

TRANSCRIPTOMIC RESPONSES TO
COAGGREGATION BETWEEN ORAL BACTERIA

NARESH V R MUTHA

INSTITUTE FOR ADVANCED STUDIES
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

**TRANSCRIPTOMIC RESPONSES TO
COAGGREGATION BETWEEN ORAL BACTERIA**

NARESH V R MUTHA

**THESIS SUBMITTED IN FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF DOCTOR OF
PHILOSOPHY**

**INSTITUTE FOR ADVANCED STUDIES
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2020

UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Naresh V R Mutha

Matric No: HHC140001

Name of Degree: Doctor of Philosophy

Title of Project Paper/Research Report/Dissertation/Thesis (“this Work”):

**TRANSCRIPTOMIC RESPONSES TO COAGGREGATION BETWEEN
ORAL BACTERIA**

Field of Study: Bioinformatics

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge, nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every right in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Naresh V R Mutha

Candidate’s Signature

Date:

Subscribed and solemnly declared before,

Witness’s Signature

Date:

Name:

Designation

TRANSCRIPTOMIC RESPONSES TO COAGGREGATION BETWEEN ORAL BACTERIA

ABSTRACT

Cell-cell interactions between genetically distinct oral bacteria form macroscopic clumps known as coaggregates. These interactions, termed coaggregation contribute to the formation of highly structured multispecies communities in oral cavity, known as oral biofilms. Individual species also compete and collaborate with other neighboring species through metabolic interactions allowing growth of other species by coaggregation. This study involves the mixing of selected pairs of oral bacteria to form coaggregates and quantifying differential gene expression levels in each pair in comparison with mono-species coaggregation. The dual RNA-Seq approach made it possible to see how bacteria respond to one another leading to different metabolic interactions for each pairing in coaggregation. In this perspective, *Streptococcus gordonii* was coaggregated with selected oral bacteria. The aim was to analyze transcription data from separate pairings of *Streptococcus gordonii* with *Fusobacterium nucleatum* (first pair), *Veillonella parvula* (second pair) and *Streptococcus oralis* (third pair) to understand gene regulation mechanisms involved in coaggregation. The data analysis involved comparison of mRNA profiles in mono-species cultures vs coaggregates. Significant differentially expressed genes and pathways in *S. gordonii* among all three bacterial pairings were compared to determine whether common mechanisms exist between oral bacterial coaggregates analyzed in this study. In the first pairing a total of 119 genes differentially expressed in *S. gordonii* following coaggregation with *F. nucleatum* whereas only 16 genes had shown differential expression in *F. nucleatum*. In both species, genes involved in amino acid and carbohydrate metabolism were strongly affected by coaggregation. In particular, one 8-gene operon in *F. nucleatum* encoding sialic acid uptake and catabolism was up-

regulated to 2-5 fold following coaggregation. In *S. gordonii*, a gene cluster encoding functions for phosphotransferase system-mediated uptake of lactose and galactose was down regulated up to 3-fold in response to coaggregation. In the second pairing a total of 272 genes differentially expressed in *V. parvula*, including 39 genes in oxidoreductases processes. In *S. gordonii*, there was a high degree of inter-sample variation. Nevertheless, 69 genes were identified as potentially regulated by coaggregation, including two phosphotransferase system transported and several other genes involved in carbohydrate metabolism. In third pairing a total of 22 genes in *S. oralis* and 72 genes in *S. gordonii* were regulated following coaggregation. A 6-gene operon encoding tryptophan in *S. oralis* was down regulated to 1.5-fold following coaggregation whereas mainly transporter genes in *S. gordonii* were up-regulated. A large cluster encoding transporters and two component (NisK/SpaK) regulatory system was upregulated to 2-4 folds in coaggregation.

Keywords: Coaggregation, Oral Biofilm, Dual RNA-Seq, *Streptococcus gordonii*

ABSTRAK

Interaksi sel-sel di antara bakteria oral membentuk gumpalan makroskopik yang dikenali sebagai ko-agregasi. Interaksi ko-agregasi menyumbang kepada pembentukan kaviti oral yang mengandungi kepelbagaian spesies di dalam komuniti yang berstruktur tinggi, turut dikenali sebagai biofilem. Setiap individu species bersaing dan berkolaborasi bersama dengan spesies lain melalui interaksi metabolisme dan membenarkan perkembangan spesies yang lain melalui ko-agregasi. Kajian ini menggabungkan pasangan bakteria terpilih bagi membentuk ko-agregasi dan kuantiti serta kadar gen ekspresi bagi setiap spesies berpasangan di dalam ko-agregasi relatif kepada keadaan kultur tulen ditentukan bagi setiap spesies bakteria. Dari perspektif ini, *Streptococcus* spp. digunakan untuk menjalankan ko-agregasi bersama spesies bakteria terpilih.

Matlamat kajian ini adalah untuk menganalisis data transkriptom dari pasangan bakteria *Streptococcus gordonii* bersama *Fusobacterium nucleatum* (pasangan pertama), *Veillonella parvula* (pasangan kedua), bersama *Streptococcus oralis* (pasangan ketiga) di dalam piring yang berasingan bagi memahami mekanisme gen regulasi yang terlibat di dalam ko-agregasi. Data analisis kajian ini melibatkan perbandingan profil mRNA di antara keadaan monokultur dan keadaan ko-agregasi. Gene perbezaan yang telah diekspres secara signifikan dan tapak laluan bagi tiga pasangan bakteria dibandingkan dan mekanisme umum yang wujud di antara bakteria oral ko-agregasi yang telah dianalisis di dalam kajian ini ditentukan. Bagi pasangan pertama, sejumlah 119 gene mengalami ekspresi berbeza dalam *S. gordonii* setelah ko-agregasi bersama *F. nucleatum* manakala hanya 16 gene telah mengalami ekspresi berbeza di dalam *F. nucleatum*. Bagi kedua-dua spesies, gen yang terlibat di dalam metabolisme amino acid telah dirangsang secara tinggi melalui ko-agregasi. Khususnya, satu 8-gen operon dalam

F. nucleatum yang mengekod pengambilan dan katabolisme asid sialik telah mengalami regulasi positif sebanyak 2-5 ganda setelah ko-agregasi. Bagi *S. gordonii*, satu cluster gen yang mengekod fungsi pengambilan laktosa dan galaktosa melalui rangsangan system fosfotransferase telah mengalami regulasi negatif sebanyak 2-5 ganda setelah ko-agregasi.

Bagi pasangan kedua, sebanyak 272 gen telah mengalami ekspresi berbeza di dalam *V. parvula*. Ini termasuklah 39 gen mengambil bahagian di dalam proses oksireduktas. Bagi *S. gordonii*, didapati variasi inter-sampel hadir pada tahap yang agak tinggi. Walaubagaimanapun, 69 gen telah dikenalpasti mempunyai potensi untuk mengalami regulasi daripada keadaan ko-agregasi. Ini termasuk di dalam sistem pengangkutan fosfotransferase and segelintir gen lain yang terlibat di dalam metabolisme karbohidrat. Di dalam pasangan ketiga, sebanyak 22 gen di dalam *S. oralis* dan sebanyak 72 genes di dalam *S. gordonii* didapati telah mengalami regulasi setelah ko-agregasi. Satu 6-gen operon yang mengekod tritofan operon di dalam *S.oralis* telah mengalami regulasi negatif sebanyak 1.5-ganda setelah ko-agregasi manakala gen yang mengalami regulasi di dalam *S. gordonii* didominasi oleh fungsi pengangkutan. Satu kluster besar yang mengekod untuk 'transporter' dan dua komponen system pengawalseliaan NisK/SpaK mengalami regulasi positif sebanyak 2-4 ganda setelah ko-agregasi.

Keywords: Ko-agregasi, Biofilem oral, Penjujukan RNA pendua, *Streptococcus gordonii*

ACKNOWLEDGEMENTS

Firstly, I would like to offer my sincere gratitude and obeisances to my supervisors, Associate Professor Dr. Geok Yuan Annie Tan, Professor Dr. Lawrence Choo Siew Woh and Dr. Nicholas Jakubovics for their invaluable advice, guidance and enormous patience in guiding me all through the project. I appreciate their willingness to teach, supervise and encourage me throughout my PhD study and the preparation of the project reports and papers.

