins identified as significant by a Likelihood Ratio Test (LRT). Rows represent proteins, clustered by their expression patterns, and columns represent individual samples. Yellow/red indicates high expression, and blue indicates low expression relative to the mean. Analysis of the full experiment from Day 0 to Day 14.

A comparison of the key transitional stages (Inoculum, Day 0, and Day 14) further clarified the overall trajectory of the host response (Fig 5.25). A distinct cluster of genes associated with epithelial integrity and homeostasis, such as cadherin 1, integrin subunit alpha 2, and epidermal growth factor receptor, were lowly expressed at Day 0 but was significantly enriched by Day 14. Conversely, other groups of genes were repressed in the final dysbiotic state. One group, including proteins involved in cellular stress and remodelling such as pleckstrin, prokineticin 2, and various histones, was highly expressed in the inoculum but became depleted upon biofilm formation. Another pattern was observed for pro-inflammatory markers, including C-X-C motif chemokines ligand 5, which was strongly induced at Day 0 before being depleted in the mature Day 14 biofilm.

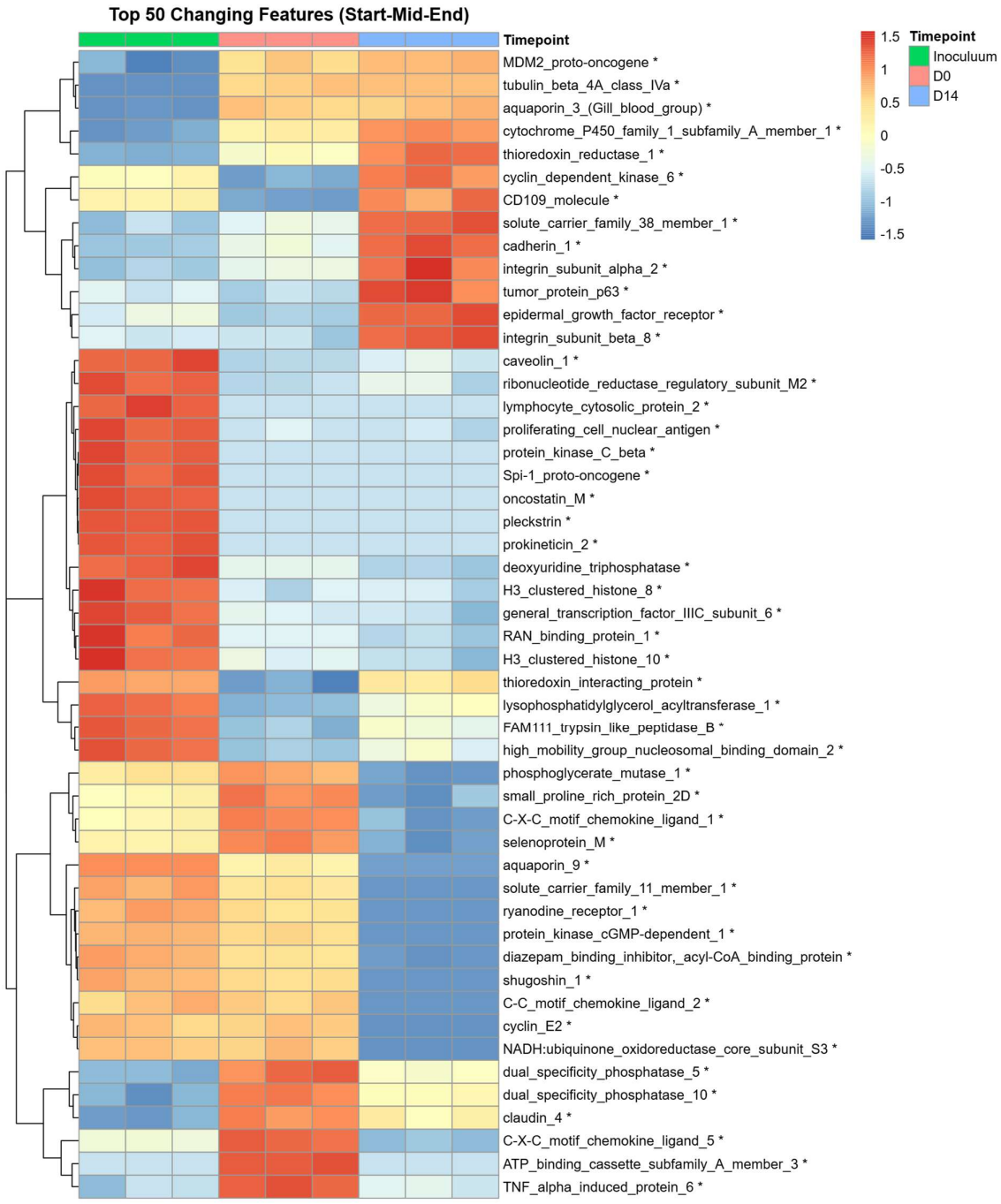


Figure 5.25. Changes in the host transcriptional activity at key timepoints. Heatmaps display the scaled and centred relative abundance of plasmid-encoded proteins identified as significant by a Likelihood Ratio Test (LRT). Rows represent proteins, clustered by their expression patterns, and columns represent individual samples. Yellow/red indicates high expression, and blue indicates low expression relative to the mean. Analysis of the key time points, inoculum, day 0 and day 14.

5.2.6.4 Pairwise Analysis of Host Gene Expression

5.2.6.4.1 Host Response to Early Biofilms (D0 vs Control)

To define the host's initial transcriptional response to the early biofilm, a pairwise comparison was performed between the Day 0 co-culture and the control. The analysis revealed an immediate host reaction. A strong pro-inflammatory signature was evident, with significant upregulation of key chemokines such as C-X-C motif chemokine ligand 5, and other immune modulators like TNF alpha induced protein 6 (Table 5.2). Concurrently, the host cells mounted an anti-apoptotic response, evidenced by the upregulation of baculoviral IAP.

In contrast, genes associated with cellular proliferation and DNA maintenance were significantly downregulated. This included critical DNA repair genes like BRCA1, cell cycle components like proliferating cell nuclear antigen and ribonucleotide reductase regulatory subunit M2, and structural proteins such as lamin B1. This transcriptional change indicates an immediate shift in cellular priorities from growth and maintenance towards a pro-inflammatory and survival-oriented state upon initial contact with the early biofilm.

Table 5.2: Key Differentially Expressed Host Genes in the planktonic vs biofilm (Day 0 vs. Control)

Feature	log2 Fold Change	Adjusted p-value
TNF alpha induced protein 6	3.07	3.50×10^{-55}
Claudin 4	2.55	5.41×10^{-93}
Calmodulin like 5	2.52	1.61×10^{-73}
Serpin family B member 2	2.44	6.24×10^{-58}
C-X-C motif chemokine ligand 5	2.40	1.20×10^{-140}
Baculoviral IAP repeat containing 3	2.09	4.47×10^{-97}
Aquaporin 3	1.69	1.68×10^{-110}
H1.5 linker histone, cluster member	-2.79	6.36×10^{-89}
Lamin B1	-2.89	2.46×10^{-108}
Proliferating cell nuclear antigen	-2.28	3.89×10^{-84}
Ribonucleotide reductase regulatory subunit M2	-3.10	4.84×10^{-141}
Thioredoxin interacting protein	-3.78	2.10×10^{-94}

Minichromosome maintenance complex component 6	-4.10	1.37×10^{-64}

5.2.6.4.2 The Host's Mid-Experiment Transitional Response (D8 vs D0)

To characterise the host's transcriptional response at the microbial Day 8, a pairwise comparison was performed between the Day 8 and Day 0 time points (Table 5.3). This analysis revealed a significant shift in host gene expression, marked by upregulation of Ankyrin 2 (ANK2), a protein involved in organising the cellular membrane architecture, suggesting a major structural change of the epithelial cells. In contrast, key components of the innate immune system were suppressed, most notably surfactant protein D, which is involved in pathogen binding and clearance. Additionally, genes related to cell cycle progression and neuronal development, such as cyclin-dependent kinase 5 regulatory subunit 2, were also significantly downregulated. This transcriptional signature suggests that by Day 8, the host epithelium is undergoing significant cellular stress and architectural changes, while some of its initial defence mechanisms are becoming compromised.

Table 5.3: Key Differentially Expressed Host Genes in the Dysbiotic Biofilm (Day 8 vs Day 0)

Feature	log2 Fold Change	Adjusted p-value
Ankyrin 2	13.92	3.95×10^{-23}
5-hydroxytryptamine receptor 1B	11.42	5.52×10^{-15}
Protein O-mannose kinase	9.21	8.35×10^{-17}
Surfactant protein D	-13.09	2.21×10^{-22}
Cyclin dependent kinase 5 regulatory subunit 2	-12.08	6.64×10^{-17}
Sodium voltage-gated channel alpha subunit 11	-12.11	8.35×10^{-17}
Caveolin 3	-10.96	1.27×10^{-11}

5.2.6.4.3 The Mature Dysbiotic Response (D14 vs D0)

The comparison of the host transcriptome in response to the mature dysbiotic biofilm (Day 14) versus the initial biofilm (Day 0) revealed a powerful and destructive signature of inflammation and tissue stress (Table 5.4). The most significantly upregulated genes were those directly involved in cellular stress responses, apoptosis, and tissue remodelling. This included a strong induction of Thioredoxin interacting protein (TXNIP), a key activator of the NLRP3 inflammasome, and TNF receptor superfamily member 21 (TNFRSF21), a death receptor that can initiate programmed cell death. This was accompanied by a massive upregulation of genes linked to pathological tissue remodelling, such as Ankyrin 2 (ANK2), indicating a drastic reorganisation of the cell's internal architecture, alongside Epidermal growth factor receptor (EGFR) and Myocardin-related transcription factor B (MRTFB).

Conversely, there was a significant downregulation of genes critical for maintaining epithelial barrier integrity. This included Keratin 6B (KRT6B) and Claudin 4 (CLDN4), a key component of the tight junctions that seal the space between cells. The loss of these structural proteins points to a fundamentally weaker and more permeable barrier. Interestingly, the analysis also shows a specific re-regulation of the chemokine response. While some initial inflammatory mediators like C-X-C motif chemokine ligand 5 and interleukin 11 were significantly downregulated compared to their levels at Day 0, the overall host response shifted towards activating inflammasome pathways and extensive tissue remodelling. Collectively, these results demonstrate that the mature dysbiotic biofilm elicits a host response characterised by the activation of potent inflammatory sensors and tissue-remodelling enzymes, while compromising epithelial barrier functions.

Table 5.4: Key Differentially Expressed Host Genes in the Mature Dysbiotic Biofilm (Day 14 vs. Day 0)

Feature	log_2 Fold Change	Adjusted p-value
Ankyrin 2	**28.29**	2.27×10^{-11}
Thioredoxin interacting protein	**2.95**	7.18×10^{-36}
Epidermal growth factor receptor	**1.55**	4.05×10^{-87}
Myocardin related transcription factor B	**1.39**	9.12×10^{-47}
TNF receptor superfamily member 21	**1.08**	7.30×10^{-44}
Interleukin 11	**-4.14**	3.16×10^{-177}
C-X-C motif chemokine ligand 5	**-3.63**	4.55×10^{-202}
Keratin 6B	**-d2.59**	3.10×10^{-86}
Claudin 4	**-0.94**	1.28×10^{-7}

5.3 Discussion

This chapter aimed to move beyond the descriptive analysis of the microbial and host 'omics data to decipher the molecular drivers of dysbiotic progression in an *in vitro* model of periodontitis. By integrating metagenomic, metatranscriptomic, and host transcriptome data, we aimed to build a model of the microbial and functional shifts that define the transition from eubiosis to dysbiosis. The findings presented here provide strong evidence for a structured, multi-stage process, highlighting the key microbial successions, functional events, and host-pathogen interactions that result in a periodontitis phenotype.

A theme that emerged is that the progression to the periodontitis-like state is a microbial process, not a random collapse of community structure. The contrast between the dynamics of alpha and beta diversity within the microbial community demonstrates this. While alpha diversity, which measures richness and evenness, remained statistically stable, the beta diversity, which measures compositional differences, revealed a significant directional shift over time. This suggests that the transition to a pathogenic state represents a highly structured process, where the specific identities of the community members are of greater consequence than the sheer number of species present. This observation aligns with the polymicrobial synergy and dysbiosis model of periodontitis, which suggests that disease arises from a restructured community rather than the action of a single pathogen (Hajishengallis and Lamont, 2012, Hajishengallis, Darveau and Curtis, 2012).

The initial transition from the planktonic inoculum to the Day 0 biofilm acted as a filter. The significant depletion of common oral inhabitants like *Neisseria perflava* and various health-associated *Streptococcus* species underscores the principle that planktonic prevalence does not predict biofilm-forming capacity. Conversely, the increase of pioneer colonisers, particularly *Actinomyces sp. oral* at the DNA level and *Streptococcus anginosus* at the RNA level, is consistent with established models of dental plaque formation. These organisms, especially *Streptococcus* and *Actinomyces* species, are recognised as the fundamental primary colonisers that create a scaffold for subsequent, later colonisers to attach and proliferate (Li *et al.*, 2004b, Radaic and Kapila, 2021) .

Following this establishment, the longitudinal analysis revealed a shift from health to disease. The early biofilm, dominated by facultative anaerobes, was gradually replaced by a consortium of obligate anaerobes renowned for their proteolytic metabolism, a marker of the subgingival environment in periodontitis (Letzelter et al., 1998). The importance of this late-stage community's features provides a biological explanation for its pathogenicity. The enrichment of species such as *Parvimonas micra, Fusobacterium vincentii*, and the keystone pathogen *Porphyromonas gingivalis* strongly aligns with the known microbiology of periodontal lesions (Neilands *et al.*, 2019, Yonezawa *et al.*, 2024). *P. micra* is a Gram-positive anaerobe frequently isolated from dental plaque in patients with chronic periodontitis and is a common constituent of mixed anaerobic infections. *P. gingivalis* is a key driver of periodontitis, containing multiple virulence factors for colonisation and host response alteration (How et al., 2016). Its emergence, both in abundance (DNA) and transcriptional activity (RNA), is a critical validation of the model's ability to reflect a clinically relevant dysbiotic state.

The beta diversity analysis identified a major structural shift occurring between Day 2 and Day 5, a critical mid-experiment transition defined by the significant depletion of early colonisers and the enrichment of several *Neisseria* species. The role of *Neisseria* in the oral microbiome is complex, while often considered commensal, its abundance shifts in dysbiotic conditions (Cui et al., 2025). This change likely reflects a critical metabolic point, where the community's primary energy source shifts from carbohydrates to the protein-rich medium. This enrichment of *Neisseria* may reflect an adaptation to the changing metabolic environment, as some species are equipped to exploit byproducts, such as lactate, generated by the early colonising community (Hoshino et al., 1976).

The functional metagenomic data reveal a community undergoing a profound strategic shift, prioritising pathogenic potential over rapid proliferation. This is evidenced by a clear functional trade-off where the genetic blueprint for core metabolic processes is sacrificed in favour of genes that facilitate a more aggressive virulence. The significant enrichment of genes for HGT, such as those for conjugal transfer and Type IV secretion systems, suggests the community is actively equipping itself to share virulence and resistance factors, thereby accelerating its adaptation to the

hostile host environment (Grohmann et al., 2018). This adaptive potential is coupled with a direct investment in pathogenicity, as shown by the increased abundance of collagen-binding domain protein genes. This signifies a genetic pivot towards direct host interaction, preparing the community to adhere to and degrade key structural components of the periodontium (Lallier et al., 2001).

This focus on virulence appears to be at a metabolic cost. The depletion of the genetic toolkit required for a healthy, proliferative state indicates a community that has moved beyond the rapid growth phase of early colonisation. This functional trade-off is a signature of dysbiosis, suggesting an adaptation to the challenging, inflammatory niche where survival and host exploitation are prioritised over simple proliferation (Maitra and Dill, 2015).

The metatranscriptomic data confirmed that this genetic potential translated into functional activity. The transcriptional profiles were initially dominated by early colonisers like *Actinomyces* and *Streptococcus*, but as the biofilm matured, their activity was replaced by that of late-stage pathogens. The massive upregulation of *Porphyromonas gingivalis* and the increased activity of *Tannerella forsythia* confirm their roles as primary functional drivers of the mature dysbiotic state. Both are core members of the "red complex," a consortium strongly associated with severe periodontitis (Suzuki, Yoneda and Hirofuji, 2013, Mohanty *et al.*, 2019)

The transcriptional comparison between Day 8 and Day 0 reveals a community undergoing a major shift. The massive upregulation of genes for mobile genetic elements, such as transposons, and a strong induction of chaperone proteins like DnaK, suggest an adaptive strategy. The upregulation of DnaK, an indicator of cellular stress, represents an essential community-wide response to protect vital machinery from the hostile host environment (Tomoyasu *et al.*, 2012, Fourie and Wilson, 2020). Simultaneously, the activation of transposons suggests the community is increasing its rate of genetic innovation to generate new genotypes better adapted to the dysbiotic niche.

The mature dysbiotic biofilm at Day 14 provides evidence for a community-level metabolic trade-off. This was defined by a downregulation of core machinery for replication and transcription,

such as RNA polymerase and DNA helicase, as well as the shutdown of early-stage metabolic pathways like carbohydrate transport via the PTS system. This redirection of resources appeared to favour the expression of late-stage pathogenic functions, such as adhesion via the collagen-binding protein (Cna protein B-type domain) and specialised carbohydrate acquisition through the highly upregulated starch-binding protein (SusD-like) (Maitra and Dill, 2015).

The analysis of the resistome and mobilome reveals a community dynamically evolving its defensive capabilities. The active resistance strategy evolved in a multi-stage process, shifting from broad-spectrum defences like antibiotic efflux in the early biofilm to more specific and efficient mechanisms of target modification and replacement in the mature pathogenic community. This suggests an adaptation to the increasingly hostile conditions of the dysbiotic niche (Reygaert, 2018). Concurrently, the mobilome analysis revealed a two-stage process. First, a biofilm-adapted plasmid pool was established at Day 0, creating an adaptive arsenal. This occurred even as the overall composition of the plasmid pool remained relatively stable after its initial establishment. This suggests the community first established an adaptive gene pool, then upregulated the machinery to share it under late-stage dysbiotic pressure.

A successful *in vitro* model must also replicate the pathological host response. The host transcriptomic analysis revealed a significant shift from a state of homeostasis to a sustained, destructive pro-inflammatory state, triggered by the microbial community's transition around Day 5-8. This was defined by a transcriptional shift away from pathways for maintaining epithelial barrier integrity toward pathways of defence, inflammation, and cellular stress.

The molecular drivers of the host's pathological state were clearly identified at Day 14. Instead of a sustained upregulation of initial inflammatory chemokines, the data point to a shift toward a destructive cellular stress response and profound tissue remodelling. A key finding was the strong induction of Thioredoxin interacting protein (TXNIP), a potent activator of the NLRP3 inflammasome, indicating the host cells are under significant oxidative stress (Ao et al., 2021). This was coupled with the upregulation of the death receptor TNFRSF21, suggesting the cells are being pushed towards apoptosis.

This internal stress response was accompanied by evidence of massive structural reorganisation (Kasof et al., 2001). The upregulation of ANK2 points to a complete rearrangement of the cell's internal cytoskeleton (Smith and Penzes, 2018). Furthermore, the induction of genes like Epidermal Growth Factor Receptor (EGFR) and Myocardin-related transcription factor B (MRTFB) suggests a dysregulated, pathological tissue remodelling process, characteristic of fibrosis and chronic inflammation (Shi et al., 2020; Venkataraman and Frieman, 2017).

Concurrently, the significant downregulation of structural proteins like Keratin 6B (KRT6B) and the tight junction component Claudin 4 (CLDN4) signifies a fundamental weakening of the host's primary physical defense (Takeuchi et al., 2022). This compromises the epithelial barrier, allowing increased penetration of bacterial products and perpetuating a destructive feedback loop.

5.3.2 Limitations and Future Directions

While this study successfully demonstrates the utility of an advanced co-culture model for dissecting periodontal disease, it is essential to acknowledge the limitations. A primary drawback is the inherent simplification of any *in vitro* system. The model cannot fully replicate the complexity of the *in vivo* environment, as key components are absent. These include a functional systemic immune system, particularly the adaptive immune response involving T-cells and B-cells, which plays a crucial role in the chronicity of periodontitis (Han et al., 2023). Furthermore, the model lacks physiological fluid dynamics, such as the constant flow of saliva and gingival crevicular fluid, which influence biofilm structure and nutrient availability (Lăzureanu et al., 2021). The cellular complexity is also simplified, omitting the direct contributions of other critical cell types like endothelial cells, osteoblasts, and osteoclasts (Brizuela and Winters, 2025).

A potential technical limitation must also be considered. Two of the Day 14 DNA samples were sequenced in a separate batch from the earlier time points. Although all data underwent standardised processing and normalisation, subtle batch effects may still influence the metagenomic profile at this final time point. Additionally, a significant dysbiotic shift was observed between the initial planktonic inoculum and the established Day 0 biofilm, which had been allowed to mature for two weeks under healthy media conditions before the start of the

experiment, but still showed early signatures of a dysbiotic shift, such as the successful colonisation by anaerobic species that would later become dominant, including *P. gingivalis*. This was potentially due to the anaerobic conditions of biofilm growth, which selected for these organisms from the initial healthy inoculum. This highlights a gap in our understanding of the very early stages of biofilm establishment.

A further consideration regarding the clinical relevance of this model lies in the sequential design of the study. As the dysbiotic shift was induced on an initial 14-day 'healthy' biofilm, it is challenging to definitively identify the effects of the periodontitis-associated media from the natural changes that occur from an additional two weeks of biofilm growth. Previous studies utilising this method for biofilm growth have avoided this issue by cultivating healthy and dysbiotic biofilms in tandem from inoculation (Naginyte et al., 2019). However, while no *in vitro* model can ever perfectly replicate the human oral cavity, the integration of complex, donor-derived biofilm with 3D stratified epithelial layers makes the system developed significantly more advanced and clinically reflective than traditional culture methods and limited multispecies models. Crucially, whether the primary driver of dysbiosis between timepoints 0 and 14 was the introduction of serum-rich media or by the extended maturation time, the resulting biofilm is an accurate representation of periodontitis. As established from the sequencing data, the Day 14 co-culture exhibits classic hallmarks of periodontitis. Taxonomically, this includes the depletion of early colonisers and enrichment of obligate anaerobes and keystone pathogens such as *P. gingivalis, T. forsythia* and *P. micra.* Functionally this is mirrored by a shift towards virulence and biofilm adaptation, such as collagen binding and transposases. Importantly, the biofilm transition successfully results in a clinically relevant host response, driving host cells to pro-inflammatory and remodelling pathways, marked by induction of TXNIP, ANK2, and EGFR alongside reduction in defences like KRT6b and CLDN4. Therefore, the dysbiotic shift of periodontitis is captured across both microbial and host responses, validating the biological relevance of the model regardless of the initiating driver. By capturing these shifts, the model provides a highly relevant transitional platform for the preclinical testing of novel therapeutics, allowing researcher to evaluate the interventions aimed at preventing or stopping dysbiosis on an

already established community, whilst understanding the effects therapeutics have on host-responses.

These limitations open several avenues for future research. A clear next step is to use this model as a platform for testing therapeutics. The ability to track both the microbial community and the host response over time makes it an ideal system for evaluating the efficacy of novel treatments, such as host-modulatory agents or antimicrobial compounds, in preventing or reversing the dysbiotic shift and mitigating the destructive host inflammatory response. Furthermore, future work could include sequencing intermediate time points between the initial inoculation and Day 0. This would provide a higher-resolution view of the critical early events in biofilm establishment and community assembly, though such an expansion was beyond the financial capabilities of the present study.

5.3.3 Conclusion

In conclusion, this study successfully leveraged metagenomics and metatranscriptomics to observe the development of a dysbiotic, periodontitis-like state *in vitro*. The results suggest a model of a structured, three-phase process with an initial establishment of a health-associated community, a mid-experiment structural transition characterised by community-wide stress and adaptation, and the final maturation into a pathogenic consortium that actively promotes a destructive host inflammatory response. This work highlights that periodontal dysbiosis is an organised microbial process that results in a functionally reprogrammed pathogenic community, which in turn elicits a specific and destructive host inflammatory cascade.

From a clinical perspective, the ultimate goal is to translate these insights into diagnostic and therapeutic tools. The identification of a distinct molecular mid-experiment structural transition and the specific host and microbial factors that define the mature pathogenic state provides a rich set of potential biomarkers and therapeutic targets. For instance, key host-derived signalling molecules, such as TXNIP, or the microbial stress-response pathways activated, could be targets for novel host-modulatory therapies aimed at preventing the transition to a destructive disease state. This work highlights the potential of sophisticated co-culture models as powerful platforms

for dissecting disease mechanisms and for preclinical evaluation of new treatments in periodontitis.

Chapter 6 General Discussion

The human oral cavity harbours a complex ecosystem where hundreds of microbial species coexist in balance that is fundamental to host health (Deo and Deshmukh, 2019). Disruptions to this equilibrium, known as dysbiosis, are the drivers of the world's most common oral diseases (Aas *et al.*, 2008, Cui, Wang and Gao, 2025). However, the nature of these dysbiotic shifts, their functional consequences, and the processes in the transition from health to disease are not understood throughout disease progression. This project aimed to dissect and compare the dysbiotic states of dental caries and periodontitis through a multi-faceted approach. The research progressed with three phases. First, a large-scale meta-analysis of publicly available metagenomic data to define the static taxonomic and functional signatures of each disease, second, the application of machine learning to quantify the predictive power of these signatures, and finally, the use of a longitudinal, multi-omics *in vitro* model to attempt to understand the mechanistic progression of periodontitis dysbiosis and the corresponding host response over time.

6.1 Defining the Signatures of Dysbiosis from Publicly Available Data

The initial phase of this project sought to establish a broad, population-level understanding of the dysbiotic states in dental caries and periodontitis. By collating and standardising publicly available shotgun metagenomic data, this work moved beyond the limitations of single-cohort studies to identify the core microbial and functional signatures that differentiate oral eubiosis and dysbiosis. A central finding was the confirmation that caries and periodontitis represent fundamentally distinct forms of dysbiosis. Periodontitis was characterised by a community-wide shift marked by a significant increase of microbial diversity as evidenced by considerably higherShannon and Simpson diversity indices ($p<0.001$) and a major compositional shift toward a known anaerobic, proteolytic community. This structural shift was statistically significant, with PERMANOVA analysis confirming that disease status accounted for a significant portion of the community variance ($R^2 = 0.078$, $p = 0.001$), driven by the enrichment of keystone pathogens such as *Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola*, the core members of the "red complex". This microbial transformation represents more than an increased abundance of pathogens but suggests the loss of beneficial functions provided by a healthy

commensal community, such as colonisation resistance against potentially pathogenic species, the maintenance of a neutral pH, and the establishment of an inflammatory environment where pathogens not only cause inflammation but proliferate under it.

In contrast, the dysbiosis associated with dental caries was not defined by a significant loss of alpha-diversity but rather by a more subtle metabolic and compositional reorganisation, with a much smaller, non-significant compositional shift (PERMANOVA $R^2 = 0.008$, $p = 0.221$). The enrichment of acidogenic and aciduric organisms, such as *Streptococcus mutans* and various Lactobacillus species defined this state. The literature on caries and diversity is mixed, with some studies reporting decreases and others reporting increases (Aas *et al.*, 2008, Gondo *et al.*, 2024). The meta-analysis performed in this study suggests that the overarching diversity is not a change in the total number of species, but rather a change in which species are dominant. This supports a model where periodontitis is the result of a broad microbial breakdown, while caries is a more specialised dysbiosis driven by the specific environmental pressure of frequent carbohydrate consumption, as in the microbial plaque hypothesis (Marsh, 2003, Rosier *et al.*, 2014b). In this model, the frequent consumption of fermentable sugars lowers the environmental pH, creating a selective pressure that only a few species can tolerate.

This work also expanded upon existing knowledge by revealing a multi-domain component to both diseases. The significant enrichment of the fungus *Candida albicans* in caries cohorts reinforces the growing view of caries as a polymicrobial disease, where cross-kingdom interactions can worsen pathogenic processes (Du et al., 2021). This aligns with *in vitro* evidence showing that *C. albicans* can coexist with *S. mutans*, co-aggregating within the biofilm and enhancing the production of extracellular glucans, which form the scaffold of the biofilm matrix, thereby increasing its structural integrity and acid tolerance (Du et al., 2021). Similarly, the detection of elevated levels of human beta- and gammaherpesviruses in periodontitis samples supports an emerging hypothesis that eukaryotic viruses may act as important cofactors in disease, perhaps by causing localised immunosuppression, altering epithelial cell cytokine profiles, modulating the local host immune environment to the benefit of pathogenic bacteria and facilitating the establishment of an anaerobic community (Shi and Gewirtz, 2018).

The functional analysis of the oral resistome and mobilome further demonstrates the distinct nature of the two diseases. The periodontitis resistome was fundamentally changed, mirroring its taxonomic collapse. It showed a significant depletion in overall ARG richness and diversity ($p<0.01$ for Observed and Chao1 indices, $p<0.001$ for Shannon), but a clinically critical enrichment of specific resistance genes, most notably various beta-lactamase genes, including multiple *TEM* and *VEB* variants often carried on mobile plasmids (Siebor et al., 2021). This finding has direct implications for the efficacy of first-line antibiotics such as amoxicillin. The caries resistome, while more stable in its overall diversity, also showed a significant restructuring of specific ARG families, particularly the MGE-associated genes *tet(W)*, *erm(B)*, and *erm(F)*, which confer resistance to tetracyclines and macrolides. Subsequent resistotyping analysis provided a novel classification, confirming these patterns by identifying a disease-associated resistome (RT2) for periodontitis while showing that the ARG profiles of caries and health were largely indistinguishable, forming a common 'core' oral resistome (RT1).

The primary limitation of this meta-analysis is its cross-sectional nature, which can identify strong correlations but cannot decipher causality or how these profiles developed. Furthermore, the reliance on metagenomics provides a view of the community's genetic potential but not its real-time functional activity. The presence of a gene does not result in its expression, which is often regulated by environmental pressures (Bengtsson-Palme et al., 2021). These limitations necessitated the subsequent phases of the project, which aimed to move from description to predictive modelling and mechanistic investigation.

6.2 Quantifying Dysbiosis with Machine Learning

Building upon the distinct signatures identified in the meta-analysis, the second phase of this project sought to quantify their predictive utility. By applying a systematic approach to both traditional and deep learning models, this work tested whether the shift observed in periodontitis would yield a more predictable signal than the subtler changes observed in dental caries.

Periodontitis emerged as a highly predictable condition. Machine learning models, particularly an AdaBoost classifier with polynomial features, consistently distinguished healthy from diseased

states with high accuracy, achieving a final test AUC-ROC of 0.86 using taxonomic profiles. The biological relevance of these models was validated by feature importance analyses, which identified known periodontal pathogens (*Aggregatibacter actinomycetemcomitans*, interactions involving *Fusobacterium* sp., and *Tannerella forsythia*) and clinically relevant ARGs (*novA* and *msbA*) as the most important features. The success of classifiers like AdaBoost, which construct a classifier by iteratively focusing on difficult-to-classify samples, suggests they are well equipped to find complex, non-linear relationships separating health and disease states within high-dimensional microbiome data (Javeed et al., 2022). This demonstrates that the community-wide shift in periodontitis is so distinct that it yields a strong, consistent signal readily learned by computational algorithms.

Conversely, the models trained to predict dental caries performed poorly. While the taxonomic profile contained a moderate predictive signal (Test AUC-ROC of 0.69), the resistome profile failed to generalise to new data, performing no better than random chance (Test AUC-ROC of 0.49). This was most clearly illustrated by the joint three-class model, which, while excellent at identifying periodontitis (correctly classifying 24 of 28 samples), frequently confused caries and healthy samples (misclassifying 8 of 19 caries samples as healthy). This provides evidence that the microbial signature of dental caries overlaps substantially with that of the healthy state, not just in membership but also in its accessory genome.

These findings have direct clinical implications. The high predictability of periodontitis suggests that the development of non-invasive, biomarker-based diagnostic tools, perhaps along with a qPCR panel targeting the top 5-10 predictive microbial or genetic features, is a realistic goal. For dental caries, however, the results indicate that DNA-level taxonomic or resistome data alone are likely insufficient for reliable prediction. This reinforces the idea that caries is primarily a disease of functional activity, the upregulation of acidogenic pathways, rather than simply genetic potential. Therefore, predictive models of caries will likely require integrating additional data types, such as metatranscriptomics or metabolomics, to capture the biofilm's active metabolic state. A significant limitation of many powerful ML models is their "black box" nature, but the growing field of Explainable AI may address this. Tools like SHAP or LIME can make models

transparent by providing prediction importance values for the influence of individual features, which is essential for translating these models into clinically trusted tools (Hassija et al., 2024).

6.3 Longitudinal Dysbiosis and the Host Response

While the first two chapters defined and quantified the static disease states, they could not explain the processes that drive the transition from health to disease. To address this, the final phase of this project utilised an advanced *in vitro* co-culture model. This system, which integrated a complex, healthy-donor-derived biofilm with a 3D stratified oral epithelial tissue, provided a platform to observe the longitudinal progression of a periodontitis-like dysbiosis and, critically, to characterise the corresponding host response over time.

The results revealed that dysbiosis development was not random but a highly structured microbial succession. The experiment demonstrated clear changes between an initial establishment phase dominated by health-associated pioneer species like *Streptococcus anginosus* and *Actinomyces sp. oral*, a critical mid-experiment transition point around Day 5-8, marked by a community-wide stress response (upregulation of chaperone proteins like DnaK), increased genetic mobility (upregulation of transposon machinery), and a metabolic shift indicated by an enrichment of *Neisseria* species, and the final maturation into a pathogenic consortium enriched with known periodontal pathogens such as *Porphyromonas gingivalis*, *Tannerella forsythia*, and *Parvimonas micra* (How, Song and Chan, 2016, Neilands *et al.*, 2019, Cui, Wang and Gao, 2025).

The multi-omics analysis of this progression provided mechanistic insights. The mature dysbiotic community was found to transcriptionally downregulate core replication machinery, such as RNA polymerase and DNA helicase, while upregulating genes for virulence and nutrient acquisition (e.g., starch-binding proteins). The active resistance strategy evolved from broad-spectrum efflux in the early biofilm to more specific mechanisms of target modification and replacement in the late stages. Furthermore, the genes for HGT, including multiple conjugal transfer proteins (*TraK, TraG, TraJ*), were transcriptionally activated in the mature biofilm, indicating that the community was actively sharing its adaptive genetic toolkit under the selective pressure of the dysbiotic environment.

Crucially, this model allowed for the integration of the host response. The host epithelial transcriptome shifted in response to the microbial transition. The microbial community's structural change around Day 8 triggered a switch in the host, from a state of homeostasis and barrier maintenance (characterised by expression of pathways for cell-cell junction assembly and EGF receptor activity) to a sustained pro-inflammatory state. The mature host response was defined by a molecular signature of severe cellular stress and pathological remodelling. This included the massive upregulation of the inflammasome activator Thioredoxin interacting protein (TXNIP) and genes involved in tissue reorganisation, like Ankyrin 2 (ANK2) and Epidermal growth factor receptor (EGFR). While initial pro-inflammatory chemokines like CXCL8 were downregulated by this stage, the sustained expression of these stress and remodelling genes. This provides direct experimental evidence for the inflammatory loop, where inflammation provides nutrients for proteolytic microbes, which in turn elicit more inflammation that drives tissue destruction in periodontitis (Sedghi et al., 2021).

While this model has inherent limitations, such as the absence of a systemic immune component (e.g., T-cells and B-cells, which are crucial for modelling chronic disease) and physiological salivary flow, it successfully recapitulates key functional and host-interactive features of periodontitis, providing a powerful platform for future studies.

6.4 Synthesis, Clinical Implications, and Future Directions

Taken together, the chapters of this thesis construct a cohesive narrative of oral dysbiosis. The meta-analysis defined distinct disease states, the machine learning analysis quantified their predictability, and the *in vitro* model explained the transition process in periodontitis. The results show a similar conclusion that periodontitis is a predictable, community-wide microbial collapse that actively drives a destructive host response, while dental caries represents a more subtle, metabolic dysbiosis that is significantly harder to distinguish from health at the genomic level.

These findings have direct implications for the future of dentistry. The high predictability of periodontitis from microbial data suggests a way for biomarker-based diagnostics. The mechanistic insights from the co-culture model, particularly the identification of the mid-stage

transition and the key molecular drivers of the host response (TXNIP, ANK2, EGFR), provide a rich set of novel targets for host-modulatory therapies. The findings on the evolution of the resistome and the activation of the HGT machinery, further supported by the meta-analysis finding of increased HGT activity from medically important pathogens such as *Enterococcus faecium* in the caries state, underscore the critical importance of the oral cavity as a reservoir for AMR. This reinforces the need for stringent antimicrobial stewardship in dentistry and situates oral health firmly within the broader One Health framework, where an ARG mobilised in the mouth can be swallowed, transferred to a gut pathogen, and enter the wider environment.

Future research should build upon these foundations. There is a pressing need for longitudinal human studies that integrate multi-omics approaches, particularly metatranscriptomic and metabolomic analyses, validate *in vitro* findings, and capture the functional activity that appears to be the key driver of caries, to identify biomarkers that predict future disease onset. A key finding from the ML analysis was the appearance of species not often associated with disease, such as *Streptococcus phages* and *Ralstonia pickettii*, as important features in classification, highlighting potentially novel targets for treatment and research. The application of explainable AI will be essential for translating "black box" algorithms into clinically trusted diagnostic tools. Finally, the advanced co-culture model developed here should be leveraged as a platform for the preclinical testing of the next generation of therapies designed not to restore microbial balance, for instance, by using prebiotics or targeted antimicrobials to disrupt the pathogenic consortium.