In addition, I would also like to thank my colleagues in Genome Informatics Research Group (GIRG) for their supports, opinions and suggestions rendered to me throughout this project. I sincerely appreciate the High Impact Research Committee for giving me financial assistance through the Graduate Research Assistantship Scheme (GRAS) for my PhD study. I would like to take this opportunity to thank the staff from the High Impact Research (HIR) secretariat and the Institute for advanced studies (IAS) for their excellent administrative works.

Additionally, I extend my thanks to University of Malaya and Ministry of Education (MOE), Malaysia which funded my study. Furthermore, I would also like to thank all those who helped in this study, in particular my friend G. Ranganath and Waleed K. Mohammed who helped me in understanding wet lab methodologies for my research. I am especially grateful for support of Professor Dr. Jennifer Ann Harikrishna for the last two semesters of my research.

I thank my parents for everything they have done for me. Without their help and support I would not have reached this point and I am enormously grateful and dedicate my thesis to “Amma” and “Dad”.

TABLE OF CONTENTS

Abstract	iii
Abstrak	v
Acknowledgements	vii
Table of Contents	viii
List of Figures	xii
List of Tables	xiv
List of Symbols and Abbreviations.....	xv
List of Appendices	xvii
CHAPTER 1: INTRODUCTION	1
1.1 Overview	1
1.2 Significance of the project.....	3
1.3 Project objectives	3
CHAPTER 2: LITERATURE REVIEW.....	4
2.1 Interspecies communication in oral biofilms	4
2.2 Significance of coaggregation in oral biofilm formation	6
2.2.1 <i>Streptococcus gordonii</i> : Coaggregation and its significance	8
2.2.2 <i>Fusobacterium nucleatum</i> : Coaggregation and its significance.....	10
2.2.3 <i>Veillonella parvula</i> : Coaggregation and its significance	12
2.2.4 <i>Streptococcus oralis</i> : Coaggregation and its significance.....	13
2.3 Importance of dual species coaggregation experiments.....	14
2.4 Metabolic interactions and associations of biofilms	15
2.5 High through-put sequencing technologies and its applications	17
2.5.1 Sequencing evolution, achievements and limitations.....	17
2.5.2 Key high-throughput sequencing platforms	17
2.5.3 Illumina sequencing platform and its significance	19
2.5.4 Dual RNA-seq approach for studying oral bacterial coaggregations.....	20

2.5.5	General work flow of RNA-Seq data analysis and related bioinformatics tools	25
2.5.5.1	FASTQ format.....	29
2.5.5.2	Quality scores.....	29
2.5.5.3	SAM/BAM format	30
CHAPTER 3: MATERIALS AND METHODS		31
3.1	Study Design	31
3.2	Library preparation and Dual RNA Sequencing.....	31
3.3	Bioinformatics analysis of transcriptomes	31
3.3.1	RNA-Seq datasets.....	33
3.3.2	RNA-Seq data generation, pre-processing and quality trimming	33
3.3.3	Alignment of pre-processed RNA-Seq reads	33
3.3.4	Gene expression quantification	34
3.3.5	Differential expression analysis	34
3.3.6	Gene enrichment analysis.....	35
3.3.7	Putative protein interaction network construction.....	35
3.4	RT-qPCR validation of RNA-Seq data	35
3.5	Common genes/pathways expressed in <i>Streptococcus gordonii</i> in response to coaggregation with <i>Fusobacterium nucleatum</i> , <i>Veillonella parvula</i> and <i>Streptococcus oralis</i>	36
CHAPTER 4: TRANSCRIPTOMIC RESPONSES TO COAGGREGATION BETWEEN <i>STREPTOCOCCUS GORDONII</i> AND <i>FUSOBACTERIUM NUCLEATUM</i> 37		
4.1	Coaggregation of <i>Streptococcus gordonii</i> and <i>Fusobacterium nucleatum</i>	37
4.2	Dual RNA-Seq data analysis	39

4.2.1	RNA-Seq data generation and pre-processing	39
4.2.2	Mapping reads to reference genome and transcript abundance estimation 40	
4.3	Differential expression analysis	42
4.4	RT-qPCR validation of RNA-Seq data	42
4.5	Discussion	43
4.5.1	Gene regulation in <i>S. gordonii</i> in response to coaggregation with <i>F. nucleatum</i>	43
4.5.2	Gene regulation in <i>F. nucleatum</i> in response to coaggregation with <i>S. gordonii</i>	49

CHAPTER 5: TRANSCRIPTOMIC RESPONSES TO COAGGREGATION BETWEEN *STREPTOCOCCUS GORDONII* AND *VEILLONELLA PARVULA*... 53

5.1	Coaggregation of <i>Streptococcus gordonii</i> and <i>Veillonella parvula</i>	53
5.2	Dual RNA-Seq data analysis	54
5.2.1	RNA-Seq data generation and pre-processing	54
5.2.2	Mapping reads to reference genome and abundance estimation	55
5.2.3	Differential expression analysis	57
5.3	RT-qPCR validation of RNA-Seq data	58
5.4	Discussion	60
5.4.1	Gene regulation in <i>S. gordonii</i> in response to coaggregation with <i>V. parvula</i>	60
5.4.2	Gene regulation in <i>V. parvula</i> in response to coaggregation with <i>S. gordonii</i>	65

CHAPTER 6: TRANSCRIPTOMIC RESPONSES TO COAGGREGATION BETWEEN *STREPTOCOCCUS GORDONII* AND *STREPTOCOCCUS ORALIS*.72

6.1	Coaggregation of <i>Streptococcus gordonii</i> and <i>Streptococcus oralis</i>	72
6.2	Dual RNA-Seq data analysis	73
6.2.1	RNA-Seq data generation and pre-processing	73
6.2.2	Mapping reads to reference genome and transcript abundance estimation 73	
6.2.3	Differential expression analysis	75
6.3	Discussion	76
6.3.1	Gene regulation in <i>S. gordonii</i> in response to coaggregation with <i>S. oralis</i> 76	
6.3.2	Gene regulation in <i>S. oralis</i> in response to coaggregation with <i>S. gordonii</i> 79	
CHAPTER 7: FINAL DISCUSSION AND CONCLUSIONS		82
References		86
List of Publications and Papers Presented		101
Appendix		102

LIST OF FIGURES

Figure 1.1: Oral bacterial coaggregation.....	1
Figure 2.1: A model of oral biofilm formation.	5
Figure 2.2: Diagram illustrating the possible roles of coaggregation in the development of multispecies biofilms.	7
Figure 2.3: Illumina sequencing workflow.	20
Figure 2.4: Single end sequencing and paired end sequencing.....	22
Figure 2.5: RNA-Seq Workflow	23
Figure 3.1: Work flow implemented for oral bacterial coaggregation studies using dual RNA-Seq approach	32
Figure 4.1: Confocal scanning laser microscopy analysis of coaggregation between <i>Streptococcus gordonii</i> and <i>Fusobacterium nucleatum</i>	37
Figure 4.2: High Resolution Analysis of Interactions between <i>Streptococcus gordonii</i> and <i>Fusobacterium nucleatum</i> cells using Transmission Electron Microscopy.....	39
Figure 4.3: Boxplots representing quality of read counts in SgFn experiment.....	41
Figure 4.4: Volcano plots representing differentially expressed genes in coaggregates vs monocultures for <i>Streptococcus gordonii</i> and <i>Fusobacterium nucleatum</i>	42
Figure 4.5: RT-qPCR validation of RNA-Seq data analysis results for <i>Streptococcus gordonii</i> and <i>Fusobacterium nucleatum</i>	43
Figure 4.6: Genes involved in DNA maintenance that were regulated in <i>S. gordonii</i> in response to coaggregation with <i>F. nucleatum</i> visualized by network	44
Figure 4.7: Genes involved in amino acid metabolism that were regulated in <i>S. gordonii</i> in response to coaggregation with <i>F. nucleatum</i> visualized by network.	45
Figure 4.8: Genes involved in carbohydrate metabolism that were regulated in <i>S. gordonii</i> in response to coaggregation with <i>F. nucleatum</i> visualized by network.....	47
Figure 4.9: Network of genes involved in carbohydrate metabolism that were regulated in <i>S. gordonii</i> in response to coaggregation with <i>F. nucleatum</i>	51
Figure 5.1: Confocal scanning laser microscopy analysis of coaggregation between <i>S. gordonii</i> and <i>V. parvula</i>	53