In conclusion, this project provides compelling evidence that the environment is a powerful selective force in driving oral disease. It demonstrates that the potential for a pathogenic state resides within the healthy oral microbiome, though it does not emerge until environmental selection. A deeper understanding of the microbial principles governing this selection process, as detailed in this work, is fundamental to advancing the future of precision oral medicine.

References

Aas, J.A., Griffen, A.L., Dardis, S.R., Lee, A.M., Olsen, I., Dewhirst, F.E., Leys, E.J., Paster, B.J., 2008. Bacteria of Dental Caries in Primary and Permanent Teeth in Children and Young Adults. J Clin Microbiol 46, 1407–1417. https://doi.org/10.1128/JCM.01410-07

Abe, K., Nomura, N., Suzuki, S., 2020. Biofilms: hot spots of horizontal gene transfer (HGT) in aquatic environments, with a focus on a new HGT mechanism. FEMS Microbiology Ecology 96, fiaa031. https://doi.org/10.1093/femsec/fiaa031

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., Narasimhan, G., 2016. Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis. Evol Bioinform Online 12, 5–16. https://doi.org/10.4137/EBO.S36436

Ahmadi, H., Ebrahimi, A., Ahmadi, F., 2021. Antibiotic Therapy in Dentistry. International Journal of Dentistry 2021, 6667624. https://doi.org/10.1155/2021/6667624

Ai, D., Huang, R., Wen, J., Li, C., Zhu, J., Xia, L.C., 2017. Integrated metagenomic data analysis demonstrates that a loss of diversity in oral microbiota is associated with periodontitis. BMC Genomics 18, 1041. https://doi.org/10.1186/s12864-016-3254-5

Ajdic, D., Chen, Z., 2012. A Novel PTS of Streptococcus mutans is Responsible for Transport of Carbohydrates with α-1,3 linkage. Molecular oral microbiology 28, 114. https://doi.org/10.1111/omi.12009

Akhtar, M., Hirt, H., Zurek, L., 2009. Horizontal transfer of the tetracycline resistance gene tetM mediated by pCF10 among Enterococcus faecalis in the house fly (Musca domestica L.) alimentary canal. Microb Ecol 58, 509–518. https://doi.org/10.1007/s00248-009-9533-9

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. https://doi.org/10.48550/arXiv.1907.10902

Alcock, B.P., Huynh, W., Chalil, R., Smith, K.W., Raphenya, A.R., Wlodarski, M.A., Edalatmand, A., Petkau, A., Syed, S.A., Tsang, K.K., Baker, S.J.C., Dave, M., McCarthy, M.C., Mukiri, K.M., Nasir, J.A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., Kwong, M., Liang, Z.C., Niu, K.C., Shan, P., Yang, J.Y.J., Gray, K.L., Hoad, G.R., Jia, B., Bhando, T., Carfrae, L.A., Farha, M.A., French, S., Gordzevich, R., Rachwalski, K., Tu, M.M., Bordeleau, E., Dooley, D., Griffiths, E., Zubyk, H.L., Brown, E.D., Maguire, F., Beiko, R.G., Hsiao, W.W.L., Brinkman, F.S.L., Van Domselaar, G., McArthur, A.G., 2023. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res 51, D690–D699. https://doi.org/10.1093/nar/gkac920

Al-Hebshi, N.N., Baraniya, D., Chen, T., Hill, J., Puri, S., Tellez, M., Hasan, N.A., Colwell, R.R., Ismail, A., 2018. Metagenome sequencing-based strain-level and functional characterization of supragingival microbiome associated with dental caries in children. J Oral Microbiol 11, 1557986. https://doi.org/10.1080/20002297.2018.1557986

Ali, K., 2024. From prescription to protection: combating antimicrobial resistance in dental practice. Evid Based Dent 25, 171–172. https://doi.org/10.1038/s41432-024-01063-x

Al-Maweri, S.A., Al-Mashraqi, A.A., Al-Qadhi, G., Al-Hebshi, N., Ba-Hattab, R., 2025. The association between the oral microbiome and hypertension: a systematic review. J Oral Microbiol 17, 2459919. https://doi.org/10.1080/20002297.2025.2459919

Ao, H., Li, H., Zhao, X., Liu, B., Lu, L., 2021. TXNIP positively regulates the autophagy and apoptosis in the rat müller cell of diabetic retinopathy. Life Sci 267, 118988. https://doi.org/10.1016/j.lfs.2020.118988

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N., Yeh, L.-S.L., 2004. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res 32, D115–D119. https://doi.org/10.1093/nar/gkh131

Aruni, A.W., Mishra, A., Dou, Y., Chioma, O., Hamilton, B.N., Fletcher, H.M., 2015. Filifactor alocis — a new emerging periodontal pathogen. Microbes Infect 17, 517–530. https://doi.org/10.1016/j.micinf.2015.03.011

Atlas, R.M., 2013. One Health: its origins and future. Curr Top Microbiol Immunol 365, 1–13. https://doi.org/10.1007/82_2012_223

Attebery, H.R., Finegold, S.M., 1969. Combined screw-cap and rubber-stopper closure for Hungate tubes (pre-reduced anaerobically sterilized roll tubes and liquid media). Appl Microbiol 18, 558–561. https://doi.org/10.1128/am.18.4.558-561.1969

Baker, J.L., Edlund, A., 2019. Exploiting the Oral Microbiome to Prevent Tooth Decay: Has Evolution Already Provided the Best Tools? Front Microbiol 9, 3323. https://doi.org/10.3389/fmicb.2018.03323

Baker, J.L., Faustoferri, R.C., Quivey, R.G., 2017. Acid-adaptive mechanisms of Streptococcus mutans–the more we know, the more we don't. Mol Oral Microbiol 32, 107–117. https://doi.org/10.1111/omi.12162

Baker, J.L., Mark Welch, J.L., Kauffman, K.M., McLean, J.S., He, X., 2024. The oral microbiome: diversity, biogeography and human health. Nat Rev Microbiol 22, 89–104. https://doi.org/10.1038/s41579-023-00963-6

Bargheet, A., Klingenberg, C., Esaiassen, E., Hjerde, E., Cavanagh, J.P., Bengtsson-Palme, J., Pettersen, V.K., 2023. Development of early life gut resistome and mobilome across gestational ages and microbiota-modifying treatments. eBioMedicine 92, 104613. https://doi.org/10.1016/j.ebiom.2023.104613

Barlow, M., 2009. What antimicrobial resistance has taught us about horizontal gene transfer. Methods Mol Biol 532, 397–411. https://doi.org/10.1007/978-1-60327-853-9_23

Baty, J.J., Stoner, S.N., Scoffield, J.A., 2022. Oral Commensal Streptococci: Gatekeepers of the Oral Cavity. J Bacteriol 204, e00257-22. https://doi.org/10.1128/jb.00257-22

Baumgardner, D.J., 2019. Oral Fungal Microbiota: To Thrush and Beyond. J Patient Cent Res Rev 6, 252–261. https://doi.org/10.17294/2330-0698.1705

Belle, V., Papantonis, I., 2021. Principles and Practice of Explainable Machine Learning. Front. Big Data 4. https://doi.org/10.3389/fdata.2021.688969

Belle, V., Papantonis, I., 2020. Principles and Practice of Explainable Machine Learning. https://doi.org/10.48550/arXiv.2009.11698

Belmok, A., de Cena, J.A., Kyaw, C.M., Damé-Teixeira, N., 2020. The Oral Archaeome: A Scoping Review. J Dent Res 99, 630–643. https://doi.org/10.1177/0022034520910435

Bengtsson-Palme, J., Jonsson, V., Heß, S., 2021. What Is the Role of the Environment in the Emergence of Novel Antibiotic Resistance Genes? A Modeling Approach. Environ Sci Technol 55, 15734–15743. https://doi.org/10.1021/acs.est.1c02977

Benítez-Páez, A., Belda-Ferre, P., Simón-Soro, A., Mira, A., 2014. Microbiota diversity and gene expression dynamics in human oral biofilms. BMC Genomics 15, 311. https://doi.org/10.1186/1471-2164-15-311

Bhaumik, D., Salzman, E., Davis, E., Blostein, F., Li, G., Neiswanger, K., Weyant, R.J., Crout, R., McNeil, D.W., Marazita, M.L., Foxman, B., 2024. Plaque Microbiome in Caries-Active and Caries-Free Teeth by Dentition. JDR Clin Trans Res 9, 61–71. https://doi.org/10.1177/23800844221121260

Blagus, R., Lusa, L., 2013. SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics 14, 106. https://doi.org/10.1186/1471-2105-14-106

Bowen, W.H., Koo, H., 2011. Biology of Streptococcus mutans- Derived Glucosyltransferases: Role in Extracellular Matrix Formation of Cariogenic Biofilms. Caries Research 45, 69–86. https://doi.org/10.1159/000324598

Bradshaw, D.J., Marsh, P.D., 1998. Analysis of pH-driven disruption of oral microbial communities in vitro. Caries Res 32, 456–462. https://doi.org/10.1159/000016487

Brancaccio, G., Balato, A., Malvehy, J., Puig, S., Argenziano, G., Kittler, H., 2024. Artificial Intelligence in Skin Cancer Diagnosis: A Reality Check. J Invest Dermatol 144, 492–499. https://doi.org/10.1016/j.jid.2023.10.004

Brizuela, M., Winters, R., 2025. Histology, Oral Mucosa, in: StatPearls. StatPearls Publishing, Treasure Island (FL).

Brodersen, J., Siersma, V.D., 2013. Long-term psychosocial consequences of false-positive screening mammography. Ann Fam Med 11, 106–115. https://doi.org/10.1370/afm.1466

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12, 59–60. https://doi.org/10.1038/nmeth.3176

Burgin, J., Ahamed, A., Cummins, C., Devraj, R., Gueye, K., Gupta, D., Gupta, V., Haseeb, M., Ihsan, M., Ivanov, E., Jayathilaka, S., Balavenkataraman Kadhirvelu, V., Kumar, M., Lathi, A., Leinonen, R., Mansurova, M., McKinnon, J., O'Cathail, C., Paupério, J., Pesant, S., Rahman, N., Rinck, G., Selvakumar, S., Suman, S., Vijayaraja, S., Waheed, Z., Woollard, P., Yuan, D., Zyoud, A., Burdett, T., Cochrane, G., 2022. The European Nucleotide Archive in 2022.

Burmeister, A.R., 2015. Horizontal Gene Transfer. Evol Med Public Health 2015, 193–194. https://doi.org/10.1093/emph/eov018

Butcher, M.C., Short, B., Veena, C.L.R., Bradshaw, D., Pratten, J.R., McLean, W., Shaban, S.M.A., Ramage, G., Delaney, C., 2022. Meta-analysis of caries microbiome studies can improve upon disease prediction outcomes. APMIS 130, 763–777. https://doi.org/10.1111/apm.13272

Carr, V.R., Shkoporov, A., Hill, C., Mullany, P., Moyes, D.L., 2021. Probing the Mobilome: Discoveries in the Dynamic Microbiome. Trends in Microbiology 29, 158–170. https://doi.org/10.1016/j.tim.2020.05.003

Cattaneo, C., Gargari, G., Koirala, R., Laureati, M., Riso, P., Guglielmetti, S., Pagliarini, E., 2019. New insights into the relationship between taste perception and oral microbiota composition. Sci Rep 9, 3549. https://doi.org/10.1038/s41598-019-40374-3

Caufield, P.W., Schön, C.N., Saraithong, P., Li, Y., Argimón, S., 2015. Oral Lactobacilli and Dental Caries. J Dent Res 94, 110S-118S. https://doi.org/10.1177/0022034515576052

Chang, S.E., Foster, S., Betts, D., Marnock, W.E., 1992. DOK, a cell line established from human dysplastic oral mucosa, shows a partially transformed non-malignant phenotype. Int J Cancer 52, 896–902. https://doi.org/10.1002/ijc.2910520612

Chen, C., Hemme, C., Beleno, J., Shi, Z.J., Ning, D., Qin, Y., Tu, Q., Jorgensen, M., He, Z., Wu, L., Zhou, J., 2018. Oral microbiota of periodontal health and disease and their changes after nonsurgical periodontal therapy. ISME J 12, 1210–1224. https://doi.org/10.1038/s41396-017-0037-1

Chen, T., Yu, W.-H., Izard, J., Baranova, O.V., Lakshmanan, A., Dewhirst, F.E., 2010. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. Database (Oxford) 2010, baq013. https://doi.org/10.1093/database/baq013

Chen, Y., Huang, Z., Tang, Z., Huang, Y., Huang, M., Liu, H., Ziebolz, D., Schmalz, G., Jia, B., Zhao, J., 2022. More Than Just a Periodontal Pathogen –the Research Progress on Fusobacterium nucleatum. Front Cell Infect Microbiol 12, 815318. https://doi.org/10.3389/fcimb.2022.815318

Chenicheri, S., R, U., Ramachandran, R., Thomas, V., Wood, A., 2017. Insight into Oral Biofilm: Primary, Secondary and Residual Caries and Phyto-Challenged Solutions. Open Dent J 11, 312–333. https://doi.org/10.2174/1874210601711010312

Christersson, L.A., Zambon, J.J., Genco, R.J., 1991. Dental bacterial plaques. Nature and role in periodontal disease. J Clin Periodontol 18, 441–446. https://doi.org/10.1111/j.1600-051x.1991.tb02314.x

Costalonga, M., Herzberg, M.C., 2014. The oral microbiome and the immunobiology of periodontal disease and caries. Immunology Letters, Microbiome influences on host immunity 162, 22–38. https://doi.org/10.1016/j.imlet.2014.08.017

Cui, Z., Wang, P., Gao, W., 2025. Microbial dysbiosis in periodontitis and peri-implantitis: pathogenesis, immune responses, and therapeutic. Front Cell Infect Microbiol 15, 1517154. https://doi.org/10.3389/fcimb.2025.1517154

Dame-Teixeira, N., Lynch, J., Yu, X., Cena, J.A., Do, T., 2025. The Caries and Caries-Free Archaeome. J Dent Res 00220345251329343. https://doi.org/10.1177/00220345251329343

Dame-Teixeira, N., Parolo, C.C.F., Maltz, M., Tugnait, A., Devine, D., Do, T., 2016. Actinomyces spp. gene expression in root caries lesions. J Oral Microbiol 8, 10.3402/jom.v8.32383. https://doi.org/10.3402/jom.v8.32383

Darrene, L.-N., Cecile, B., 2016. Experimental Models of Oral Biofilms Developed on Inert Substrates: A Review of the Literature. Biomed Res Int 2016, 7461047. https://doi.org/10.1155/2016/7461047

Davis-Turak, J., Courtney, S.M., Hazard, E.S., Glen, W.B., da Silveira, W., Wesselman, T., Harbin, L.P., Wolf, B.J., Chung, D., Hardiman, G., 2017. Genomics pipelines and data integration: challenges and opportunities in the research setting. Expert Rev Mol Diagn 17, 225–237. https://doi.org/10.1080/14737159.2017.1282822

Dawson, R.J.P., Locher, K.P., 2006. Structure of a bacterial multidrug ABC transporter. Nature 443, 180–185. https://doi.org/10.1038/nature05155

DeGruttola, A.K., Low, D., Mizoguchi, A., Mizoguchi, E., 2016. Current understanding of dysbiosis in disease in human and animal models. Inflamm Bowel Dis 22, 1137–1150. https://doi.org/10.1097/MIB.0000000000000750

Delgado-Baquerizo, M., Hu, H.-W., Maestre, F.T., Guerra, C.A., Eisenhauer, N., Eldridge, D.J., Zhu, Y.-G., Chen, Q.-L., Trivedi, P., Du, S., Makhalanyane, T.P., Verma, J.P., Gozalo, B., Ochoa, V., Asensio, S., Wang, L., Zaady, E., Illán, J.G., Siebe, C., Grebenc, T., Zhou, X., Liu, Y.-R., Bamigboye, A.R., Blanco-Pastor, J.L., Duran, J., Rodríguez, A., Mamet, S., Alfaro, F., Abades, S., Teixido, A.L., Peñaloza-Bojacá, G.F., Molina-Montenegro, M.A., Torres-Díaz, C., Perez, C., Gallardo, A., García-Velázquez, L., Hayes, P.E., Neuhauser, S., He, J.-Z., 2022. The global distribution and environmental drivers of the soil antibiotic resistome. Microbiome 10, 219. https://doi.org/10.1186/s40168-022-01405-w

Deo, P.N., Deshmukh, R., 2019. Oral microbiome: Unveiling the fundamentals. Journal of Oral and Maxillofacial Pathology : JOMFP 23, 122. https://doi.org/10.4103/jomfp.JOMFP_304_18

Dewhirst, F.E., Chen, T., Izard, J., Paster, B.J., Tanner, A.C.R., Yu, W.-H., Lakshmanan, A., Wade, W.G., 2010. The Human Oral Microbiome. J Bacteriol 192, 5002–5017. https://doi.org/10.1128/JB.00542-10

Dini, C., Costa, R.C., Bertolini, M., Shibli, J.A., Feres, M., Klein, M.I., de Avila, É.D., Souza, J.G.S., Barão, V.A.R., 2023. In-vitro polymicrobial oral biofilm model represents clinical microbial profile and disease progression during implant-related infections. J Appl Microbiol 134, lxad265. https://doi.org/10.1093/jambio/lxad265

Dinis, M., Agnello, M., Cen, L., Shokeen, B., He, X., Shi, W., Wong, D.T.W., Lux, R., Tran, N.C., 2022. Oral Microbiome: Streptococcus mutans/Caries Concordant-Discordant Children. Front Microbiol 13, 782825. https://doi.org/10.3389/fmicb.2022.782825

Dion, M.B., Oechslin, F., Moineau, S., 2020. Phage diversity, genomics and phylogeny. Nat Rev Microbiol 18, 125–138. https://doi.org/10.1038/s41579-019-0311-5

Do, T., Devine, D., Marsh, P.D., 2013. Oral biofilms: molecular analysis, challenges, and future prospects in dental diagnostics. Clin Cosmet Investig Dent 5, 11–19. https://doi.org/10.2147/CCIDE.S31005

Donlan, R.M., Costerton, J.W., 2002. Biofilms: Survival Mechanisms of Clinically Relevant Microorganisms. Clin Microbiol Rev 15, 167–193. https://doi.org/10.1128/CMR.15.2.167-193.2002

Dorst, M., Zeevenhooven, N., Wilding, R., Mende, D., Brandt, B.W., Zaura, E., Hoekstra, A., Sheraton, V.M., 2024. FAIR compliant database development for human microbiome data samples. Front Cell Infect Microbiol 14, 1384809. https://doi.org/10.3389/fcimb.2024.1384809

Du, Q., Ren, B., He, J., Peng, X., Guo, Q., Zheng, L., Li, J., Dai, H., Chen, V., Zhang, L., Zhou, X., Xu, X., 2021. Candida albicans promotes tooth decay by inducing oral microbial dysbiosis. ISME J 15, 894–908. https://doi.org/10.1038/s41396-020-00823-8

Duran-Pinedo, A.E., 2021. Metatranscriptomic analyses of the oral microbiome. Periodontol 2000 85, 28–45. https://doi.org/10.1111/prd.12350

Ellabaan, M.M.H., Munck, C., Porse, A., Imamovic, L., Sommer, M.O.A., 2021. Forecasting the dissemination of antibiotic resistance genes across bacterial genomes. Nat Commun 12, 2435. https://doi.org/10.1038/s41467-021-22757-1

Eribe, E.R.K., Olsen, I., 2008. Leptotrichia species in human infections. Anaerobe 14, 131–137. https://doi.org/10.1016/j.anaerobe.2008.04.004

Falsetta, M.L., Klein, M.I., Colonne, P.M., Scott-Anne, K., Gregoire, S., Pai, C.-H., Gonzalez-Begne, M., Watson, G., Krysan, D.J., Bowen, W.H., Koo, H., 2014. Symbiotic Relationship between Streptococcus mutans and Candida albicans Synergizes Virulence

of Plaque Biofilms In Vivo. Infect Immun 82, 1968–1981. https://doi.org/10.1128/IAI.00087-14

Faran Ali, S.M., Tanwir, F., 2012. Oral microbial habitat a dynamic entity. J Oral Biol Craniofac Res 2, 181–187. https://doi.org/10.1016/j.jobcr.2012.07.001

Fässler, D., Heinken, A., Hertel, J., 2025. Characterising functional redundancy in microbiome communities via relative entropy. Comput Struct Biotechnol J 27, 1482–1497. https://doi.org/10.1016/j.csbj.2025.03.012

Fernández-Trapote, E., Oliveira, M., Cobo-Díaz, J.F., Alvarez-Ordóñez, A., 2024. The resistome of the food chain: A One Health perspective. Microb Biotechnol 17, e14530. https://doi.org/10.1111/1751-7915.14530

Forsberg, K.J., Reyes, A., Wang, B., Selleck, E.M., Sommer, M.O.A., Dantas, G., 2012. The shared antibiotic resistome of soil bacteria and human pathogens. Science 337, 1107–1111. https://doi.org/10.1126/science.1220761

Fourie, K.R., Wilson, H.L., 2020. Understanding GroEL and DnaK Stress Response Proteins as Antigens for Bacterial Diseases. Vaccines (Basel) 8, 773. https://doi.org/10.3390/vaccines8040773

Furquim, C.P., Caruth, L., Chandrasekaran, G., Cucchiara, A., Kallan, M.J., Martin, L., Feres, M., Bittinger, K., Divaris, K., Glessner, J., Kantarci, A., Giannobile, W., Verma, S.S., Teles, F., 2025. Developing Predictive Models for Periodontitis Progression Using Artificial Intelligence: A Longitudinal Cohort Study. J Clin Periodontol 52, 1478–1490. https://doi.org/10.1111/jcpe.14194

Galperin, M.Y., Koonin, E.V., 2004. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic Acids Res 32, 5452–5463. https://doi.org/10.1093/nar/gkh885

Gendron, R., Grenier, D., Maheu-Robert, L., 2000. The oral cavity as a reservoir of bacterial pathogens for focal infections. Microbes Infect 2, 897–906. https://doi.org/10.1016/s1286-4579(00)00391-9

Geng, M., Li, M., Li, Y., Zhu, J., Sun, C., Wang, Y., Chen, W.-H., 2024. A universal oral microbiome-based signature for periodontitis. iMeta 3, e212. https://doi.org/10.1002/imt2.212

Ghannoum, M.A., Jurevic, R.J., Mukherjee, P.K., Cui, F., Sikaroodi, M., Naqvi, A., Gillevet, P.M., 2010. Characterization of the Oral Fungal Microbiome (Mycobiome) in Healthy Individuals. PLoS Pathog 6, e1000713. https://doi.org/10.1371/journal.ppat.1000713

Goetting-Minesky, M.P., Godovikova, V., Fenno, J.C., 2021. Approaches to Understanding Mechanisms of Dentilisin Protease Complex Expression in Treponema denticola. Front Cell Infect Microbiol 11, 668287. https://doi.org/10.3389/fcimb.2021.668287

Gondo, T., Hiraishi, N., Takeuchi, A., Moyes, D., Shimada, Y., 2024. Comparative analysis of microbiome in coronal and root caries. BMC Oral Health 24, 869. https://doi.org/10.1186/s12903-024-04670-3

Graham, D.W., Bergeron, G., Bourassa, M.W., Dickson, J., Gomes, F., Howe, A., Kahn, L.H., Morley, P.S., Scott, H.M., Simjee, S., Singer, R.S., Smith, T.C., Storrs, C., Wittum, T.E., 2019. Complexities in understanding antimicrobial resistance across domesticated animal, human, and environmental systems. Ann N Y Acad Sci 1441, 17–30. https://doi.org/10.1111/nyas.14036

Grieneisen, L., Blekhman, R., Archie, E., 2023. How longitudinal data can contribute to our understanding of host genetic effects on the gut microbiome. Gut Microbes 15, 2178797. https://doi.org/10.1080/19490976.2023.2178797

Grohmann, E., Christie, P.J., Waksman, G., Backert, S., 2018. Type IV secretion in Gram-negative and Gram-positive bacteria. Mol Microbiol 107, 455–471. https://doi.org/10.1111/mmi.13896

Haffajee, A.D., Socransky, S.S., 2005. Microbiology of periodontal diseases: introduction. Periodontology 2000 38, 9–12. https://doi.org/10.1111/j.1600-0757.2005.00112.x

Hajishengallis, G., Darveau, R.P., Curtis, M.A., 2012. The keystone-pathogen hypothesis. Nat Rev Microbiol 10, 717–725. https://doi.org/10.1038/nrmicro2873

Hajishengallis, G., Lamont, R.J., 2012. Beyond the red complex and into more complexity: the polymicrobial synergy and dysbiosis (PSD) model of periodontal disease etiology. Mol Oral Microbiol 27, 409–419. https://doi.org/10.1111/j.2041-1014.2012.00663.x

Han, N., Liu, Yitong, Du, J., Xu, J., Guo, L., Liu, Yi, 2023. Regulation of the Host Immune Microenvironment in Periodontitis and Periodontal Bone Remodeling. Int J Mol Sci 24, 3158. https://doi.org/10.3390/ijms24043158

Hao, T., Zhang, R., Zhao, T., Wu, J., Leung, W.K., Yang, J., Sun, W., 2023. Porphyromonas gingivalis infection promotes inflammation via inhibition of the AhR signalling pathway in periodontitis. Cell Prolif 56, e13364. https://doi.org/10.1111/cpr.13364

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., Hussain, A., 2024. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. Cogn Comput 16, 45–74. https://doi.org/10.1007/s12559-023-10179-8

He, J., Kim, D., Zhou, X., Ahn, S.-J., Burne, R.A., Richards, V.P., Koo, H., 2017. RNA-Seq Reveals Enhanced Sugar Metabolism in Streptococcus mutans Co-cultured with Candida albicans within Mixed-Species Biofilms. Front Microbiol 8, 1036. https://doi.org/10.3389/fmicb.2017.01036

Healey, G.R., Murphy, R., Brough, L., Butts, C.A., Coad, J., 2017. Interindividual variability in gut microbiota and host response to dietary interventions. Nutr Rev 75, 1059–1080. https://doi.org/10.1093/nutrit/nux062

Hewitt, B., Batt, J., Shelton, R.M., Cooper, P.R., Landini, G., Lucas, R.A., Wiench, M., Milward, M.R., 2022. A 3D Printed Device for In Vitro Generation of Stratified Epithelia at the Air-Liquid Interface. Tissue Eng Part C Methods 28, 599–609. https://doi.org/10.1089/ten.TEC.2022.0130

Hollingshead, C.M., Brizuela, M., 2023. Antibiotic Prophylaxis in Dental and Oral Surgery Practice, in: StatPearls [Internet]. StatPearls Publishing.

Home [WWW Document], n.d. . Advanced Research Computing. URL https://arc.leeds.ac.uk/ (accessed 10.19.25).

Hoshino, E., Yamada, T., Araya, S., 1976. Lactate degradation by a strain of Neisseria isolated from human dental plaque. Arch Oral Biol 21, 677–683. https://doi.org/10.1016/0003-9969(76)90142-4

Hou, K., Wu, Z.-X., Chen, X.-Y., Wang, J.-Q., Zhang, D., Xiao, C., Zhu, D., Koya, J.B., Wei, L., Li, J., Chen, Z.-S., 2022. Microbiota in health and diseases. Sig Transduct Target Ther 7, 1–28. https://doi.org/10.1038/s41392-022-00974-4

How, K.Y., Song, K.P., Chan, K.G., 2016. Porphyromonas gingivalis: An Overview of Periodontopathic Pathogen below the Gum Line. Front Microbiol 7, 53. https://doi.org/10.3389/fmicb.2016.00053

Hsu, T.Y., Nzabarushimana, E., Wong, D., Luo, C., Beiko, R.G., Langille, M., Huttenhower, C., Nguyen, L.H., Franzosa, E.A., 2023. Profiling novel lateral gene transfer events in the human microbiome. bioRxiv 2023.08.08.552500. https://doi.org/10.1101/2023.08.08.552500

Hubrecht, R.C., Carter, E., 2019. The 3Rs and Humane Experimental Technique: Implementing Change. Animals (Basel) 9, 754. https://doi.org/10.3390/ani9100754

Imamura, T., 2003. The role of gingipains in the pathogenesis of periodontal disease. J Periodontol 74, 111–118. https://doi.org/10.1902/jop.2003.74.1.111

Iniesta, M., Vasconcelos, V., Sanz, M., Herrera, D., 2024. Supra- and Subgingival Microbiome in Gingivitis and Impact of Biofilm Control: A Comprehensive Review. Antibiotics (Basel) 13, 571. https://doi.org/10.3390/antibiotics13060571

Jandhyala, S.M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M., Reddy, D.N., 2015. Role of the normal gut microbiota. World J Gastroenterol 21, 8787–8803. https://doi.org/10.3748/wjg.v21.i29.8787

Javeed, A., Dallora, A.L., Berglund, J.S., Anderberg, P., 2022. An Intelligent Learning System for Unbiased Prediction of Dementia Based on Autoencoder and Adaboost Ensemble Learning. Life (Basel) 12, 1097. https://doi.org/10.3390/life12071097

Jia, G., Zhi, A., Lai, P.F.H., Wang, G., Xia, Y., Xiong, Z., Zhang, H., Che, N., Ai, L., 2018. The oral microbiota – a mechanistic role for systemic diseases. Br Dent J 224, 447–455. https://doi.org/10.1038/sj.bdj.2018.217

Jiang, S., Zha, Y., Zhao, T., Jin, X., Zhu, R., Wei, S., Wang, R., Song, Y., Li, L., Lyu, J., Hu, W., Zhang, D., Wang, M., Zhang, Y., 2023. Antimicrobial peptide temporin derivatives inhibit biofilm formation and virulence factor expression of Streptococcus mutans. Front Microbiol 14, 1267389. https://doi.org/10.3389/fmicb.2023.1267389

Johnson, J.M., Khoshgoftaar, T.M., 2019. Survey on deep learning with class imbalance. Journal of Big Data 6, 27. https://doi.org/10.1186/s40537-019-0192-5

Joseph, S., Aduse-Opoku, J., Hashim, A., Hanski, E., Streich, R., Knowles, S.C.L., Pedersen, A.B., Wade, W.G., Curtis, M.A., 2021. A 16S rRNA Gene and Draft Genome Database for the Murine Oral Bacterial Community. mSystems 6, 10.1128/msystems.01222-20. https://doi.org/10.1128/msystems.01222-20

Kageyama, S., Hama, S., Furuta, M., Asakawa, M., Kawano, S., Ninomiya, T., Takeshita, T., 2025. Performance of salivary microbiota in detecting periodontitis using a machine learning approach. Front. Cell. Infect. Microbiol. 15. https://doi.org/10.3389/fcimb.2025.1631798

Kåhrström, C.T., 2013. Mismatch excels when ampicillin runs low. Nat Rev Microbiol 11, 299–299. https://doi.org/10.1038/nrmicro3017

Kalpana, B., Prabhu, P., Bhat, A.H., Senthilkumar, A., Arun, R.P., Asokan, S., Gunthe, S.S., Verma, R.S., 2020. Bacterial diversity and functional analysis of severe early childhood caries and recurrence in India. Sci Rep 10, 21248. https://doi.org/10.1038/s41598-020-78057-z

Kans, J., 2025. Entrez Direct: E-utilities on the Unix Command Line, in: Entrez Programming Utilities Help [Internet]. National Center for Biotechnology Information (US).

Kasof, G.M., Lu, J.J., Liu, D., Speer, B., Mongan, K.N., Gomes, B.C., Lorenzi, M.V., 2001. Tumor necrosis factor-alpha induces the expression of DR6, a member of the TNF receptor family, through activation of NF-kappaB. Oncogene 20, 7965–7975. https://doi.org/10.1038/sj.onc.1204985

Katz, K., Shutov, O., Lapoint, R., Kimelman, M., Brister, J.R., O'Sullivan, C., 2022. The Sequence Read Archive: a decade more of explosive growth. Nucleic Acids Res 50, D387–D390. https://doi.org/10.1093/nar/gkab1053

Khelaifia, S., Virginie, P., Belkacemi, S., Tassery, H., Terrer, E., Aboudharam, G., 2023. Culturing the Human Oral Microbiota, Updating Methodologies and Cultivation Techniques. Microorganisms 11, 836. https://doi.org/10.3390/microorganisms11040836

Kilian, M., Chapple, I.L.C., Hannig, M., Marsh, P.D., Meuric, V., Pedersen, A.M.L., Tonetti, M.S., Wade, W.G., Zaura, E., 2016. The oral microbiome – an update for oral healthcare professionals. Br Dent J 221, 657–666. https://doi.org/10.1038/sj.bdj.2016.865

Kim, E.-H., Kim, S., Kim, H.-J., Jeong, H., Lee, J., Jang, J., Joo, J.-Y., Shin, Y., Kang, J., Park, A.K., Lee, J.-Y., Lee, S., 2020. Prediction of Chronic Periodontitis Severity Using Machine Learning Models Based On Salivary Bacterial Copy Number. Front Cell Infect Microbiol 10, 571515. https://doi.org/10.3389/fcimb.2020.571515

Kim, L., Lavrinienko, A., Sebechlebska, Z., Stoltenberg, S., Bokulich, N.A., 2025. Tier-based standards for FAIR sequence data and metadata sharing in microbiome research. https://doi.org/10.1101/2025.02.06.636914

Kim, R., 2023. Advanced Organotypic In Vitro Model Systems for Host–Microbial Coculture. Biochip J 1–27. https://doi.org/10.1007/s13206-023-00103-5

Kingma, D.P., Ba, J., 2017. Adam: A Method for Stochastic Optimization. https://doi.org/10.48550/arXiv.1412.6980

Kodikara, S., Ellul, S., Lê Cao, K.-A., 2022. Statistical challenges in longitudinal microbiome data analysis. Brief Bioinform 23, bbac273. https://doi.org/10.1093/bib/bbac273

Kolenbrander, P.E., Andersen, R.N., Blehert, D.S., Egland, P.G., Foster, J.S., Palmer, R.J., 2002. Communication among Oral Bacteria. Microbiol Mol Biol Rev 66, 486–505. https://doi.org/10.1128/MMBR.66.3.486-505.2002

Könönen, E., Fteita, D., Gursoy, U.K., Gursoy, M., 2022. Prevotella species as oral residents and infectious agents with potential impact on systemic conditions. J Oral Microbiol 14, 2079814. https://doi.org/10.1080/20002297.2022.2079814

Koo, H., Falsetta, M.L., Klein, M.I., 2013. The Exopolysaccharide Matrix. J Dent Res 92, 1065–1073. https://doi.org/10.1177/0022034513504218

Koo, H., Xiao, J., Klein, M.I., Jeon, J.G., 2010. Exopolysaccharides Produced by Streptococcus mutans Glucosyltransferases Modulate the Establishment of Microcolonies within Multispecies Biofilms. J Bacteriol 192, 3024–3032. https://doi.org/10.1128/JB.01649-09

Krawczyk, B., 2016. Learning from imbalanced data: open challenges and future directions. Prog Artif Intell 5, 221–232. https://doi.org/10.1007/s13748-016-0094-0

Kriaa, A., Mariaule, V., De Rudder, C., Jablaoui, A., Sokol, H., Wilmes, P., Maguin, E., Rhimi, M., 2024. From animal models to gut-on-chip: the challenging journey to capture inter-individual variability in chronic digestive disorders. Gut Microbes 16, 2333434. https://doi.org/10.1080/19490976.2024.2333434

Krieger, M., AbdelRahman, Y.M., Choi, D., Palmer, E.A., Yoo, A., McGuire, S., Kreth, J., Merritt, J., 2024. Stratification of Fusobacterium nucleatum by local health status in the oral cavity defines its subspecies disease association. Cell Host Microbe 32, 479-488.e4. https://doi.org/10.1016/j.chom.2024.02.010

Krishnan, K., Chen, T., Paster, B., 2017. A practical guide to the oral microbiome and its relation to health and disease. Oral Dis 23, 276–286. https://doi.org/10.1111/odi.12509

Kunath, B.J., De Rudder, C., Laczny, C.C., Letellier, E., Wilmes, P., 2024. The oral–gut microbiome axis in health and disease. Nat Rev Microbiol 22, 791–805. https://doi.org/10.1038/s41579-024-01075-5

Kurtzman, G.M., Horowitz, R.A., Johnson, R., Prestiano, R.A., Klein, B.I., 2022. The systemic oral health connection: Biofilms. Medicine (Baltimore) 101, e30517. https://doi.org/10.1097/MD.0000000000030517

Lagier, J.-C., Edouard, S., Pagnier, I., Mediannikov, O., Drancourt, M., Raoult, D., 2015. Current and Past Strategies for Bacterial Culture in Clinical Microbiology. Clin Microbiol Rev 28, 208–236. https://doi.org/10.1128/CMR.00110-14

Lallier, T.E., Yukna, R., St Marie, S., Moses, R., 2001. The putative collagen binding peptide hastens periodontal ligament cell attachment to bone replacement graft materials. J Periodontol 72, 990–997. https://doi.org/10.1902/jop.2001.72.8.990

Lamont, R.J., Koo, H., Hajishengallis, G., 2018. The oral microbiota: dynamic communities and host interactions. Nat Rev Microbiol 16, 745–759. https://doi.org/10.1038/s41579-018-0089-x

Lao, J., Guédon, G., Lacroix, T., Charron-Bourgoin, F., Libante, V., Loux, V., Chiapello, H., Payot, S., Leblond-Bourget, N., 2020. Abundance, Diversity and Role of ICEs and IMEs in the Adaptation of Streptococcus salivarius to the Environment. Genes (Basel) 11, 999. https://doi.org/10.3390/genes11090999

Larsson, D.G.J., Flach, C.-F., 2022. Antibiotic resistance in the environment. Nat Rev Microbiol 20, 257–269. https://doi.org/10.1038/s41579-021-00649-x