Figure 5.2: High resolution analysis of interactions between <i>S. gordonii</i> and <i>V. parvula</i> cells using Transmission Electron Microscopy.....	54
Figure 5.3: Box plots showing transcript abundance before and after normalization in SgVp experiment.....	56
Figure 5.4 : Volcano plots representing differentially expressed genes in coaggregates vs monocultures for <i>S. gordonii</i> and <i>V. parvula</i>	57
Figure 5.5 Validation of RNA-Seq data by RT-qPCR.....	59
Figure 5.6 Genes regulated in <i>S. gordonii</i> in coaggregation with <i>V. parvula</i>	61
Figure 5.7: Putative protein interaction networks in the ‘oxidoreductase’ category of the <i>V. parvula</i> differentially expressed gene dataset.....	66
Figure 5.8: DNA metabolism genes expressed in <i>V. parvula</i> in coaggregation with <i>S. gordonii</i>	68
Figure 5.9: Amino acid metabolism genes expressed in <i>V. parvula</i> in coaggregation with <i>S. gordonii</i>	69
Figure 6.1 Confocal scanning laser microscopy analysis of coaggregation between <i>S. gordonii</i> and <i>S. oralis</i>	72
Figure 6.2: Boxplots representing quality of read counts in SgSo experiment.....	75
Figure 6.3: Volcano plots representing differentially expressed genes in coaggregates vs monocultures for <i>S. gordonii</i> and <i>S. oralis</i>	76
Figure 6.4 Network of genes regulated in <i>S. oralis</i> in coaggregation with <i>S. gordonii</i> . 77	
Figure 6.5 Network of genes regulated in <i>S. gordonii</i> in coaggregation with	80
Figure 7.1 Venn diagram to represent overlapping <i>S. gordonii</i> genes between three comparisons	82

LIST OF TABLES

Table 4.1: Alignment statistics of mixed and monoculture transcriptomes in <i>S. gordonii</i> and <i>F. nucleatum</i> coaggregation experiments.....	40
Table 4.2: DNA metabolic process genes of interest regulated in <i>S. gordonii</i> in response to coaggregation with <i>F. nucleatum</i>	44
Table 4.3: Amino acid metabolic process genes of interest regulated in <i>S. gordonii</i> in response to coaggregation with <i>F. nucleatum</i>	46
Table 4.4: Carbohydrate metabolic process genes of interest regulated in <i>S. gordonii</i> in response to coaggregation with <i>F. nucleatum</i>	48
Table 4.5 Genes regulated in <i>F. nucleatum</i> following coaggregation with <i>S. gordonii</i> .	50
Table 4.6 Sialic acid metabolism genes regulated in <i>F. nucleatum</i> following coaggregation with <i>S. gordonii</i>	50
Table 5.1: Alignment statistics of mixed and monoculture transcriptomes in <i>S. gordonii</i> and <i>V. parvula</i> in coaggregation experiment	55
Table 5.2: DNA metabolism genes regulated in <i>V. parvula</i> in response to coaggregation with <i>S. gordonii</i>	69
Table 5.3: Amino acid metabolism genes regulated in <i>V. parvula</i> in response to coaggregation with <i>S. gordonii</i>	70
Table 6.1 Alignment statistics of mixed and monoculture transcriptomes of <i>S. gordonii</i> and <i>S. oralis</i> in coaggregation experiment.....	74
Table 6.2 Ribosomal proteins regulated in <i>S. oralis</i> in response to coaggregation with <i>S. gordonii</i>	77
Table 6.3 Tryptophan metabolism genes regulated in <i>S. oralis</i> in response to coaggregation with <i>S. gordonii</i>	78
Table 6.4 Tryptophan metabolism genes regulated in <i>S. oralis</i> in response to coaggregation with <i>S. gordonii</i>	79
Table 7.1 List of common <i>S. gordonii</i> genes between three comparisons.....	83

LIST OF SYMBOLS AND ABBREVIATIONS

HTS	:	High-throughput sequencing technologies
NGS	:	Next generation sequencing technologies
qRT-PCR	:	Quantitative reverse transcription– Polymerase Chain Reaction
RNA-Seq	:	RNA Sequencing
WTSS	:	Whole transcriptome shotgun sequencing
dRNA-Seq	:	Dual RNA-Sequencing
Bp	:	Base pair
SAM/BAM	:	Sequence alignment map/ Binary alignment map
NB	:	Negative Binomial
BGI	:	Beijing Genomic Institute
DTT	:	Dithiothreitol
THYE	:	Todd-Hewitt Yeast Extract
FAB	:	Fastidious Anaerobe Broth
PBS	:	Phosphate Buffered Saline
CSLM	:	Confocal Laser Scanning Microscopy
Ex	:	Extraction
Em	:	Emission
OD	:	Optical density
BHYG	:	Brain Heart infusion, Yeast extract and sodium glutamate
SgFn	:	<i>Streptococcus gordonii</i> - <i>Fusobacterium nucleatum</i>
SgVp	:	<i>Streptococcus gordonii</i> – <i>Veillonella parvula</i>
SgSo	:	<i>Streptococcus gordonii</i> – <i>Streptococcus oralis</i>

NCBI	:	National Centre for Biotechnology Information
HSP	:	High scoring segment pair
STRING	:	Search Tool for the Retrieval of Interacting Genes/Proteins
3D	:	Three dimensional
TEM	:	Transmission electron microscopy
SBF-SEM	:	Serial block-face scanning electron microscopy
PRJNA	:	Bioproject Accession Number
KEGG	:	Kyoto Encyclopedia of genes and genomes
TMM	:	Trimmed Mean of M values
PcoA	:	Principle coordinate Analysis
FDR	:	False Discovery Rate
SRA	:	Sequenced Read Archive
SAGE	:	Serial Analysis of Gene Expression
MDS-plot	:	Multi-dimensional scaling plot

LIST OF APPENDICES

Appendix A: R script used for differential expression analysis.....	101
Appendix B: R scripts for Boxplots.....	102
Appendix C: RT-PCR validations.....	103
Appendix D: R script for Volcano plots.....	104
Appendix E: Principal component analysis of biological repeats for monoculture and coaggregation samples.....	105
Appendix F: R script for MDS plot.....	106
Appendix G: Multi-dimensional scaling analysis of biological repeats for batch-1 (Old) and batch-2 monoculture and coaggregation samples.....	107

CHAPTER 1: INTRODUCTION

1.1 Overview

Oral bacteria live in structurally and functionally organized communities on the surfaces of the hard and soft tissues in the mouth (Bowen *et al.*, 2018). In any individual approximately 700 different species of oral bacteria are harboured and adapted to colonize to form biofilms in oral cavity (Huang *et al.*, 2011). Bacterial adhesion to surfaces is an important step in colonization and oral biofilm formation (Kolenbrander *et al.*, 2010). Oral bacterial cells adherence to immobilized bacteria is called coadhesion, and binding of bacteria in cell suspension is called coaggregation. Direct physical contact of two individual bacterial species with their neighbouring co-inhabitants within a microbial community could initiate a signalling cascade and achieve modulation of gene expression (Figure 1.1) in accordance with different species (Parashar *et al.*, 2015). These intergenic pairings are important in the sequential formation of oral biofilms and plaque formation.

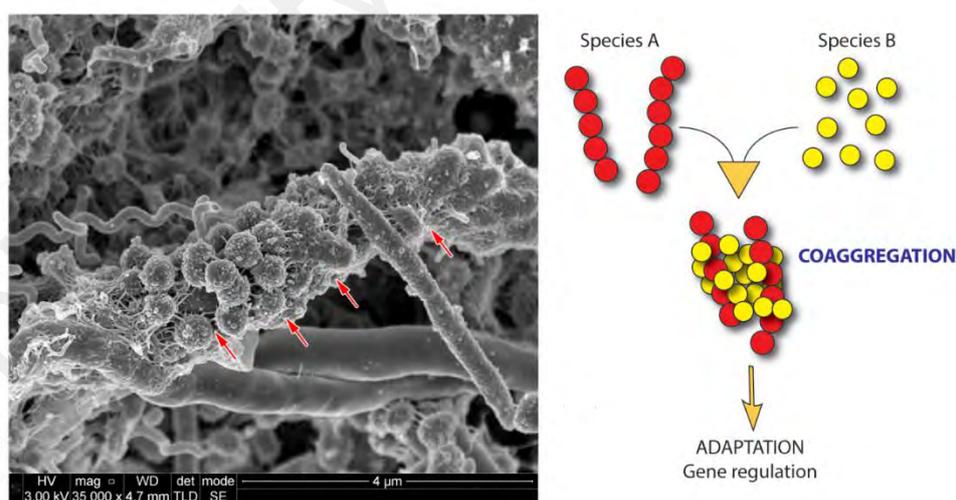


Figure 1.1: Oral bacterial coaggregation.

The panel on the left shows a scanning electron micrograph of dental plaque. The central structure (arrowed) is a well-known type of coaggregation called a ‘corn-cob’, where spherical cells (cocci) surround a large rod-shaped cell. The cartoon emphasises the role of coaggregation as an active process that leads to changes in the biology of the system.

Many studies have employed microarrays and proteomics to elucidate gene regulation during the formation of mixed-species biofilms (Hendrickson *et al.*, 2014; Hendrickson *et al.*, 2012; Jakubovics *et al.*, 2008a; Liu *et al.*, 2011; Luppens *et al.*, 2008) but, so far not many sought to look at the transcriptomes of both simultaneously using high through-put sequencing approaches. Recently, Dual RNA-Seq has been used to study transcriptional changes in *Candida albicans* (fungus) and *S. gordonii* following cell-cell interactions (Dutton *et al.*, 2016).

Streptococci are considered as one of the main group of early colonizers, subsequently, *Fusobacteria* and *Veillonella* bind by coaggregation to these pioneer species. *Streptococcus gordonii* is considered to be among the first species to colonize and compete more effectively than any other tested oral streptococci for adherence (Rosan *et al.*, 2000). From this perspective we study *Streptococcus gordonii* coaggregations with other species in the context of oral cavity, in order to investigate how bacteria respond genetically to interactions with other species. The results of this analysis could be useful to show how oral microbes participate respond genetically to co-inhabitants through cell contact dependent physical interactions, metabolic interdependencies, as well as coordinative signalling systems to establish and maintain balanced microbial communities. High-throughput sequencing technologies have accelerated knowledge on the oral microbiome. The dual RNA-Seq approach made it possible to see whether bacterial coaggregation lead to different responses depending on the pairings.

In this study, cells with saliva as the sole nutrient were used to determine the role of coaggregation in biofilm architecture and composition that occurred in cells that were inoculated either with individual species or simultaneously with coaggregates of mixed species. Four bacterial species were chosen as the inocula for the mixed-species biofilms; three of these species have been identified as primary colonizers of the tooth

surface. These early colonizers include the facultative anaerobes *Streptococci gordonii*, *Veillonella parvula* and *Streptococcus oralis*. The fourth species used in this study, *Fusobacterium nucleatum*, an obligate anaerobe, is one of the most abundant gram-negative bacteria in subgingival oral biofilms (Moore *et al.*, 1994). Together, these four species exemplify a wide range of metabolic and physiological characteristics that serve to represent a diverse array of oral bacteria.

1.2 Significance of the project

Coaggregation followed by dual RNA-Seq provides very good model to analyze the early responses of bacterial interactions with taxonomically distinct cells. This study demonstrates the feasibility of using dual RNA-seq approach for identifying and studying the genes involved in coaggregation of oral biofilm formation. The simultaneous sequencing of the genomes of two oral bacteria involved in biofilm formation has been successfully applied to decipher transcriptional changes involved in coaggregate formation. This study identified a registry of genes and pathways in mixed culture that play important roles in coaggregation mechanism.

1.3 Project objectives

The objectives of this study were

- To identify the genes that are regulated in response to coaggregation of *Streptococcus gordonii* with *Fusobacterium nucleatum*, *Veillonella parvula* and *Streptococcus oralis* each paired separately.
- To annotate differentially expressed genes in each pair:
 - Streptococcus gordonii* DL1 – *Fusobacterium nucleatum* ATCC25586
 - Streptococcus gordonii* DL1 – *Veillonella parvula* PK1910
 - Streptococcus gordonii* DL1 – *Streptococcus oralis* 34
- To identify whether there are common genes/responses in three pairs of coaggregation.

CHAPTER 2: LITERATURE REVIEW

2.1 Interspecies communication in oral biofilms

Oral bacteria interact with their environment by attaching to surfaces and establish to mixed-species communities. Adherent bacteria sense their neighbours and exchange messages through signalling molecules and metabolites developing into micro colonies in oral cavity termed as oral biofilms (Parashar *et al.*, 2015). These bacterial biofilms are highly complex microbial communities and up to 700 different phlotypes are natural colonizers of the human mouth (Whittaker *et al.*, 1996). These bacterial species are thought to play important roles in the maintenance of oral health and in the aetiology of oral human diseases (Socransky *et al.*, 2002).

The foundations of biofilms are laid by the pioneer bacteria genera attaching to the tooth surface, predominantly *Streptococcus* spp., *Veillonella* spp., *Actinomyces* spp., *Haemophilus* spp., *Neisseria* spp. and several other oral bacteria (Figure 2.1) providing binding sites for co-adhesion with other bacteria (Foster *et al.*, 2004), Watnick *et al.*, 2000). The Initial stage of oral biofilm formation starts by the recognition of binding proteins (α -amylase and proline-rich glycoproteins/proteins) in the acquired pellicle by planktonic bacterial cells and attachment to surfaces of the oral cavity by excreted polymeric substances (EPS) or indirectly bind to other bacterial cells that have already colonized (Jakubovics, 2010). It is believed that these early colonizers promote the establishment of other bacteria by providing specific binding sites either directly or through salivary glycoproteins binding to the pioneer mircoorganisms for subsequent bacterial colonization and promote biofilm development. Later colonizing bacteria recognize polysaccharide or protein receptors on the pioneer bacterial cell surface and attach to them. Bacteria coaggregate, forming the typical corn cob forms, bristle brush forms, or other forms in mature oral biofilm.

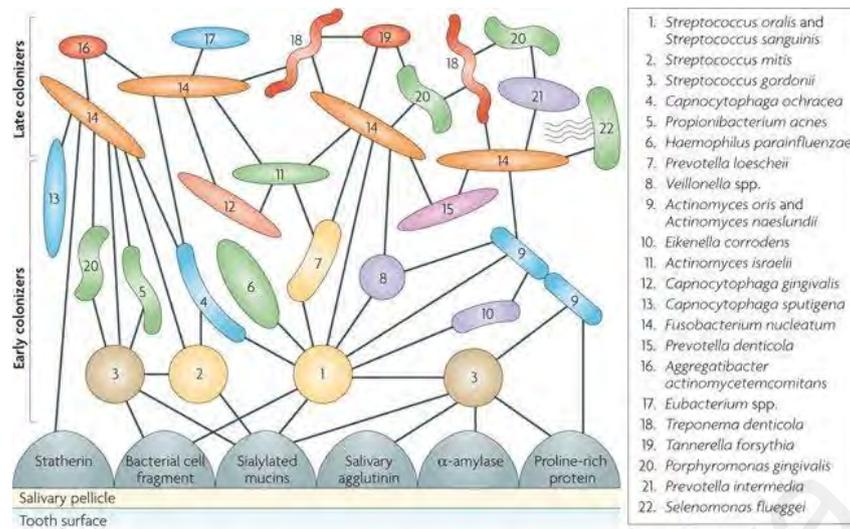


Figure 2.1: A model of oral biofilm formation.
 adapted from (Kolenbrander *et al.*, 2010)

Subsequent attached bacteria species/late colonizers include *Fusobacterium nucleatum*, *Treponema* spp., *Porphyromonas gingivalis* and *Aggregatibacter actinomycetemcomitans* (Kolenbrander *et al.*, 2002). A biofilm provides many advantages to the reproduction, metabolism and defense from other bacteria or the host. Oral bacteria such as viridans streptococci can colonize the tooth surface by binding to the complex proteinaceous pellicle (Kolenbrander *et al.*, 2002; Rogers *et al.*, 2001). *Fusobacterium nucleatum* is thought to play a central role in the maturation of dental plaque because it coaggregates with almost all oral bacteria, including early colonizers and later colonizers. Hence, *F. nucleatum* forms a coaggregation bridge between oral bacteria that do not naturally coaggregate with each other. A previous comprehensive review has provided concrete evidence that can help identify specific interspecies interactions in the co-aggregation among oral bacteria (Kolenbrander *et al.*, 2002). For oral bacteria, nutrients are available from saliva, gingival crevicular fluid, food containing sugars, food debris, and metabolic products of other bacteria. The advantage of using aggregates rather than biofilms is that there is no need for scraping cells from a surface to recover RNA, which could potentially lead to changes in RNA levels as a result of sample processing.