Lăzureanu, P.C., Popescu, F., Tudor, A., Stef, L., Negru, A.G., Mihăilă, R., 2021. Saliva pH and Flow Rate in Patients with Periodontal Disease and Associated Cardiovascular Disease. Med Sci Monit 27, e931362-1-e931362-13. https://doi.org/10.12659/MSM.931362

Lee, E., Priutt, E., Woods, S., Quick, A., King, S., McLellan, L.K., Shields, R.C., 2024. Genomic analysis of conjugative and chromosomally integrated mobile genetic elements in oral streptococci. Appl Environ Microbiol 90, e0136024. https://doi.org/10.1128/aem.01360-24

Lemos, J.A., Quivey, R.G., Koo, H., Abranches, J., 2013. Streptococcus mutans: a new Gram-positive paradigm? Microbiology (Reading) 159, 436–445. https://doi.org/10.1099/mic.0.066134-0

Lerminiaux, N.A., Cameron, A.D.S., 2019. Horizontal transfer of antibiotic resistance genes in clinical environments. Can J Microbiol 65, 34–44. https://doi.org/10.1139/cjm-2018-0275

Letzelter, C., Croute, F., Pianezzi, B., Roques, C., Soleilhavoup, J.P., 1998. Supernatant cytotoxicity and proteolytic activity of selected oral bacteria against human gingival fibroblasts in vitro. Arch Oral Biol 43, 15–23. https://doi.org/10.1016/s0003-9969(97)00095-2

Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, J., Helmerhorst, E.J., Leone, C.W., Troxler, R.F., Yaskell, T., Haffajee, A.D., Socransky, S.S., Oppenheim, F.G., 2004. Identification of early microbial colonizers in human dental biofilm. J Appl Microbiol 97, 1311–1318. https://doi.org/10.1111/j.1365-2672.2004.02420.x

Li, J., Zhao, Y.-N., Wang, Y.-H., Chen, Y., Hou, J.-L., Wang, D.-Y., Shi, L., Shen, J., 2025. Animal models of oral infectious diseases. Front Oral Health 6, 1571492. https://doi.org/10.3389/froh.2025.1571492

Li, X., Brejnrod, A., Thorsen, J., Zachariasen, T., Trivedi, U., Russel, J., Vestergaard, G.A., Stokholm, J., Rasmussen, M.A., Sørensen, S.J., 2023. Differential responses of the gut microbiome and resistome to antibiotic exposures in infants and adults. Nat Commun 14, 8526. https://doi.org/10.1038/s41467-023-44289-6

Li, X., Kolltveit, K.M., Tronstad, L., Olsen, I., 2000. Systemic Diseases Caused by Oral Infection. Clin Microbiol Rev 13, 547–558. https://doi.org/10.1128/cmr.13.4.547-558.2000

Li, X., Liu, Y., Yang, X., Li, C., Song, Z., 2022. The Oral Microbiota: Community Composition, Influencing Factors, Pathogenesis, and Interventions. Front Microbiol 13, 895537. https://doi.org/10.3389/fmicb.2022.895537

Linden, S.K., Sutton, P., Karlsson, N.G., Korolik, V., McGuckin, M.A., 2008. Mucins in the mucosal barrier to infection. Mucosal Immunol 1, 183–197. https://doi.org/10.1038/mi.2008.5

Liu, Y., Tang, H., Lin, Z., Xu, P., 2015. Mechanisms of acid tolerance in bacteria and prospects in biotechnology and bioremediation. Biotechnol Adv 33, 1484–1492. https://doi.org/10.1016/j.biotechadv.2015.06.001

Lockhart, P.B., Brennan, M.T., Sasser, H.C., Fox, P.C., Paster, B.J., Bahrani-Mougeot, F.K., 2008. Bacteremia Associated with Tooth Brushing and Dental Extraction. Circulation 117, 3118–3125. https://doi.org/10.1161/CIRCULATIONAHA.107.758524

Longman, L.P., Pearce, P.K., McGowan, P., Hardy, P., Martin, M.V., 1991. Antibiotic-resistant oral streptococci in dental patients susceptible to infective endocarditis. J Med Microbiol 34, 33–37. https://doi.org/10.1099/00222615-34-1-33

Loos, B.G., Dyer, D.W., Whittam, T.S., Selander, R.K., 1993. Genetic structure of populations of Porphyromonas gingivalis associated with periodontitis and other oral infections. Infect Immun 61, 204–212. https://doi.org/10.1128/iai.61.1.204-212.1993

López, E., Blázquez, J., 2009. Effect of Subinhibitory Concentrations of Antibiotics on Intrachromosomal Homologous Recombination in Escherichia coli. Antimicrob Agents Chemother 53, 3411–3415. https://doi.org/10.1128/AAC.00358-09

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. https://doi.org/10.48550/arXiv.1705.07874

Lunde, T.M., Hjerde, E., Al-Haroni, M., 2021. Prevalence, diversity and transferability of the Tn916-Tn1545 family ICE in oral streptococci. J Oral Microbiol 13, 1896874. https://doi.org/10.1080/20002297.2021.1896874

Lynge Pedersen, A.M., Belstrøm, D., 2019. The role of natural salivary defences in maintaining a healthy oral microbiota. J Dent 80 Suppl 1, S3–S12. https://doi.org/10.1016/j.jdent.2018.08.010

Ma, Y., Jiang, Z., Wang, Y., Pan, L., Liu, K., Xia, R., Yuan, L., Cheng, X., 2025. Tongue coating microbiota-based machine learning for diagnosing digestive system tumours. J Oral Microbiol 17, 2487645. https://doi.org/10.1080/20002297.2025.2487645

Madsen, J.S., Burmølle, M., Hansen, L.H., Sørensen, S.J., 2012. The interconnection between biofilm formation and horizontal gene transfer. FEMS Immunol Med Microbiol 65, 183–195. https://doi.org/10.1111/j.1574-695X.2012.00960.x

Mahmood, M.K., Fatih, M.T., Kurda, H.A., Mahmood, N.K., Shareef, F.U., Faraidun, H., Tassery, H., Tardivo, D., Lan, R., Noori, Z.F., Qadir, B.H., Hassan, A.D., 2024. Role of viruses in periodontitis: An extensive review of herpesviruses, human immunodeficiency virus, coronavirus-19, papillomavirus and hepatitis viruses. World J Virol 13, 99070. https://doi.org/10.5501/wjv.v13.i4.99070

Maier, T., 2023. Oral Microbiome in Health and Disease: Maintaining a Healthy, Balanced Ecosystem and Reversing Dysbiosis. Microorganisms 11, 1453. https://doi.org/10.3390/microorganisms11061453

Mainjot, A., D'Hoore, W., Vanheusden, A., Van Nieuwenhuysen, J.-P., 2009. Antibiotic prescribing in dental practice in Belgium. International Endodontic Journal 42, 1112–1117. https://doi.org/10.1111/j.1365-2591.2009.01642.x

Maitra, A., Dill, K.A., 2015. Bacterial growth laws reflect the evolutionary importance of energy efficiency. Proc Natl Acad Sci U S A 112, 406–411. https://doi.org/10.1073/pnas.1421138111

Mark Welch, J.L., Ramírez-Puebla, S.T., Borisy, G.G., 2020. Oral Microbiome Geography: Micron-Scale Habitat and Niche. Cell Host Microbe 28, 160–168. https://doi.org/10.1016/j.chom.2020.07.009

Marsh, P.D., 2003. Are dental diseases examples of ecological catastrophes? Microbiology (Reading) 149, 279–294. https://doi.org/10.1099/mic.0.26082-0

Marsh, P.D., 1999. Microbiologic aspects of dental plaque and dental caries. Dent Clin North Am 43, 599–614, v–vi.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12. https://doi.org/10.14806/ej.17.1.200

Martínez, A., Kuraji, R., Kapila, Y.L., 2021. The human oral virome: Shedding light on the dark matter. Periodontol 2000 87, 282–298. https://doi.org/10.1111/prd.12396

Mathew, J., Kshirsagar, R., Abidin, D.Z., Griffin, J., Kanarachos, S., James, J., Alamaniotis, M., Fitzpatrick, M.E., 2023. A comparison of machine learning methods to classify radioactive elements using prompt-gamma-ray neutron activation data. Sci Rep 13, 9948. https://doi.org/10.1038/s41598-023-36832-8

Matsui, R., Cvitkovitch, D., 2010. Acid tolerance mechanisms utilized by Streptococcus mutans. Future Microbiol 5, 403–417. https://doi.org/10.2217/fmb.09.129

Maxim, L.D., Niebo, R., Utell, M.J., 2014. Screening tests: a review with examples. Inhal Toxicol 26, 811–828. https://doi.org/10.3109/08958378.2014.955932

May, A., Brandt, B.W., El-Kebir, M., Klau, G.W., Zaura, E., Crielaard, W., Heringa, J., Abeln, S., 2016. metaModules identifies key functional subnetworks in microbiome-related disease. Bioinformatics 32, 1678–1685. https://doi.org/10.1093/bioinformatics/btv526

McDaniel, J., McDaniel, S., Samiano, B.J., Marrujo, M., Kingsley, K., Howard, K.M., 2021. Microbial Screening Reveals Oral Site-Specific Locations of the Periodontal Pathogen Selenomonas noxia. Curr Issues Mol Biol 43, 353–364. https://doi.org/10.3390/cimb43010029

McMurdie, P.J., Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8, e61217. https://doi.org/10.1371/journal.pone.0061217

Michaelis, C., Grohmann, E., 2023. Horizontal Gene Transfer of Antibiotic Resistance Genes in Biofilms. Antibiotics (Basel) 12, 328. https://doi.org/10.3390/antibiotics12020328

Mohanty, R., Asopa, S.J., Joseph, M.D., Singh, B., Rajguru, J.P., Saidath, K., Sharma, U., 2019. Red complex: Polymicrobial conglomerate in oral flora: A review. J Family Med Prim Care 8, 3480–3486. https://doi.org/10.4103/jfmpc.jfmpc_759_19

Montelongo-Jauregui, D., Lopez-Ribot, J.L., 2018. Candida Interactions with the Oral Bacterial Microbiota. J Fungi (Basel) 4, 122. https://doi.org/10.3390/jof4040122

Moradali, M.F., Davey, M.E., 2021. Metabolic plasticity enables lifestyle transitions of Porphyromonas gingivalis. NPJ Biofilms Microbiomes 7, 46. https://doi.org/10.1038/s41522-021-00217-4

Moradi, J., Berggreen, E., Bunæs, D.F., Bolstad, A.I., Bertelsen, R.J., 2025. Microbiome composition and metabolic pathways in shallow and deep periodontal pockets. Sci Rep 15, 12926. https://doi.org/10.1038/s41598-025-97531-0

Mountcastle, S.E., Cox, S.C., Sammons, R.L., Jabbari, S., Shelton, R.M., Kuehne, S.A., 2020. A review of co-culture models to study the oral microenvironment and disease. J Oral Microbiol 12, 1773122. https://doi.org/10.1080/20002297.2020.1773122

Moussa, L.G., Sonmez, O.F., Bedi, R., 2025. Incorporating oral health into the One Health agenda: contributions and implications. Science in One Health 4, 100111. https://doi.org/10.1016/j.soh.2025.100111

Munita, J.M., Arias, C.A., 2016. Mechanisms of Antibiotic Resistance. Microbiol Spectr 4. https://doi.org/10.1128/microbiolspec.VMBF-0016-2015

Naginyte, M., Do, T., Meade, J., Devine, D.A., Marsh, P.D., 2019. Enrichment of periodontal pathogens from the biofilms of healthy adults. Sci Rep 9, 5491. https://doi.org/10.1038/s41598-019-41882-y

Nakazawa, F., Poco, S.E., Ikeda, T., Sato, M., Kalfas, S., Sundqvist, G., Hoshino, E., 1999. Cryptobacterium curtum gen. nov., sp. nov., a new genus of gram-positive anaerobic rod isolated from human oral cavities. Int J Syst Bacteriol 49 Pt 3, 1193–1200. https://doi.org/10.1099/00207713-49-3-1193

Nativel, B., Couret, D., Giraud, P., Meilhac, O., d'Hellencourt, C.L., Viranaïcken, W., Da Silva, C.R., 2017. Porphyromonas gingivalis lipopolysaccharides act exclusively through TLR4 with a resilience between mouse and human. Sci Rep 7, 15789. https://doi.org/10.1038/s41598-017-16190-y

Naureen, Z., Dautaj, A., Anpilogov, K., Camilleri, G., Dhuli, K., Tanzi, B., Maltese, P.E., Cristofoli, F., De Antoni, L., Beccari, T., Dundar, M., Bertelli, M., 2020. Bacteriophages presence in nature and their role in the natural selection of bacterial populations. Acta Biomed 91, e2020024. https://doi.org/10.23750/abm.v91i13-S.10819

Nearing, J.T., Comeau, A.M., Langille, M.G.I., 2021. Identifying biases and their potential solutions in human microbiome studies. Microbiome 9, 113. https://doi.org/10.1186/s40168-021-01059-0

Neilands, J., Davies, J.R., Bikker, F.J., Svensäter, G., 2019. Parvimonas micra stimulates expression of gingipains from Porphyromonas gingivalis in multi-species communities. Anaerobe 55, 54–60. https://doi.org/10.1016/j.anaerobe.2018.10.007

Nemoto, T.K., Ohara Nemoto, Y., 2021. Dipeptidyl-peptidases: Key enzymes producing entry forms of extracellular proteins in asaccharolytic periodontopathic bacterium Porphyromonas gingivalis. Mol Oral Microbiol 36, 145–156. https://doi.org/10.1111/omi.12317

Nguyen, T.A., Kim, H.Y., Stocker, S., Kidd, S., Alastruey-Izquierdo, A., Dao, A., Harrison, T., Wahyuningsih, R., Rickerts, V., Perfect, J., Denning, D.W., Nucci, M., Cassini, A., Beardsley, J., Gigante, V., Sati, H., Morrissey, C.O., Alffenaar, J.-W., 2024. Pichia kudriavzevii (Candida krusei): A systematic review to inform the World Health Organisation priority list of fungal pathogens. Med Mycol 62, myad132. https://doi.org/10.1093/mmy/myad132

Novielli, P., Romano, D., Magarelli, M., Bitonto, P.D., Diacono, D., Chiatante, A., Lopalco, G., Sabella, D., Venerito, V., Filannino, P., Bellotti, R., De Angelis, M., Iannone, F., Tangaro, S., 2024. Explainable artificial intelligence for microbiome data analysis in colorectal cancer biomarker identification. Front. Microbiol. 15. https://doi.org/10.3389/fmicb.2024.1348974

Oh, J., Kim, S., Lee, C., Cha, J.-H., Yang, S.Y., Im, S.G., Park, C., Jang, B.C., Choi, S.-Y., 2023. Preventing Vanishing Gradient Problem of Hardware Neuromorphic System by Implementing Imidazole-Based Memristive ReLU Activation Neuron. Adv Mater 35, e2300023. https://doi.org/10.1002/adma.202300023

Olivares Pacheco, J.A., Bernardini, A., Garcia-Leon, G., Corona, F., Sanchez, M.B., Martinez, J.L., 2013. The intrinsic resistome of bacterial pathogens. Front. Microbiol. 4. https://doi.org/10.3389/fmicb.2013.00103

Oscarsson, J., Claesson, R., Lindholm, M., Höglund Åberg, C., Johansson, A., 2019. Tools of Aggregatibacter actinomycetemcomitans to Evade the Host Response. J Clin Med 8, 1079. https://doi.org/10.3390/jcm8071079

Pais, R.J., Botelho, J., Machado, V., Alcoforado, G., Mendes, J.J., Alves, R., Bessa, L.J., 2025. Exploring AI-Driven Machine Learning Approaches for Optimal Classification of Peri-Implantitis Based on Oral Microbiome Data: A Feasibility Study. Diagnostics (Basel) 15, 425. https://doi.org/10.3390/diagnostics15040425

Papoutsoglou, G., Tarazona, S., Lopes, M.B., Klammsteiner, T., Ibrahimi, E., Eckenberger, J., Novielli, P., Tonda, A., Simeon, A., Shigdel, R., Béreux, S., Vitali, G., Tangaro, S., Lahti, L., Temko, A., Claesson, M.J., Berland, M., 2023. Machine learning approaches in microbiome research: challenges and best practices. Front Microbiol 14, 1261889. https://doi.org/10.3389/fmicb.2023.1261889

Papp, M., Solymosi, N., 2022. Review and Comparison of Antimicrobial Resistance Gene Databases. Antibiotics (Basel) 11, 339. https://doi.org/10.3390/antibiotics11030339

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. https://doi.org/10.48550/arXiv.1912.01703

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Peng, L., Wang, L., Chen, L., Shen, Z., 2025. Machine Learning-Driven Prediction Models for Brodalumab Therapeutic Effect and Response Speed in Plaque Psoriasis. Psoriasis (Auckl) 15, 429–442. https://doi.org/10.2147/PTT.S531925

Peng, X., Cheng, L., You, Y., Tang, C., Ren, B., Li, Y., Xu, X., Zhou, X., 2022. Oral microbiota in human systematic diseases. Int J Oral Sci 14, 1–11. https://doi.org/10.1038/s41368-022-00163-7

Perreten, V., Boerlin, P., 2003. A new sulfonamide resistance gene (sul3) in Escherichia coli is widespread in the pig population of Switzerland. Antimicrob Agents Chemother 47, 1169–1172. https://doi.org/10.1128/AAC.47.3.1169-1172.2003

Postma, P.W., Lengeler, J.W., Jacobson, G.R., 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol Rev 57, 543–594. https://doi.org/10.1128/mr.57.3.543-594.1993

Powis, G., Meuillet, E.J., Indarte, M., Booher, G., Kirkpatrick, L., 2023. Pleckstrin Homology [PH] domain, structure, mechanism, and contribution to human disease. Biomed Pharmacother 165, 115024. https://doi.org/10.1016/j.biopha.2023.115024

Prabhu Matondkar, S., Yavagal, C., Kugaji, M., Bhat, K.G., 2020. Quantitative assessment of Scardovia wiggsiae from dental plaque samples of children suffering from severe early childhood caries and caries free children. Anaerobe 62, 102110. https://doi.org/10.1016/j.anaerobe.2019.102110

Przymus, P., Rykaczewski, K., Martín-Segura, A., Truu, J., Carrillo De Santa Pau, E., Kolev, M., Naskinova, I., Gruca, A., Sampri, A., Frohme, M., Nechyporenko, A., 2025. Deep learning in microbiome analysis: a comprehensive review of neural network models. Front. Microbiol. 15, 1516667. https://doi.org/10.3389/fmicb.2024.1516667

Puthenveetil, R., Kumar, S., Caimano, M.J., Dey, A., Anand, A., Vinogradova, O., Radolf, J.D., 2017. The major outer sheath protein forms distinct conformers and multimeric complexes in the outer membrane and periplasm of Treponema denticola. Sci Rep 7, 13260. https://doi.org/10.1038/s41598-017-13550-6

Radaic, A., Kapila, Y.L., 2021. The oralome and its dysbiosis: New insights into oral microbiome-host interactions. Comput Struct Biotechnol J 19, 1335–1360. https://doi.org/10.1016/j.csbj.2021.02.010

Rajasekaran, J.J., Krishnamurthy, H.K., Bosco, J., Jayaraman, V., Krishna, K., Wang, T., Bei, K., 2024. Oral Microbiome: A Review of Its Impact on Oral and Systemic Health. Microorganisms 12, 1797. https://doi.org/10.3390/microorganisms12091797

Rams, T.E., Sautter, J.D., van Winkelhoff, A.J., 2020. Antibiotic Resistance of Human Periodontal Pathogen Parvimonas micra Over 10 Years. Antibiotics (Basel) 9, 709. https://doi.org/10.3390/antibiotics9100709

Rath, S., Bal, S.C.B., Dubey, D., 2021. Oral Biofilm: Development Mechanism, Multidrug Resistance, and Their Effective Management with Novel Techniques. Rambam Maimonides Medical Journal 12, e0004. https://doi.org/10.5041/RMMJ.10428

Rathee, M., Sapra, A., 2025. Dental Caries, in: StatPearls. StatPearls Publishing, Treasure Island (FL).