2.2 Significance of coaggregation in oral biofilm formation

Communication among microorganisms is essential for initial colonization and subsequent biofilm formation requires physical contact between colonizing bacteria in oral cavity (Kolenbrander, 1988). Most of the cultured species of oral bacteria physically interact and adhere among themselves to display specific recognition patterns with their respective partner cells (Bos *et al.*, 1994). This recognition between genetically distinct cells in suspension and resultant clumping is called coaggregation (Kolenbrander, 1988). Recognition between a suspended cell type and one already attached to a substratum is termed coadhesion (Bos *et al.*, 1994). Bacteria sense and respond to their external environment and coaggregation has been shown to lead to specific gene regulation responses that impact on interspecies metabolic interactions, small molecule chemical signalling between species and cellular growth (Gibbons *et al.*, 1970). All of these interactions are important in biofilms and are believed to contribute to dental plaque development. These phenomena are likely to be less important in free-living environments as any cells that detach from the oral surface are washed away and swallowed. Loads of research demonstrated that coaggregation was a common phenomenon between a broad range of genera from dental plaque.

Coaggregation interactions are believed to contribute to the development of biofilms in two mechanisms. Firstly, by planktonic cells in suspension specifically recognizing and adhering to generically distinct cells in the developing biofilm. Secondly, by prior coaggregation in suspension of secondary colonizers followed by subsequent adhesion of this coaggregate to the developing biofilm. In both cases bacterial cells in suspension specifically adhere to cells in the biofilm in a process known as coadhesion in which these cells later become part of biofilm community (Figure 2.2).

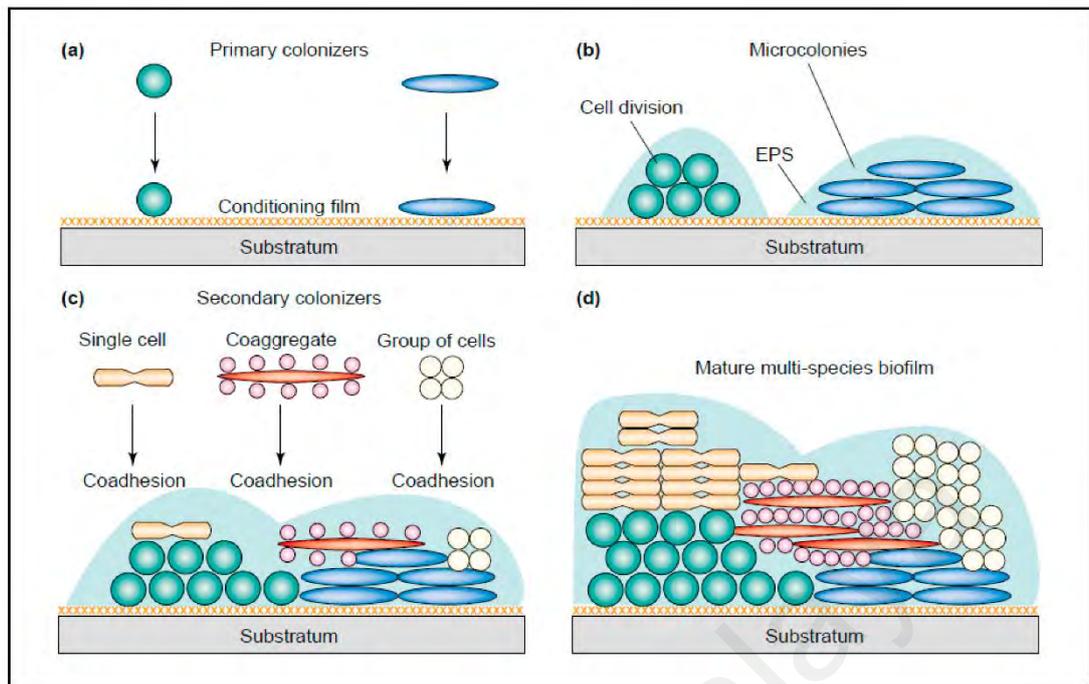


Figure 2.2: Diagram illustrating the possible roles of coaggregation in the development of multispecies biofilms.

(a) Primary colonization (b) cell growth (c) coadhesion of single cells and (d) maturation (Lof et al., 2017).

The partnerships in coaggregation are generally specific and primary colonizers can coaggregate with each other but only with later colonizers. The correlation between the coaggregation ability and the temporal sequence of bacterial integration into dental plaque very strongly implicates coaggregation as a process closely linked with development. The coaggregation adhesins expressed by the dental plaque indicates that they have evolved in parallel, as they are all different in composition from each other. It also show that many of the adhesins are not located on the cell wall but are found associated with external appendages thus enabling cells to make more effective contact with prospective partners. It is also believed that dental plaque microorganisms can express more than one coaggregation adhesin simultaneously on the cell surface.

Coaggregation among human oral bacteria was first described by Gibbons & Nygard, (1970). Coaggregation can be measured by several methods including visual inspection of clumps or coaggregates after mixing dense suspensions of two cell types (Gibbons *et al.*, 1970), turbidometric measurement of supernatant after slow speed centrifugation to

pellet the coaggregates (McIntire *et al.*, 1978), filtration through specific pore size to separate single cells from coaggregates (Lancy *et al.*, 1980), distribution of radiolabelled cells of one cell type in coaggregates and supernatant after slow-speed centrifugation (Kolenbrander *et al.*, 1986), and binding of a radiolabelled or fluorescently labelled cell type to partner cells immobilized on a nitrocellulose membrane (Lamont *et al.*, 1990). Microscopically, the clumps of cells formed consist of a network of interacting cell types. The recognition may be intrageneric, intergeneric or multigeneric in nature (Kolenbrander, 1989). In all three kinds of coaggregations, the cells appear to interact independently of other cells in the population. These interactions play critical role in biofilm architecture (Foster *et al.*, 2004; Guggenheim *et al.*, 2001; Kolenbrander, 2000). The stability of this biofilm depends on changes in environmental factors, as well as on interspecies communication (e.g., quorum-sensing signalling) that modulates the stability and composition of the biofilm (McNab *et al.*, 2003; Rickard *et al.*, 2006; Tomlin *et al.*, 2005). Microbial cell-cell interactions in the flora are believed to play an integral role in the development of dental plaque.

2.2.1 *Streptococcus gordonii*: Coaggregation and its significance

The first bacteria to colonize the pellicle on the tooth surface are streptococci and Gram-positive rods such as *Actinomyces naeslundii*. Within the first 4 h of plaque formation, *Streptococcus mitis*, *Streptococcus sanguinis* (previously *Streptococcus sanguis*) and *Streptococcus oralis* represent 60–90% of the cultivable streptococci (Nyvad *et al.*, 1990a). Streptococci are among the predominant flora in the oral cavity. Certain species of streptococci are predominant colonizers of the acquired pellicle, strongly adhering to the saliva-coated tooth surfaces. These pioneer colonizers initiate dental plaque and mediate the acquisition of other microorganisms to the biofilm.

Streptococcus gordonii is one of the early colonizers and is naturally transformable, it is widely used to model bacterial adhesion to different surfaces, including the tooth,

epithelial cells and other bacterial species. Some adhesion- related proteins (adhesins) of *Streptococcus gordonii* are regulated by environmental changes. With the coaggregation adhesins expressed by the dental plaque primary colonizer *Streptococcus gordonii* DL1 extensively studied, this strain carries some distinct proteins involved in coaggregation interactions. In some cases, the location of the coaggregation adhesins on the cell surface has been identified. The unnamed protein adhesin might use lipoteichoic acid as 'scaffolding' for its presentation to its partner microorganism (Clemans *et al.*, 1999) and it is responsible for intrageneric, galactoside-inhibitable coaggregation with other streptococcal species. The Has protein adhesin is a lectin-like protein associated with fibrillar structures and mediates adhesion to components of saliva, polymorphonuclear leukocytes and coaggregation with species of oral bacteria (Takahashi *et al.*, 2002; Yajima *et al.*, 2008).

Another protein also found on *S. gordonii* DL1, CshA, is a fibrillar protein anchored into the cell wall that can be seen by negative staining in the transmission electron microscope as peritrichous, 70-nm-long surface fibrils (McNab *et al.*, 1999; McNab *et al.*, 1996). These thin, sparse, peritrichous fibrils contain several repeating amino-acid sequences and it has been suggested that the coaggregation adhesin responsible for coaggregation with *Actinomyces naeslundii* is at the distal end of the fibrillar molecule (McNab *et al.*, 1999). Two other coaggregation adhesins, also expressed by *S. gordonii* DL1, mediate coaggregation to *Actinomyces naeslundii* and the periodontal pathogen *Porphyromonas gingivalis* and these adhesins have been designated SspA and SspB (Egland *et al.*, 2001). SspA and SspB (antigen I/II polypeptides) are multifunctional and, as well as mediating coaggregation, also mediate adhesion to salivary glycoproteins, fibronectin and collagen (Demuth *et al.*, 1996; Demuth *et al.*, 2002). The expression of these two adhesins is enhanced when *S. gordonii* DL1 is exposed to human saliva (Du *et al.*, 2000). The structural locations of SspA and SspB are unknown.

The *S. gordonii* DL1 coaggregation proteins are multifunctional and can often interact with host oral surfaces and host proteins. In addition to providing a surface for secondary colonizing microorganisms such as *Fusobacterium* spp. or *Porphyromonas gingivalis* (Periasamy *et al.*, 2009), Oral streptococci may form hubs for coaggregation-mediated interaction networks since many of them possess multiple adhesin proteins on the cell surface that mediate interactions with a wide variety of host or bacterial receptors (Kolenbrander *et al.*, 2010; Nobbs *et al.*, 2011). In these respects, the ubiquitous oral bacterium *S. gordonii* is able to interact with a wide range of coaggregation partners, including *Actinomyces oris* (Back *et al.*, 2015; Mohammed *et al.*, 2018) *S. oralis* (Demuth *et al.*, 1996), *F. nucleatum* (Foster *et al.*, 2004; Lima *et al.*, 2017), *Veillonella* spp. (Zhou *et al.*, 2015a), *Candida albicans* (Arzmi *et al.*, 2015). *S. gordonii* was used as model microorganism in order to further study community interactions within the oral cavity. Subsequently, to determine if transcriptional alterations present in *S. gordonii* in response to bacteria associated with it.

2.2.2 *Fusobacterium nucleatum*: Coaggregation and its significance

Fusobacterium nucleatum, a Gram-negative anaerobe in oral cavity forms a bridge between oral bacteria that do not naturally coaggregate with each other, it can aggregate with both aerobic and obligate anaerobic species, allowing the other species to live together (Kolenbrander, 1989; Metzger *et al.*, 2009). Coaggregation bridges usually refer to a structure of one bacterial species with two or more different receptors which can be recognized by different adhesions of two or more different bacterial species. Those bacterial species could aggregate by attaching to the first bacterial species; i.e., *F. nucleatum* is one of the best known coaggregation bridge species that facilitates streptococcal and obligate anaerobes aggregation. Without *F. nucleatum*, the number of obligate anaerobic species would decrease sharply in the oral cavity (Bradshaw *et al.*,

1998), even as thus it plays significant role in the development and maturation of dental plaque biofilm (Shanitzki *et al.*, 1998).

The coaggregation between *F. nucleatum* and other Gram-negative bacterial species are governed by lectin-carbohydrate receptor interactions (Rosen *et al.*, 2006). *F. nucleatum* possesses at least three distinct multifunctional adhesin molecules, which together mediate coaggregation with other genera from plaque and mediate adhesion to some host surfaces (Shanitzki *et al.*, 1998). *F. nucleatum* can help others survive by depleting acid and improving local environment and can produce organic acids by fermenting glutamate and aspartate providing neutral environment for other acid-sensitive bacterial species (Takahashi, 2003).

F. nucleatum supports *Porphyromonas gingivalis* growth by providing a capnophilic environment when growing in an oxygenated and CO₂ depleted environment. (Diaz *et al.*, 2002). In a two stage chemostat system, the survival of obligate anaerobic microorganisms in an aerated environment was improved by the coaggregation with *F. nucleatum* (Bradshaw *et al.*, 1998). *F. nucleatum* interspecies interactions and the identification of a number of binding partners, to date, only two fusobacterial large outer membrane proteins (OMPs), RadD and Fap2, have been characterized at a molecular level for their role as adhesins in binding to a variety of Gram-positive species (Kaplan *et al.*, 2014; Kaplan *et al.*, 2009), and *Porphyromonas gingivalis*, respectively (Copenhagen-Glazer *et al.*, 2015). Both, RadD and Fap2 are members of the autotransporter family of proteins, which are the largest known family of virulence factors expressed by Gram-negative bacteria (Henderson *et al.*, 2004). Autotransporters account for numerous biological functions including adhesion (Bullard *et al.*, 2007; Lipski *et al.*, 2007), cell-to-cell aggregation (Heras *et al.*, 2014; Klemm *et al.*, 2004) and biofilm formation (Sherlock *et al.*, 2004; Valle *et al.*, 2008), *F. nucleatum* was used a model microorganism in order to further study community interactions within the oral

cavity. Subsequently, to determine if transcriptional alterations present in *F. nucleatum* in response to bacteria associated with it.

2.2.3 *Veillonella parvula*: Coaggregation and its significance

The genus *Veillonella* consists of eleven Gram-negative species that are normal constituents of the mouth and gut of mammals (Dewhirst *et al.*, 2010). *Veillonella* are small strictly anaerobic cocci that lack flagella, spores and capsules, formerly belonging to the class clostridia, these firmicutes were then reclassified to the negativicutes class due to their peculiar cell wall composition that stains gram negative. This feature is unique for negativicutes since most firmicutes are gram-positive microorganisms (Mays *et al.*, 1982; Rogosa *et al.*, 1964) and are nutritionally fastidious by being able to utilize various organic molecules for energy, including numerous vitamins and amino acids. It was reported that oral *Veillonella*, found throughout the entire oral cavity, comprise as much as 10% of the bacterial community initially colonizing the enamel (Diaz *et al.*, 2006). It was reported that oral *Veillonella* play a central role as early colonizers that establish multispecies oral biofilm communities (Periasamy *et al.*, 2010). Members of the genus *Veillonella* gain energy from utilization of short-chain organic acids and have been isolated from the oral cavity and intestinal tract of humans and other animals (Keijsers *et al.*, 2008; Mancl *et al.*, 2013).

As mentioned earlier communication is most significant interaction in biofilm, one such method of communication is metabolic communication, *Veillonella parvula* is considered a 'benevolent organism' in dental caries because it metabolizes lactic acid, produced by streptococci, into weaker acids, such as propionic acid and acetic acid with a reduced ability to solubilize enamel (Gross *et al.*, 2012). Thus, a food chain could develop between these bacteria with the end-product of one microorganism serving as a source of energy for the other. Another form of biofilm communication is gene regulation, an example of this type of communication is *V. parvula* changing *S. mutans*

gene expression to make it more resistant to antimicrobial agents like chlorhexidine. *Veillonella* spp. creates an upregulation of the ribosome-associated chaperone *ropA* and many genes coding for ribosomal proteins to make this possible (Luppens *et al.*, 2008). Another study showed that interactions between these two species were shown to influence gene regulation *in vitro* forming dual species biofilms, where diffusible-signal exchange between coaggregation partners (Egland *et al.*, 2004).