Renson, A., Jones, H.E., Beghini, F., Segata, N., Zolnik, C.P., Usyk, M., Moody, T.U., Thorpe, L., Burk, R., Waldron, L., Dowd, J.B., 2019. Sociodemographic variation in the oral microbiome. Ann Epidemiol 35, 73-80.e2. https://doi.org/10.1016/j.annepidem.2019.03.006

Reygaert, W.C., 2018. An overview of the antimicrobial resistance mechanisms of bacteria. AIMS Microbiol 4, 482–501. https://doi.org/10.3934/microbiol.2018.3.482

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. https://doi.org/10.48550/arXiv.1602.04938

Roberts, M.C., 1996. Tetracycline resistance determinants: mechanisms of action, regulation of expression, genetic mobility, and distribution. FEMS Microbiol Rev 19, 1–24. https://doi.org/10.1111/j.1574-6976.1996.tb00251.x

Roca, C., Alkhateeb, A.A., Deanhardt, B.K., Macdonald, J.K., Chi, D.L., Wang, J.R., Wolfgang, M.C., 2024. Saliva sampling method influences oral microbiome composition and taxa distribution associated with oral diseases. PLoS One 19, e0301016. https://doi.org/10.1371/journal.pone.0301016

Rosier, B.T., De Jager, M., Zaura, E., Krom, B.P., 2014. Historical and contemporary hypotheses on the development of oral diseases: are we there yet? Front Cell Infect Microbiol 4, 92. https://doi.org/10.3389/fcimb.2014.00092

Rösing, C.K., Randall, C., Giacaman, R.A., 2023. Editorial: Dental caries and periodontal diseases as non-communicable chronic diseases. Front Oral Health 3, 1113029. https://doi.org/10.3389/froh.2022.1113029

Ruiz-Mojica, C.A., Brizuela, M., 2025. Viral Infections of the Oral Mucosa, in: StatPearls. StatPearls Publishing, Treasure Island (FL).

Ryan, M.P., Pembroke, J.T., Adley, C.C., 2006. Ralstonia pickettii: a persistent gram-negative nosocomial infectious organism. J Hosp Infect 62, 278–284. https://doi.org/10.1016/j.jhin.2005.08.015

Saarela, M., Podgorelec, V., 2024. Recent Applications of Explainable AI (XAI): A Systematic Literature Review. Applied Sciences 14, 8884. https://doi.org/10.3390/app14198884

Saeys, Y., Inza, I., Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. Bioinformatics 23, 2507–2517. https://doi.org/10.1093/bioinformatics/btm344

Šakarnytė, L., Šiugždinienė, R., Žymantienė, J., Ruzauskas, M., 2023. Comparison of Oral Microbial Composition and Determinants Encoding Antimicrobial Resistance in Dogs and Their Owners. Antibiotics (Basel) 12, 1554. https://doi.org/10.3390/antibiotics12101554

Salam, M.A., Al-Amin, M.Y., Salam, M.T., Pawar, J.S., Akhter, N., Rabaan, A.A., Alqumber, M.A.A., 2023. Antimicrobial Resistance: A Growing Serious Threat for Global Public Health. Healthcare 11, 1946. https://doi.org/10.3390/healthcare11131946

Sánchez, M.C., Llama-Palacios, A., Blanc, V., León, R., Herrera, D., Sanz, M., 2011. Structure, viability and bacterial kinetics of an in vitro biofilm model using six bacteria from the subgingival microbiota. J Periodontal Res 46, 252–260. https://doi.org/10.1111/j.1600-0765.2010.01341.x

Sandholm, L., 1986. Proteases and their inhibitors in chronic inflammatory periodontal disease. J Clin Periodontol 13, 19–26. https://doi.org/10.1111/j.1600-051x.1986.tb01409.x

Santacroce, L., Passarelli, P.C., Azzolino, D., Bottalico, L., Charitos, I.A., Cazzolla, A.P., Colella, M., Topi, S., Godoy, F.G., D'Addona, A., 2023. Oral microbiota in human health and disease: A perspective. Exp Biol Med (Maywood) 248, 1288–1301. https://doi.org/10.1177/15353702231187645

Santonocito, S., Giudice, A., Polizzi, A., Troiano, G., Merlo, E.M., Sclafani, R., Grosso, G., Isola, G., 2022. A Cross-Talk between Diet and the Oral Microbiome: Balance of Nutrition on Inflammation and Immune System's Response during Periodontitis. Nutrients 14, 2426. https://doi.org/10.3390/nu14122426

Schmidt, T.S., Hayward, M.R., Coelho, L.P., Li, S.S., Costea, P.I., Voigt, A.Y., Wirbel, J., Maistrenko, O.M., Alves, R.J., Bergsten, E., de Beaufort, C., Sobhani, I., Heintz-Buschart, A., Sunagawa, S., Zeller, G., Wilmes, P., Bork, P., 2019. Extensive transmission of microbes along the gastrointestinal tract. eLife 8, e42693. https://doi.org/10.7554/eLife.42693

Sedghi, L.M., Bacino, M., Kapila, Y.L., 2021. Periodontal Disease: The Good, The Bad, and The Unknown. Front Cell Infect Microbiol 11, 766944. https://doi.org/10.3389/fcimb.2021.766944

Şenel, S., 2021. An Overview of Physical, Microbiological and Immune Barriers of Oral Mucosa. Int J Mol Sci 22, 7821. https://doi.org/10.3390/ijms22157821

Seymour, R.A., Heasman, P.A., 1995. Tetracyclines in the management of periodontal diseases. A review. J Clin Periodontol 22, 22–35. https://doi.org/10.1111/j.1600-051x.1995.tb01767.x

Sharif, F., 2012. Antimicrobial prescribing for general dental practitioners. Br Dent J 213, 484–484. https://doi.org/10.1038/sj.bdj.2012.1035

Shi, Z., Gewirtz, A.T., 2018. Together Forever: Bacterial–Viral Interactions in Infection and Immunity. Viruses 10, 122. https://doi.org/10.3390/v10030122

Shi, Z., Ren, M., Rockey, D.C., 2020. Myocardin and myocardin-related transcription factor-A synergistically mediate actin cytoskeletal-dependent inhibition of liver fibrogenesis. Am J Physiol Gastrointest Liver Physiol 318, G504–G517. https://doi.org/10.1152/ajpgi.00302.2019

Siebor, E., de Curraize, C., Varin, V., Magallon, A., Neuwirth, C., 2021. Mobilisation of plasmid-mediated blaVEB-1 gene cassette into distinct genomic islands of Proteus mirabilis after ceftazidime exposure. J Glob Antimicrob Resist 27, 26–30. https://doi.org/10.1016/j.jgar.2021.07.011

Silva, H., 2021. Tobacco Use and Periodontal Disease—The Role of Microvascular Dysfunction. Biology (Basel) 10, 441. https://doi.org/10.3390/biology10050441

SILVA, N., ABUSLEME, L., BRAVO, D., DUTZAN, N., GARCIA-SESNICH, J., VERNAL, R., HERNÁNDEZ, M., GAMONAL, J., 2015. Host response mechanisms in periodontal diseases. J Appl Oral Sci 23, 329–355. https://doi.org/10.1590/1678-775720140259

Smith, K.R., Penzes, P., 2018. Ankyrins: roles in synaptic biology and pathology. Mol Cell Neurosci 91, 131–139. https://doi.org/10.1016/j.mcn.2018.04.010

Solbiati, J., Frias-Lopez, J., 2018. Metatranscriptome of the Oral Microbiome in Health and Disease. J Dent Res 97, 492–500. https://doi.org/10.1177/0022034518761644

Spatafora, G., Li, Y., He, X., Cowan, A., Tanner, A.C.R., 2024. The Evolving Microbiome of Dental Caries. Microorganisms 12, 121. https://doi.org/10.3390/microorganisms12010121

Spigaglia, P., Barbanti, F., Mastrantonio, P., 2005. Horizontal Transfer of Erythromycin Resistance from Clostridium difficile to Butyrivibrio fibrisolvens. Antimicrob Agents Chemother 49, 5142–5145. https://doi.org/10.1128/AAC.49.12.5142-5145.2005

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 15, 1929–1958.

Stebliankin, V., Sazal, M., Valdes, C., Mathee, K., Narasimhan, G., 2022. A novel approach for combining the metagenome, metaresistome, metareplicome and causal inference to determine the microbes and their antibiotic resistance gene repertoire that contribute to dysbiosis. Microb Genom 8, mgen000899. https://doi.org/10.1099/mgen.0.000899

Sukumar, S., Rahmanyar, Z., El Jurf, H.Q., Akil, W.S., Hussain, J., Martin, F.E., Ekanayake, K., Martinez, E., 2024. Mapping the oral resistome: a systematic review. J Med Microbiol 73, 001866. https://doi.org/10.1099/jmm.0.001866

Sukumar, S., Wang, F., Simpson, C.A., Willet, C.E., Chew, T., Hughes, T.E., Bockmann, M.R., Sadsad, R., Martin, F.E., Lydecker, H.W., Browne, G.V., Davis, K.M., Bui, M., Martinez, E., Adler, C.J., 2023. Development of the oral resistome during the first decade of life. Nat Commun 14, 1291. https://doi.org/10.1038/s41467-023-36781-w

Suzuki, N., Yoneda, M., Hirofuji, T., 2013. Mixed Red-Complex Bacterial Infection in Periodontitis. Int J Dent 2013, 587279. https://doi.org/10.1155/2013/587279

Takahashi, N., Nyvad, B., 2011. The role of bacteria in the caries process: ecological perspectives. J Dent Res 90, 294–303. https://doi.org/10.1177/0022034510379602

Takeuchi, H., Nakamura, E., Yamaga, S., Amano, A., 2022. Porphyromonas gingivalis Infection Induces Lipopolysaccharide and Peptidoglycan Penetration Through Gingival Epithelium. Front Oral Health 3, 845002. https://doi.org/10.3389/froh.2022.845002

Tan, X., Wang, Y., Gong, T., 2023. The interplay between oral microbiota, gut microbiota and systematic diseases. J Oral Microbiol 15, 2213112. https://doi.org/10.1080/20002297.2023.2213112

Tang, K.W.K., Millar, B.C., Moore, J.E., 2023. Antimicrobial Resistance (AMR). British Journal of Biomedical Science 80, 11387. https://doi.org/10.3389/bjbs.2023.11387

Tang, X., Shang, J., Ji, Y., Sun, Y., 2023. PLASMe: a tool to identify PLASMid contigs from short-read assemblies using transformer. Nucleic Acids Res 51, e83. https://doi.org/10.1093/nar/gkad578

Teoh, L., Löffler, C., Mun, M., Agnihotry, A., Kaur, H., Born, K., Thompson, W., 2025. A Systematic Review of Dental Antibiotic Stewardship Interventions. Community Dent Oral Epidemiol 53, 245–255. https://doi.org/10.1111/cdoe.13009

The Sequence Read Archive (SRA) [WWW Document], n.d. URL https://www.ncbi.nlm.nih.gov/sra/docs/ (accessed 8.12.25).

Thurnheer, T., Karygianni, L., Flury, M., Belibasakis, G.N., 2019. Fusobacterium Species and Subspecies Differentially Affect the Composition and Architecture of Supra- and Subgingival Biofilms Models. Front Microbiol 10, 1716. https://doi.org/10.3389/fmicb.2019.01716

Tomoyasu, T., Tabata, A., Imaki, H., Tsuruno, K., Miyazaki, A., Sonomoto, K., Whiley, R.A., Nagamune, H., 2012. Role of Streptococcus intermedius DnaK chaperone system in stress tolerance and pathogenicity. Cell Stress Chaperones 17, 41–55. https://doi.org/10.1007/s12192-011-0284-4

Topçuoğlu, B.D., Lesniak, N.A., Ruffin, M.T., Wiens, J., Schloss, P.D., 2020. A Framework for Effective Application of Machine Learning to Microbiome-Based Classification Problems. mBio 11, 10.1128/mbio.00434-20. https://doi.org/10.1128/mbio.00434-20

Torun, F.M., Virreira Winter, S., Doll, S., Riese, F.M., Vorobyev, A., Mueller-Reif, J.B., Geyer, P.E., Strauss, M.T., 2023. Transparent Exploration of Machine Learning for Biomarker Discovery from Proteomics and Omics Data. J Proteome Res 22, 359–367. https://doi.org/10.1021/acs.jproteome.2c00473

Uddin, T.M., Chakraborty, A.J., Khusro, A., Zidan, B.R.M., Mitra, S., Emran, T.B., Dhama, K., Ripon, M.K.H., Gajdács, M., Sahibzada, M.U.K., Hossain, M.J., Koirala, N., 2021. Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects. J Infect Public Health 14, 1750–1766. https://doi.org/10.1016/j.jiph.2021.10.020

Utter, D.R., Borisy, G.G., Eren, A.M., Cavanaugh, C.M., Mark Welch, J.L., 2020. Metapangenomics of the oral microbiome provides insights into habitat adaptation and cultivar diversity. Genome Biology 21, 293. https://doi.org/10.1186/s13059-020-02200-2

V, C., W, B., O, H., Philip, K., 2002. SMOTE. Journal of Artificial Intelligence Research. https://doi.org/10.5555/1622407.1622416

Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J., 2019. Machine learning algorithm validation with a limited sample size. PLOS ONE 14, e0224365. https://doi.org/10.1371/journal.pone.0224365

van de Guchte, M., Blottière, H.M., Doré, J., 2018. Humans as holobionts: implications for prevention and therapy. Microbiome 6, 81. https://doi.org/10.1186/s40168-018-0466-8

van Hoek, A.H.A.M., Mevius, D., Guerra, B., Mullany, P., Roberts, A.P., Aarts, H.J.M., 2011. Acquired Antibiotic Resistance Genes: An Overview. Front Microbiol 2, 203. https://doi.org/10.3389/fmicb.2011.00203

Väyrynen, E., Tirkkonen, O., Tiensuu, H., Suutala, J., Anttonen, V., Laitala, M.-L., Kukkola, K., Karki, S., 2025. A Machine Learning Algorithm With an Oversampling Technique in Limited Data Scenarios for the Prediction of Present and Future Restorative Treatment Need: Development and Validation Study. JMIR Med Inform 13, e75117. https://doi.org/10.2196/75117

Venkataraman, T., Frieman, M.B., 2017. The role of epidermal growth factor receptor (EGFR) signaling in SARS coronavirus-induced pulmonary fibrosis. Antiviral Res 143, 142–150. https://doi.org/10.1016/j.antiviral.2017.03.022

Veseli, I., Chen, Y.T., Schechter, M.S., Vanni, C., Fogarty, E.C., Watson, A.R., Jabri, B., Blekhman, R., Willis, A.D., Yu, M.K., Fernàndez-Guerra, A., Füssel, J., Eren, A.M., 2025. Microbes with higher metabolic independence are enriched in human gut microbiomes under stress. eLife 12, RP89862. https://doi.org/10.7554/eLife.89862

Vila, T., Rizk, A.M., Sultan, A.S., Jabra-Rizk, M.A., 2019. The power of saliva: Antimicrobial and beyond. PLoS Pathog 15, e1008058. https://doi.org/10.1371/journal.ppat.1008058

Vujkovic-Cvijin, I., Sklar, J., Jiang, L., Natarajan, L., Knight, R., Belkaid, Y., 2020. Host variables confound gut microbiota studies of human disease. Nature 587, 448–454. https://doi.org/10.1038/s41586-020-2881-9

Wade, W.G., 2000. Resilience of the oral microbiome. Periodontol 2000 86, 113–122. https://doi.org/10.1111/prd.12365

Wang, Z., Chen, Y., Li, H., Yue, Y., Yu, H., 2025. Exploring oral microbiome in oral squamous cell carcinoma across environment-associated sample types. Microbiol Spectr 13, e0085224. https://doi.org/10.1128/spectrum.00852-24

Warinner, C., Rodrigues, J.F.M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R.Y., Fiddyment, S., Speller, C., Hendy, J., Charlton, S., Luder, H.U., Salazar-García, D.C., Eppler, E., Seiler, R., Hansen, L.H., Castruita, J.A.S., Barkow-Oesterreicher, S., Teoh, K.Y., Kelstrup, C.D., Olsen, J.V., Nanni, P., Kawai, T., Willerslev, E., von Mering, C., Lewis, C.M., Collins, M.J., Gilbert, M.T.P., Rühli, F., Cappellini, E., 2014. Pathogens and host immunity in the ancient human oral cavity. Nat Genet 46, 336–344. https://doi.org/10.1038/ng.2906

Wei, F., Sun, X., Gao, Y., Dou, H., Liu, Y., Su, L., Luo, H., Zhu, C., Zhang, Q., Tong, P., Ren, W., Xun, Z., Guo, R., Guan, Y., Li, S., Qi, Y., Qin, J., Chen, F., Zheng, S., 2021. Is oral microbiome of children able to maintain resistance and functional stability in response to short-term interference of ingesta? Protein Cell 12, 502–510. https://doi.org/10.1007/s13238-020-00774-y

Wei, F., Wu, Z., Li, G., Sun, X., Shi, X., Tan, L., Ai, T., Qu, L., Zheng, S., 2025. Ensemble learning for microbiome-based caries diagnosis: multi-group modeling and biological interpretation from salivary and plaque metagenomic data. BMC Oral Health 25, 1188. https://doi.org/10.1186/s12903-025-06590-2

Willis, J.R., Gabaldón, T., 2020. The Human Oral Microbiome in Health and Disease: From Sequences to Ecosystems. Microorganisms 8, 308. https://doi.org/10.3390/microorganisms8020308

Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation 1, 67–82. https://doi.org/10.1109/4235.585893

Wright, G.D., 2007. The antibiotic resistome: the nexus of chemical and genetic diversity. Nat Rev Microbiol 5, 175–186. https://doi.org/10.1038/nrmicro1614

Wu, T.T., Xiao, J., Sohn, M.B., Fiscella, K.A., Gilbert, C., Grier, A., Gill, A.L., Gill, S.R., 2021. Machine Learning Approach Identified Multi-Platform Factors for Caries Prediction in Child-Mother Dyads. Front Cell Infect Microbiol 11, 727630. https://doi.org/10.3389/fcimb.2021.727630