In another study, adding *V. parvula* PK1910 (formerly *V. atypica* PK1910, (Hughes *et al.*, 1990) into *S. gordonii* - *S. mutans* mixed culture demonstrated rescue from effect of H₂O₂ (Liu *et al.*, 2011). In a very recent model it was proved that *Veillonella* species function as a physical bridge connecting pioneer and later colonizers through cell-cell coaggregation and also as middle player in a relay of nutrient flow from pioneer to later colonizers (Liu *et al.*, 2012; Zhou *et al.*, 2015a; Zhou *et al.*, 2015b). Two early-colonizing species of the dental plaque biofilm, *S. gordonii* and *V. parvula*, are an interesting pair of coaggregating microorganisms because their physiologies suggest that they participate in metabolic communication. Therefore, understanding the interactions between *S. gordonii* and *V. parvula* in the early stage of oral biofilm formation may be important to prevent these oral infectious diseases. However, detailed roles of oral *Veillonella* species in biofilm formation have not been fully clarified.

2.2.4 *Streptococcus oralis*: Coaggregation and its significance

Streptococcus oralis, one of commensal bacteria inhabiting oral cavity, belongs to the oral viridans group streptococci. *S. oralis* is a numerically important member of the oral microbiota, isolated from all intra-oral surfaces and a pioneer microorganism involved in the primary colonization of the dentition (Nyvad *et al.*, 1990b). It has been implicated as a potential causative microorganism of human cardiovascular diseases including infective endocarditis and atherosclerosis. *S. oralis* has coaggregation receptor polysaccharides (RPS) on the cell wall. The RPS function as receptors for the lectin-like

surface adhesins on other members of the oral biofilm community (Yoshida *et al.*, 2006). The structural, functional and molecular properties of streptococcal RPS support a recognition role of these cell surface molecules in oral biofilm formation. It is evident from studies of the cell wall polysaccharides on *S. oralis* and related viridans group streptococci that function as receptors for lectin-like adhesins on other members of the dental plaque biofilm community (Palmer *et al.*, 2003).

Recent work indicated that the introduction of *Candida albicans* to the oral cavity of mice enhances mucosal biofilm formation by *S. oralis* (Xu *et al.*, 2014), a ubiquitous commensal of oral mucosal surfaces in healthy humans (Diaz *et al.*, 2012a), which lacks the ability to form biofilms on its own *in vitro* and *in vivo* (Diaz *et al.*, 2012b). *S. oralis* in an *S. gordonii* dominated biofilm could enable *S. oralis* for introduction of a more favourable partner, such as *Actinomyces naeslundii*, to join the biofilm community (Palmer *et al.*, 2001). During initial dental plaque formation, the ability of a species to grow when others cannot would be advantageous, and enhanced growth through interspecies and intergeneric cooperation could be critical. Therefore, understanding the interactions between *S. gordonii* and *S. oralis* in the early stage of oral biofilm formation may be important to prevent these oral infectious diseases. However, detailed roles of oral *S. oralis* species in biofilm formation have not been fully clarified.

2.3 Importance of dual species coaggregation experiments

Biofilms can be formed by single, dual and/or multiple species of microorganisms and may constitute a single layer or three-dimensional structures. Mature biofilms represent a highly organized ecosystem ensuring the exchange of nutrients and metabolites. The dynamic process of biofilm formation is predominantly characterized by initial reversible attachment of planktonic cells, cell aggregation and colonization of surfaces. Extensive studies have been conducted on dual-species batch cultures and shown cooperation and beneficial interactions between bacteria growing as dual species

biofilms and some others have shown cooperation, competition and neutral interactions (Simoes *et al.*, 2007). While single-species biofilms have been studied extensively and there is need to know more about mixed-species biofilms and their interactions. However, in several studies, of which the majority focused on biofilms in the oral cavity, it has been demonstrated that different bacterial species in biofilms affect one another positively as well as negatively. Interactions beneficial to one or more strains or species include coaggregation of cells (Rickard *et al.*, 2003; Yamada *et al.*, 2005), conjugation (Ghigo, 2001), and protection of one or several species from eradication when the biofilm is exposed to antimicrobial compounds. Dual species biofilms are relatively stable and easy to handle in *in vitro*. When compared to mono-species biofilm, dual-species include the interaction between two species and provide a means for studying more complex microbial ecosystems. Under a more complex environment dual species biofilm seems to survive better when compared with mono-species (Kara *et al.*, 2006). The presence of second species may alter the composition and viscosity of the extracellular polysaccharide matrix and may slow down the penetration of antimicrobial into biofilm (Burmolle *et al.*, 2006). Physiological changes may occur when two species are able to transfer conjugative plasmids and thus share protective mechanisms (Behnke *et al.*, 2012) or support each other by complementing enzymes that are necessary to manage environmental challenges (Shu *et al.*, 2003). It is different from the *in vivo* multi-species plaque biofilms in both survival and pathogenic potential.

2.4 Metabolic interactions and associations of biofilms

Microbial species in the oral cavity form communities that establish a variety of micro-niches and inter- and intra species interactions. The interactions between different microbial species can be classified into three main groups: physical, chemical and metabolic interactions. Physical interactions, both non-specific (electrostatic and hydrophobic) and specific (protein-protein) interactions, occur among different types of

bacterial cells with almost all known oral microbes (also known as coaggregation and co-adhesion). These interactions make sure that bacterial cells are closely together and enable them to communicate through signaling and genetic exchange (Kolenbrander, 2000).

Oral bacteria benefit from the metabolic activity of nearby bacteria, and gain access to nutrients. Therefore, bacteria often co-localize with others that are metabolically compatible (Wright *et al.*, 2013). In the oral cavity bacteria often rely on metabolic cooperation of other bacteria resulting in web of metabolic exchanges between bacteria (Jenkinson, 2011). Within a community, different kinds of metabolic interactions can occur. Cooperative metabolic interactions occur when metabolites produced by one species are consumed by other (Bik *et al.*, 2010; Eglund *et al.*, 2004) and competitive metabolic interactions happen when two species consume the same resource (Jakubovics *et al.*, 2008b). Although quorum sensing is referred to as a form of chemical interaction, it could be seen as form of metabolic interaction. This process regulates numerous important biological processes in bacteria by releasing signal molecules (Suntharalingam *et al.*, 2005; Ueda *et al.*, 2009).

Microbial metabolic networks help to understand about factors that control community function, adaptation and development of biofilm. These microbial interactions also describe cooperative, competitive, or neutral interactions that occur between microbes. Complexities associated with understanding microbial networks include specifying taxonomic composition, genetic potential, metabolic activity and also functional adaptability of every microbe in community (Vanwonterghem *et al.*, 2014). Many metabolic pathways are used by bacteria and consequently different bacterial species require many different combinations of nutrients. Some bacteria can synthesise all of their components provided with sources of energy and certain elements (primarily C, O, N, H, P, S but also K, Mg, Ca, Fe, Zn and Mn). While some other bacteria need

many of their amino acids provided ready-made. There are many highly specific stratified environments in the oral cavity providing all sorts of nutrients and environmental niches that are created by the complex bacterial communities living in the mouth.

2.5 High through-put sequencing technologies and its applications

2.5.1 Sequencing evolution, achievements and limitations

First generation Sanger sequencing transformed biology to decipher the genetic blue print from single genes to whole genomes, from prokaryotes to eukaryote genomes. Sanger sequencing has been used for various applications such as mutation discovery, genotyping and Serial Analysis of Gene Expression (SAGE) for measuring gene expression levels; more significantly, it was used to complete the Human Genome Project (International Human Genome Sequencing, 2004). Currently, even with the availability of multiple high-throughput sequencing technologies (HTS), Sanger sequencing is still considered to be the gold standard method (with the highest raw base accuracy) and hence is still commonly used to validate variants or mutations identified using high-throughput sequencing technologies. Sanger sequencing had over many technical improvements in throughput, accuracy, safety, robustness and sensitivity for further scaling-up throughput and minimizing cost of sequencing to meet the growing demands of biological research community.