Xu, M., Yamada, M., Li, M., Liu, H., Chen, S.G., Han, Y.W., 2007. FadA from Fusobacterium nucleatum utilizes both secreted and nonsecreted forms for functional oligomerization

for attachment and invasion of host cells. J Biol Chem 282, 25000–25009. https://doi.org/10.1074/jbc.M611567200

Yang, J.-X., Peng, Y., Yu, Q.-Y., Yang, J.-J., Zhang, Y.-H., Zhang, H.-Y., Adams, C.A., Willing, C.E., Wang, C., Li, Q.-S., Han, X.-G., Gao, C., 2024. Gene horizontal transfers and functional diversity negatively correlated with bacterial taxonomic diversity along a nitrogen gradient. NPJ Biofilms Microbiomes 10, 128. https://doi.org/10.1038/s41522-024-00588-4

Yonezawa, H., Kokubu, E., Kikuchi, Y., Ishihara, K., 2024. Complete genome sequence of Fusobacterium vincentii strain TDC100 isolated from an apical periodontitis lesion. Microbiol Resour Announc 13, e0119723. https://doi.org/10.1128/mra.01197-23

Yu, O.Y., Zhao, I.S., Mei, M.L., Lo, E.C.-M., Chu, C.-H., 2017. Dental Biofilm and Laboratory Microbial Culture Models for Cariology Research. Dent J (Basel) 5, 21. https://doi.org/10.3390/dj5020021

Yuan, D., Ahamed, A., Burgin, J., Cummins, C., Devraj, R., Gueye, K., Gupta, D., Gupta, V., Haseeb, M., Ihsan, M., Ivanov, E., Jayathilaka, S., Kadhirvelu, V.B., Kumar, M., Lathi, A., Leinonen, R., McKinnon, J., Meszaros, L., O'Cathail, C., Ouma, D., Paupério, J., Pesant, S., Rahman, N., Rinck, G., Selvakumar, S., Suman, S., Sunthornyotin, Y., Ventouratou, M., Vijayaraja, S., Waheed, Z., Woollard, P., Zyoud, A., Burdett, T., Cochrane, G., 2024. The European Nucleotide Archive in 2023. Nucleic Acids Res 52, D92–D97. https://doi.org/10.1093/nar/gkad1067

Zhang, H., Zheng, X., Huang, Y., Zou, Y., Zhang, T., Repo, M.A., Yin, M., You, Y., Jie, Z., Xu, W.-A., 2025. Novel potential biomarkers for predicting childhood caries via metagenomic analysis. Front Cell Infect Microbiol 15, 1522970. https://doi.org/10.3389/fcimb.2025.1522970

Zhang, J.S., Chu, C.-H., Yu, O.Y., 2022. Oral Microbiome and Dental Caries Development. Dent J (Basel) 10, 184. https://doi.org/10.3390/dj10100184

Zhao, A., Sun, J., Liu, Y., 2023. Understanding bacterial biofilms: From definition to treatment strategies. Front Cell Infect Microbiol 13, 1137947. https://doi.org/10.3389/fcimb.2023.1137947

Appendix:

Appendix 0.1: Email confirming the successful ethical approval for the collection of healthy participant oral samples

DREC ref: 150322/JL/349

Study title: Development of an in vitro co-culture biofilm model to decipher human-oral microbiome interactions and investigate mechanisms for prevention of dysbiosis and antimicrobial resistance

Thank you for resubmitting your amended documents to the Dental Research Ethics Committee (DREC). The documents have been reviewed, and I am pleased to inform you that your application has been approved.

Documents reviewed

Document name	Version number/date
Ethics application form	Dated 10.05.2022
Protocol	Version 1 23.03.2022
Participant information sheet	Version 2 10.05.2022
Consent form	Version 2 10.05.2022
Email invitation	Version 2 10.05.2022

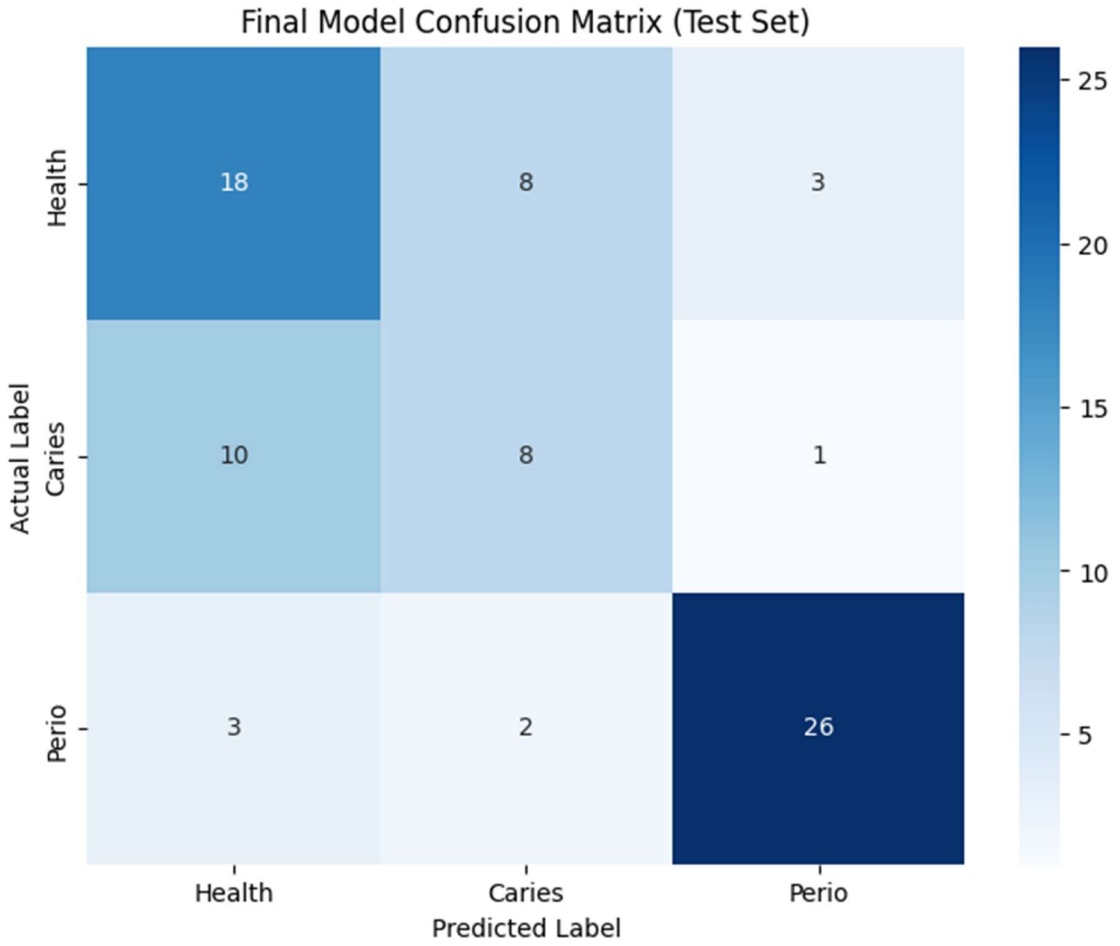
Appendix 3.1: Summary of Study Design, Inclusion Criteria, and Available Metadata of PAD

NCBI SRA Accession Number	Core Investigation	Age Selection Criteria	Disease Classification Metric	DNA Extraction Protocol	Criteria and Available Metadata
PRJNA712952	Fluoride & arginine dentifrice impacts on plaque	≤16 years old	Caries Active = ICDAS 1-3 Caries Free = ICDAS 0	Ambion mirVana miRNA isolation kit. DNA was eluted from the organic phase of the RNA extraction and then ethanol precipitated.	Excluded: recent antibiotic use (< 3 months), advanced periodontal or systemic diseases, altered salivary function, orthodontic appliances, pregnancy, and immunosuppression. Available metadata: Sex.
PRJNA383868	Heritability & ecology of supragingival plaque	5-11-year-old twins	ICDAS II	phenol/chloroform isoamyl alcohol extraction	Twin cohort. Available metadata: Sex, twin status (monozygotic/dizygotic), and dietary sugar frequency.
PRJNA766357	Pit and fissure caries etiology	Adolescents with caries and healthy	Caries Active = DMFT≥1 Caries Free = DMFT=0	CTAB method	Excluded: recent antibiotic use (< 3 months), systemic infections, gingivitis, and orthodontic appliances. Available metadata: Age, oral hygiene habits, and dietary sugar frequency.

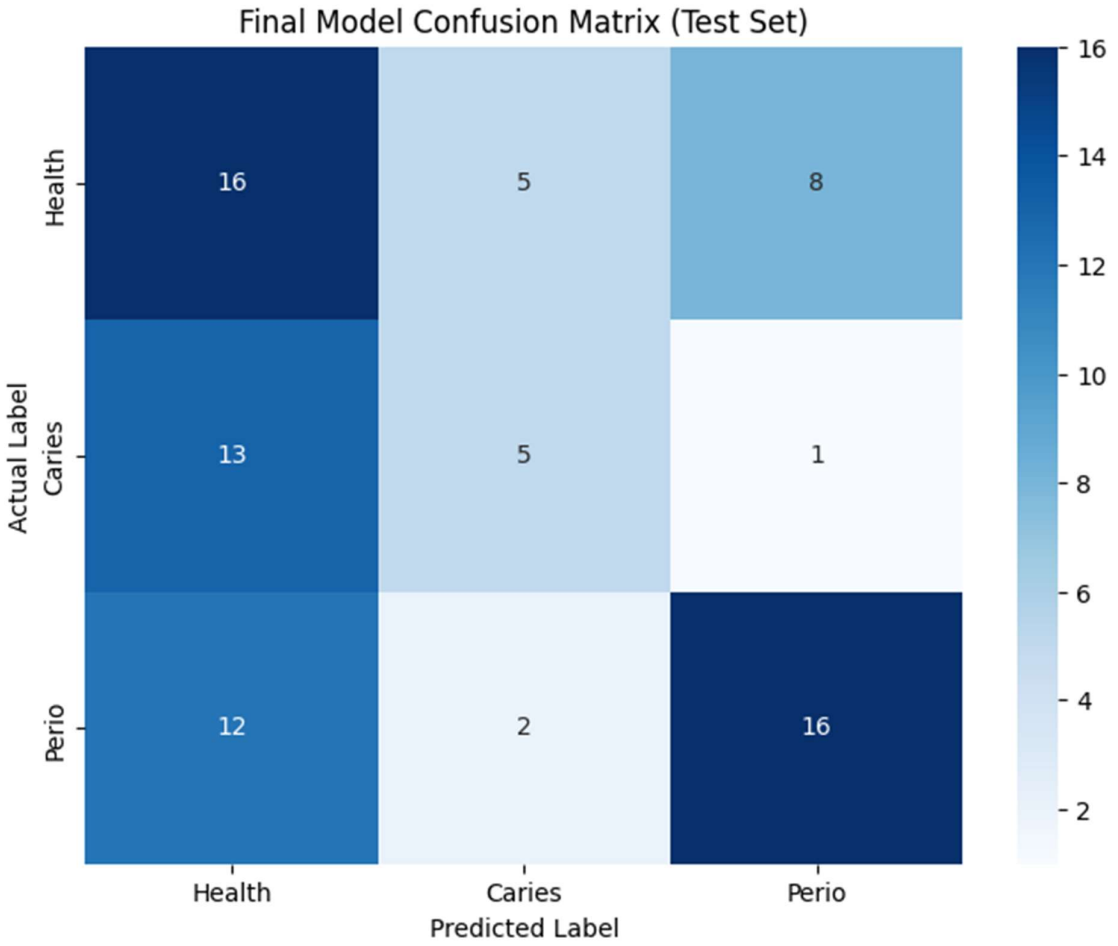
PRJNA255922	Periodontitis in Type 2 Diabetes	≥18 years old	Periodontitis = probing depth of ≥5 mm, gingival index of ≥1 and with bleeding on probing	QIAamp DNA Micro kit	Systemically healthy individuals. Excluded: recent antibiotic use (< 6 months) and history of smoking. Available metadata: Sex (predominantly female).
PRJNA528558	Diagnosis and prognosis of periodontitis in the subgingival microbiome	37 to 65 years old	Periodontitis- gingival index, recession of gums, attachment level, pocket depth, and bleeding on probing. No values explicitly stated. Healthy- no values explicitly stated.	QIAamp DNA microkit	Excluded: recent antibiotic use (< 6 months), history of smoking, and diabetes. Available metadata: Age and sex.
PRJNA625082	Niche partitioning & enigmatic uncultured taxa	21 to 55 years old	No values explicitly stated.	PowerSoil DNA Isolation kit	No specific inclusion/exclusion criteria or additional metadata reported.
PRJNA230363	Phage-bacteria interaction networks	30-65 years old	Periodontitis - no less than 4 mm probing depth and 6 mm attachment loss in at least four non-	QIAamp DNA Mini Kit	Systemically healthy. Excluded: recent antibiotic use (< 3 months), smoking, dental appliances, and prior periodontal therapy. Available metadata: Sex,

			adjacent interproximal sites. Healthy - no probing depth or attachment loss exceeding 2 mm at any site.		smoking, alcohol use, and systemic disease status.
PRJNA932553	Biosynthetic gene clusters in periodontitis	≥18 years old	Clinical examination of probing depth, bleeding on probing and gingival recession.	QIAamp DNA Mini Kit	Excluded: smokers, medications, pregnancy, and orthodontic treatment. Available metadata: Sex (matched pairs) and systemic disease status (diabetes mellitus).
PRJNA508385	Subgingival microbiome in CPT2DM	18-80 years old	Healthy- clinical attachment loss ≤ 1 mm, probing pocket depths ≤ 3 mm, and mean gingival index < 1. Periodontitis was classified based on both phenotype and disease stage and grade.	DNeasy UltraClean Microbial Kit	Excluded: recent antibiotic use (< 3 months), pregnancy, immunosuppression (HIV), previous periodontal therapy, and specific long-term medications. Available metadata: Smoking status (in controls) and systemic disease status (type 2 diabetes).
PRJEB42701	Periodontitis severity	40–80 years	Dutch Periodontal Screening Index (DPSI)	Dneasy UltraClean Microbial Kit	Excluded: recent antibiotic use or infection (< 1 month), cardiovascular disease, autoimmune conditions,

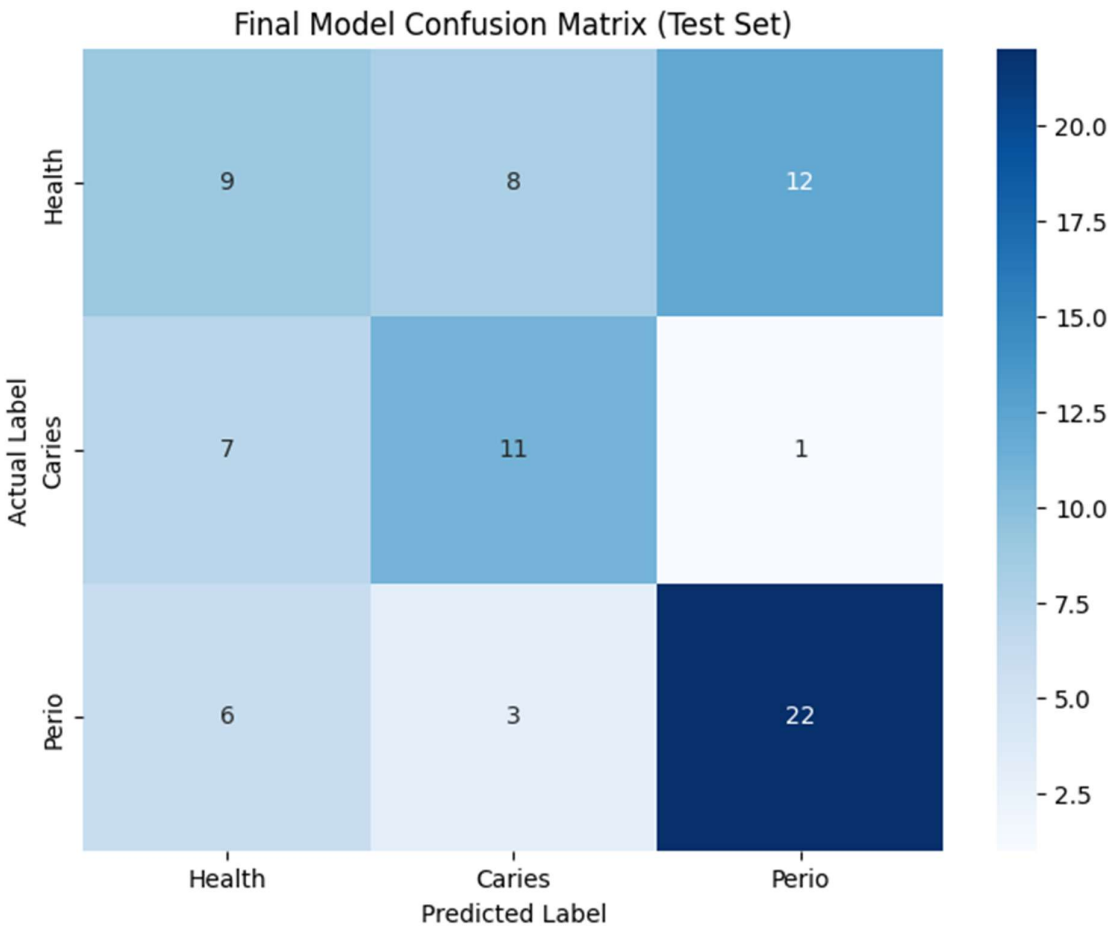
	and probing pocket depths. Periodontitis = DPSI > 3 Healthy = DPSI = 0-2	kidney or liver disease, and chronic immunomodulatory drug use. Available metadata: Systemic health status and medication use.



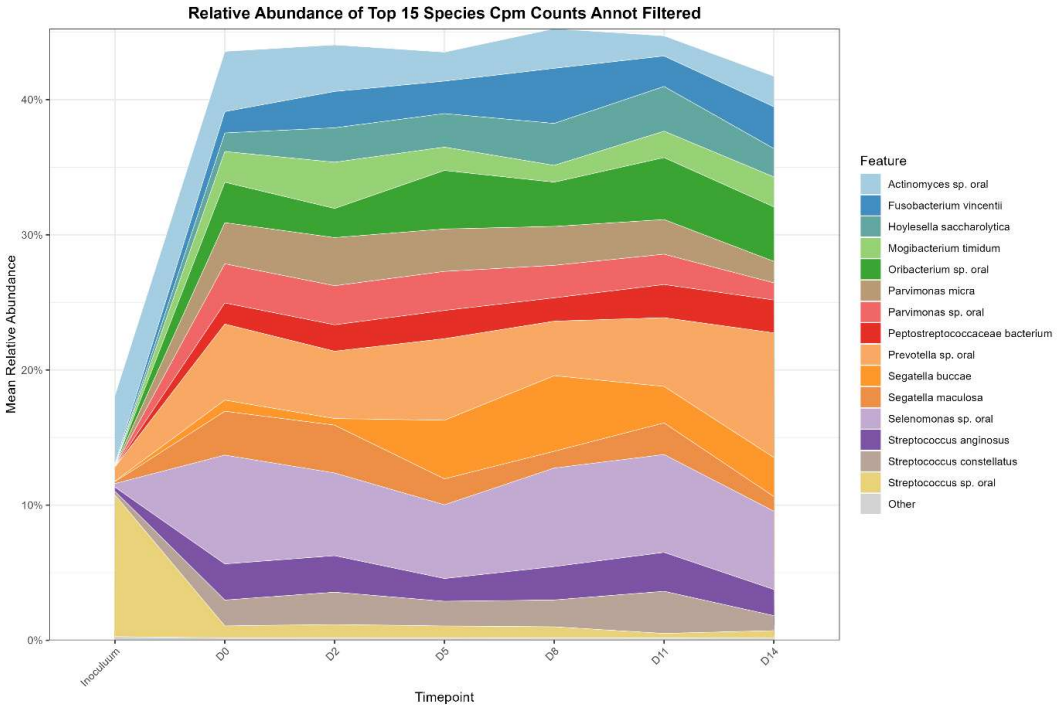
Appendix 4.1. Performance of a multi-class model deep learning for classifying samples as Caries, Health, or Periodontitis with species data. Confusion matrix for the final model on the held-out test set, achieving a multi-class AUC of 0.771.



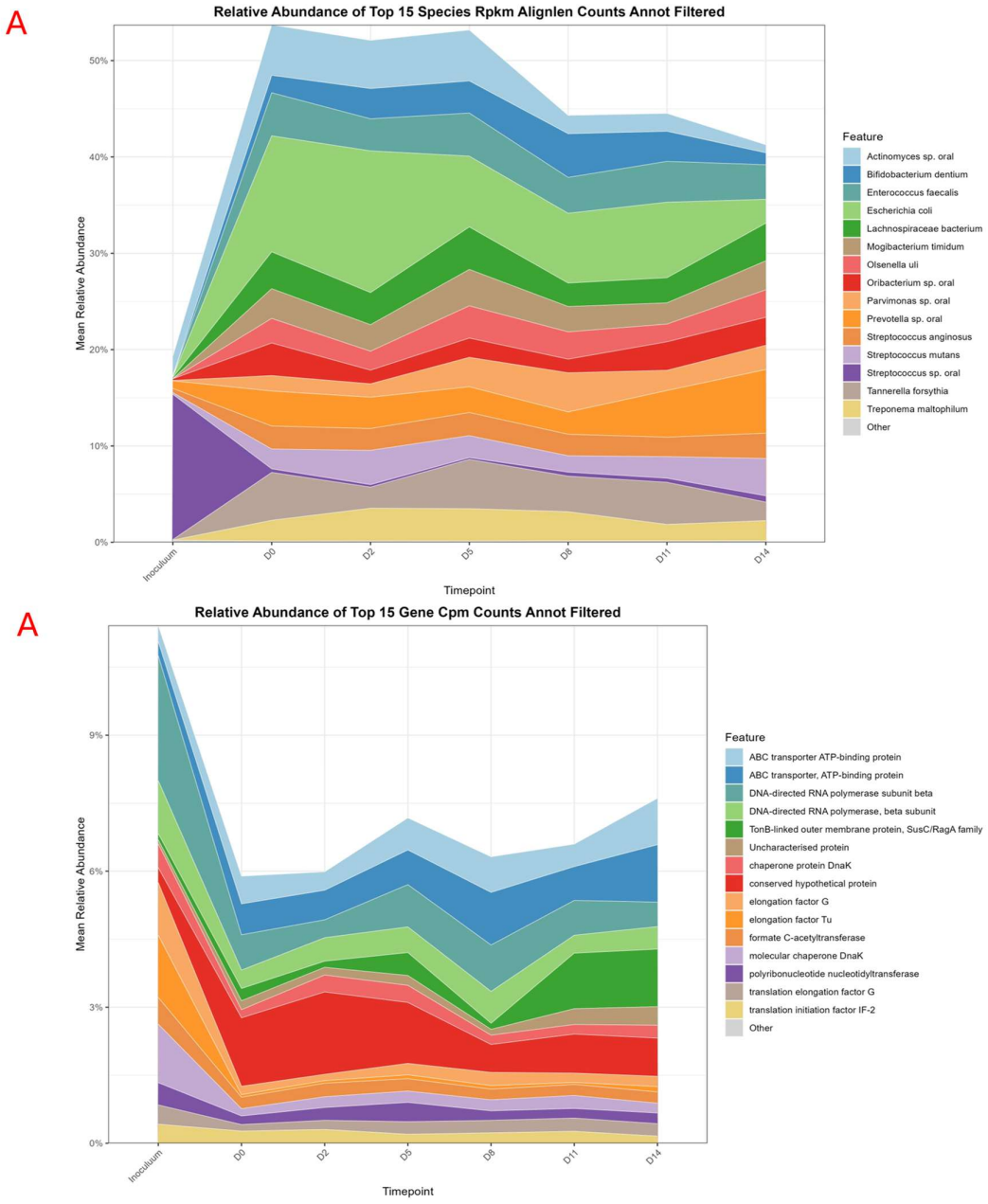
Appendix 4.2. Performance of a multi-class model deep learning for classifying samples as Caries, Health, or Periodontitis with ARG data. Confusion matrix for the final model on the held-out test set, achieving a multi-class AUC of 0.651.



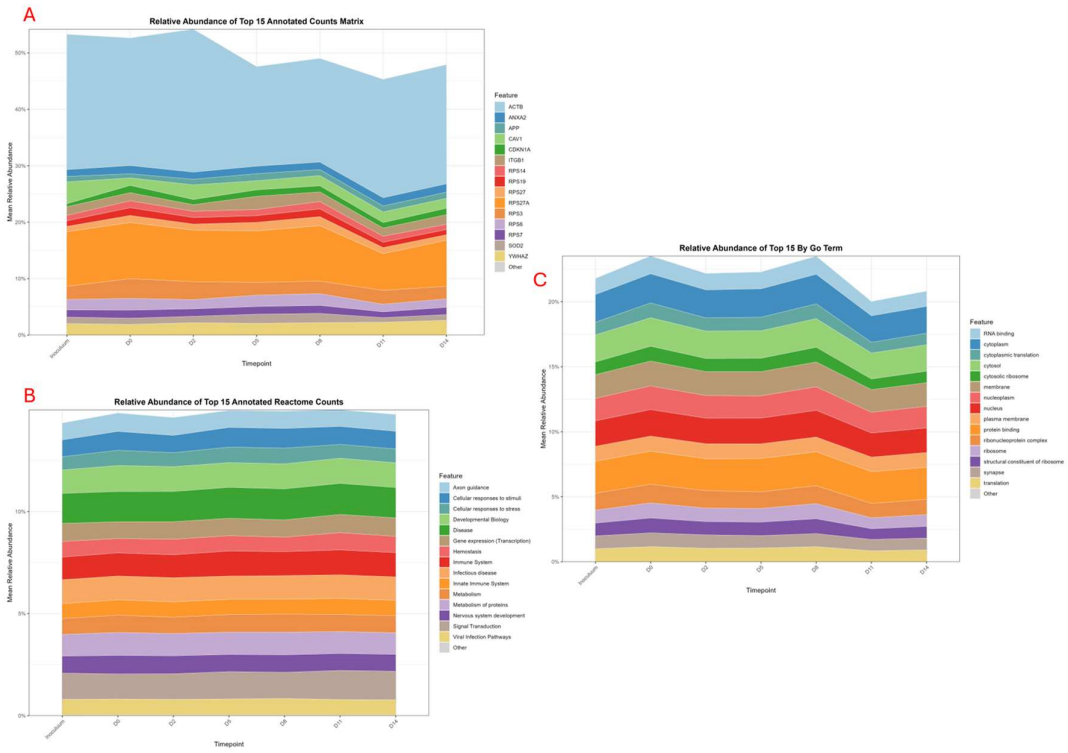
Appendix 4.3. Performance of a multi-class model deep learning for classifying samples as Caries, Health, or Periodontitis with species data. Confusion matrix for the final model on the held-out test set, achieving a multi-class AUC of 0.706.



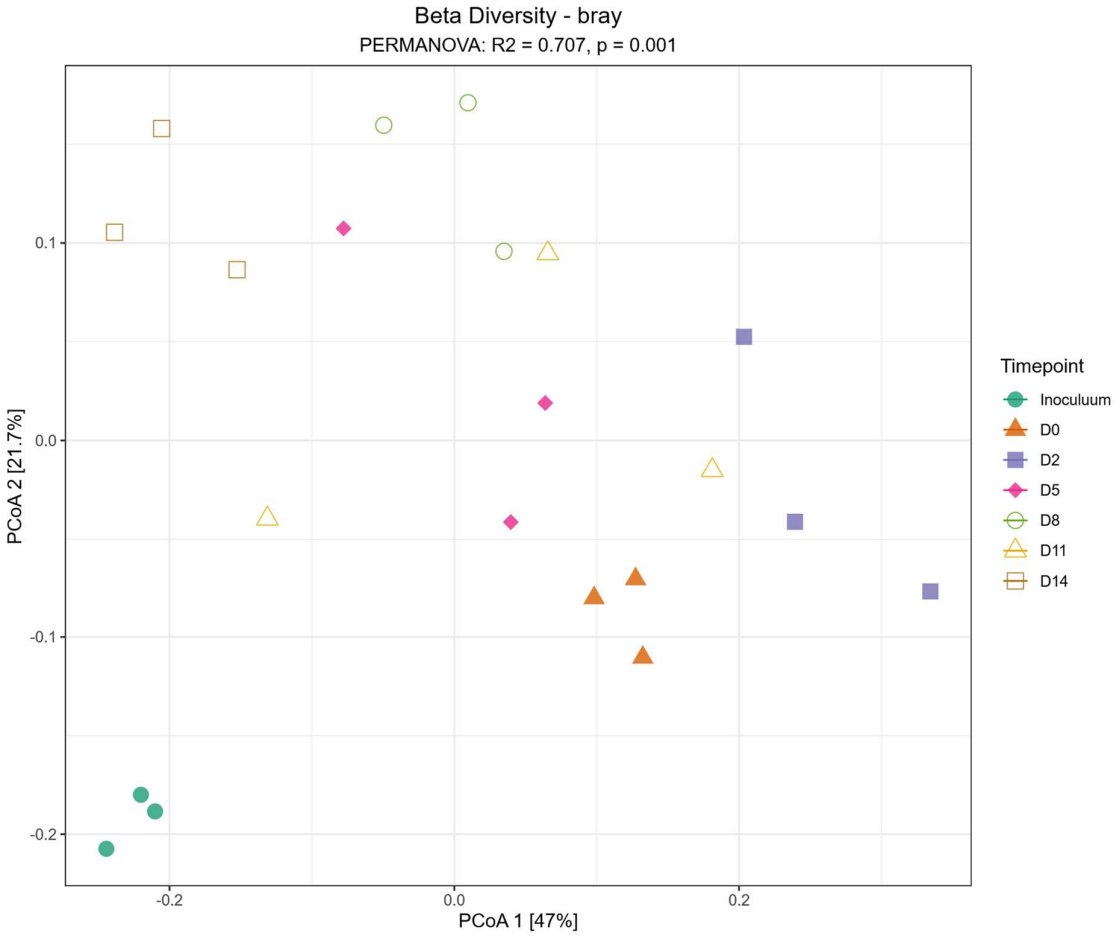
Appendix 5.1. Changes in the relative abundance of the top 15 microbial species. The stacked area chart displays the mean relative abundance of the most prevalent species at each timepoint over the 14-day in vitro dysbiosis experiment. Each coloured area corresponds to a specific species.



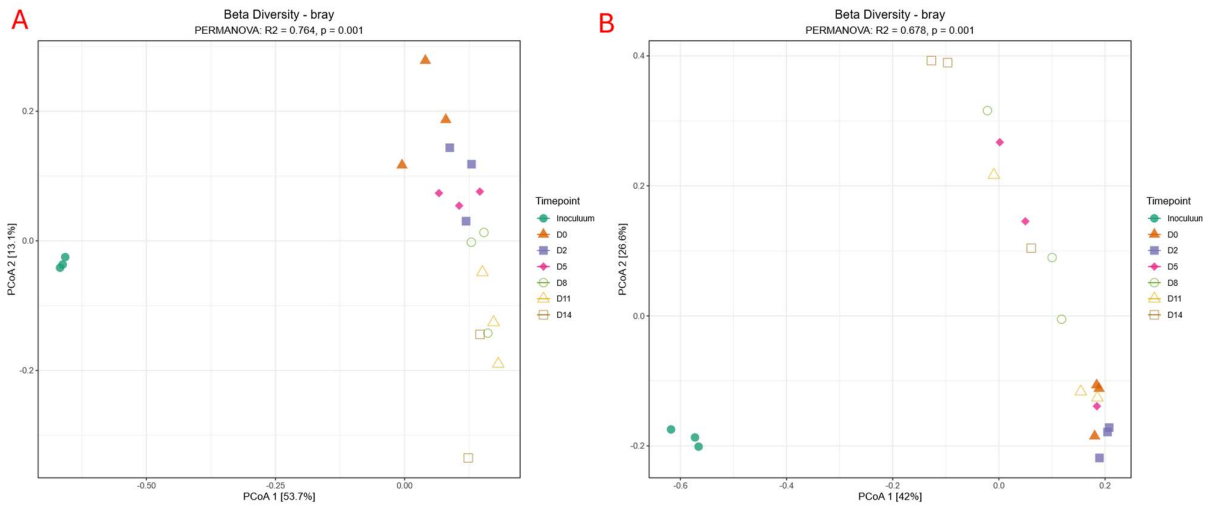
Appendix 5.2. Changes in the relative abundance of the top 15 most transcriptionally active species and gene functions. Stacked area charts displaying the mean relative abundance of the most prevalent features at each timepoint. Each coloured area corresponds to a specific feature, and the 'Other' category comprises all remaining taxa. **(A)** Top 15 most active species. **(B)** Top 15 most expressed gene functions.



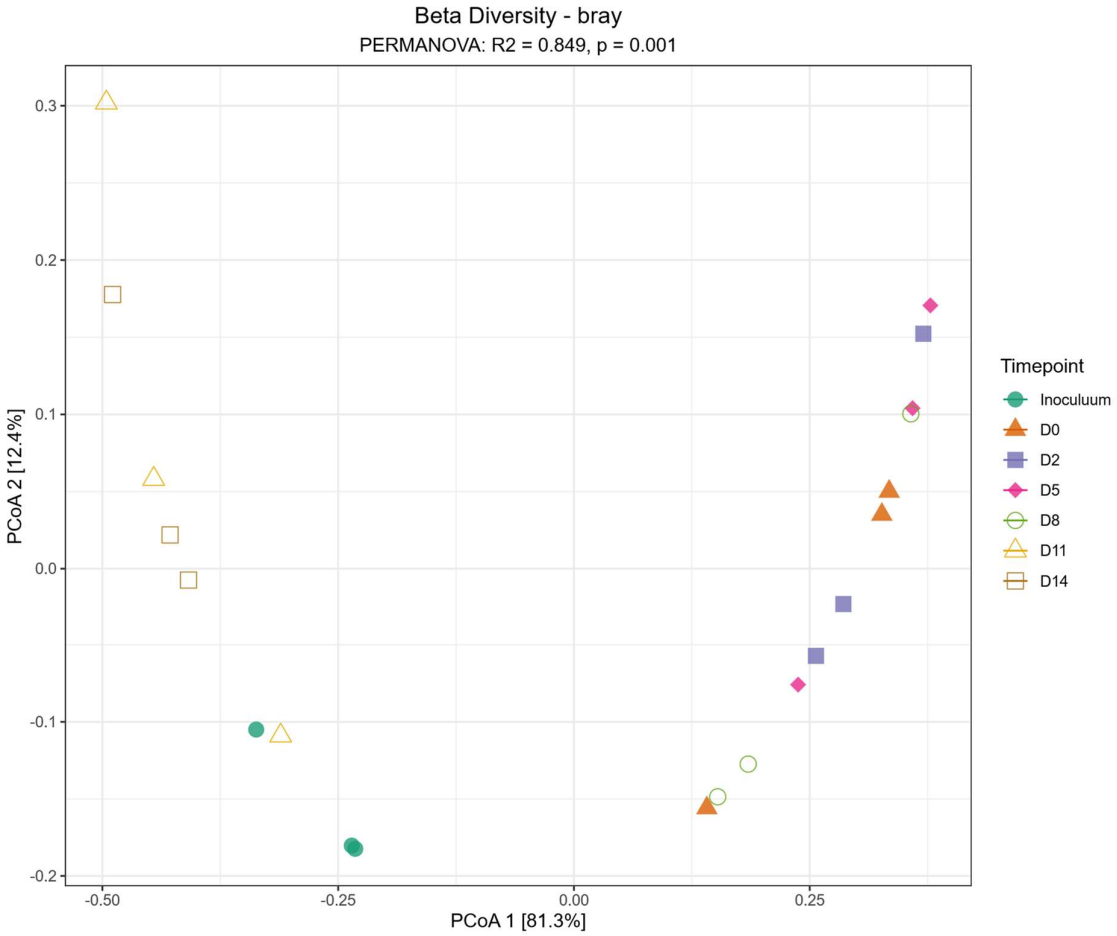
Appendix 5.3. Overview of the most transcriptionally active host genes and pathways over time. Stacked area charts displaying the mean relative abundance of the most expressed features at each timepoint. **(A)** Top 16 most active host genes. **(B)** Top 16 most active Reactome pathways. **(C)** Top 16 most active Gene Ontology (GO) terms.



Appendix 5.4. ARG diversity dynamics over a 14-day in vitro dysbiosis model. Beta diversity shown as a Principal Coordinate Analysis (PCoA) plot of Bray-Curtis dissimilarities. Each point represents a sample, coloured and shaped by timepoint. The plot illustrates a significant shift in community structure over time (PERMANOVA, $R^2 = 0.707$, $p = 0.001$).



Appendix 5.5. Plasmid diversity dynamics over a 14-day in vitro dysbiosis model. A) DNA Beta diversity shown as a Principal Coordinate Analysis (PCoA) plot of Bray-Curtis dissimilarities. Each point represents a sample, coloured and shaped by timepoint. The plot illustrates a significant shift in community structure over time (PERMANOVA, $R^2 = 0.784$, $p = 0.001$). **B) RNA** Beta diversity shown as a Principal Coordinate Analysis (PCoA) plot of Bray-Curtis dissimilarities. Each point represents a sample, coloured and shaped by timepoint. The plot illustrates a significant shift in community structure over time (PERMANOVA, $R^2 = 0.676$, $p = 0.001$).



Appendix 5.6. Host gene diversity dynamics over a 14-day in vitro dysbiosis model. A) DNA Beta diversity shown as a Principal Coordinate Analysis (PCoA) plot of Bray-Curtis dissimilarities. Each point represents a sample, coloured and shaped by timepoint. The plot illustrates a significant shift in community structure over time (PERMANOVA, $R^2 = 0.849$, $p = 0.001$)