2.5.2 Key high-throughput sequencing platforms

High throughput sequencing technology is second/next generation sequencing technology launched by Roche/454 company, Illumina/Solexa company and ABI/Solid company. In contrast to Sanger sequencing, next generation sequencing technologies (NGS) are characterized by their ability to perform massively parallel sequencing of up to hundred millions of sequence reads. Currently, high-throughput sequencing has its applications in multi-level researches on genomics, transcriptomics and epigenomics

changing the approach of problems in basic and translational researches and created many possibilities. These technologies provide important data for research studies, such as disclosure of genetic information and regulation of gene expressions.

With appearance of Roche's 454 technology in 2005, Illumina's Solexa technology in 2006 and ABI's SOLiD technology, HTS has enormous developments, thus amounts of genetic information successively revealed allowing to explore the research of life in detail, to uncover the huge diversity of novel genes that are previously inaccessible, to understand nucleic acid therapeutics, to better integrate biological information for a complete picture of health and disease at a personalized level and to move to advance. Therefore, a number of bioinformatics methods and softwares has been created to accelerate HTS was to be applied for researches on genomics, transcriptomics and epigenetics. Further, it has also brought new challenges for bioinformatics in effective processing and analyzing massive data and extracting valuable bio-information from it, and also an important factor to decide whether or not HTS technology plays a major role in the scientific exploration.

The Core idea of three major sequencing platforms, Roche/454 pyrosequencing, Illumina/Solexa sequencing of polymer synthesis and ABI/SOLiD ligase sequencing technologies is sequencing by synthesis. When generating a new complementary strand of cDNA, they either add normal dNTP through enzymatic cascade reaction to catalyze substrates to excite fluorescence (Roche/454), or directly into the fluorescently labeled dNTP (Illumina/Solexa) or semi-degenerate primers (ABI/SOLiD), then when generating or connecting to synthesize the complementary chain, the substrates will release fluorescent signal. By capturing the optical signal and converting to a sequencing peak, it can be converted again to the sequence information of complementary strand.

2.5.3 Illumina sequencing platform and its significance

Illumina produces a suite of sequencers (MiSeq, NextSeq and the HiSeq series) optimized for a variety of throughputs and turnaround times. The MiSeq and HiSeqs are the most established platforms. The MiSeq is designed as a fast, personal benchtop sequencer, with runtime as low as possible and outputs intended for targeted sequencing and sequencing of small genomes. The HiSeq series, on the other hand is engineered for high-throughput applications, yielding output of 1TB data. While many sequencing platforms are available, Illumina's platforms (<http://www.illumina.com/>) currently dominate the HTS market (Quail *et al.*, 2012).

The sequencing process involves clonal amplification of adaptor-ligated DNA fragments on the surface of a glass slide (Bentley *et al.*, 2008). Illumina's bridge amplification allows for generation of small "clusters" with an identical sequence to be analyzed. Clusters formed on an Illumina flow cell create multiple hybridization steps allowing multiple sequencing start points (Figure 2.3). This allows the sequencing of both ends of the original template molecule known as paired-end sequencing. Sequencing information from paired-end reads play an important role in Illumina's technology by increasing the output from a sequencing run, identifying variants in RNA-seq, and to deduplicate (remove duplicate copy) reads originating from the same original template.

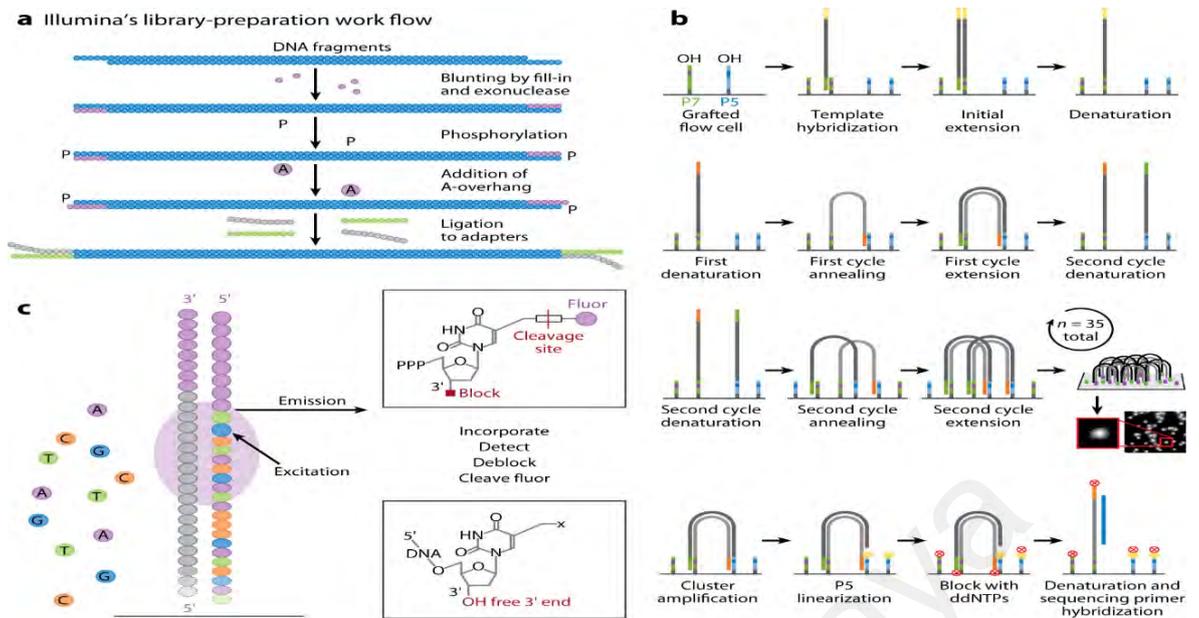


Figure 2.3: Illumina sequencing workflow.

(a) Illumina library preparation construction process (b) Illumina cluster generation by bridge amplification (c) Sequencing by synthesis with reversible dye terminators (Mardis, 2013).

In Illumina sequencing however, all four nucleotides are available during incorporation which can lead to an overall substitution error rate of 0.11% (Minoche *et al.*, 2011). Both Ion Torrent (<http://www.iontorrent.com/>) and 454 (<http://www.454.com>) employ the use of polymerase chain reactions to amplify DNA within an emulsified droplet. Sequencing information is either correlated with light (in 454) or hydrogen ion (in Ion Torrent) detection during each nucleotide incorporation event. If multiple nucleotide incorporation events occur, this is interpreted as stretches in the sequence of particular nucleotide (homopolymers). All HTS technologies have difficulties sequencing homopolymers, but using Ion Torrent and 454, it could prove otherwise as the nucleotides lack a blocking moiety, resulting in entire incorporation of homopolymers during sequencing run.

2.5.4 Dual RNA-seq approach for studying oral bacterial coaggregations

The advent of second generation sequencing technologies has created many opportunities to improve functional genomics experiments, including quantitative gene expression studies. Most previous transcriptional analysis methods have relied on

hybridization of targeted oligonucleotides to particular loci for their sequence specificity: either primer binding to targeted oligonucleotides to particular loci for their sequence specificity: either primers binding to target cDNA in quantitative reverse transcription polymerase chain reaction (qRT-PCR), labeled probes binding to RNA in Northern blotting or hybridization of cDNA to probes on microarray chips. The Next generation sequencing (NGS) technologies applied to analysis of the transcriptome is referred to as RNA Sequencing (RNA-Seq) or Whole Transcriptome Shotgun Sequencing (WTSS) as it is quantitative, and can generate a global profile of an microorganism in a single analysis.

The basic workflow of RNA-Seq has several steps. The first step is to convert the population of RNA extracted from an microorganism under a test condition to a library of cDNA fragments with adaptors attached to one or both ends. The adaptors are short sequences that are specific to a platform. The second step is to sequence the molecules (cDNA with adaptors), which can be done with or without amplification (e.g. PCR), in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing). Single-end read sequencing involves sequencing fragments from one side, while pair-end sequencing fragments from both sides (Figure 2.4). Any high-throughput sequencing technology platforms explained above can be used for RNA-Seq. The last step is either mapping the short sequences to a reference genome or reference transcripts, or assembling the sequences *denovo* without a reference genomic sequence, to produce a genome-scale transcription map that consists of the transcript structure and/or level of expression of each gene. The short sequences referred to as reads, are typically 30-400 bp, depending on the sequencing technology used.

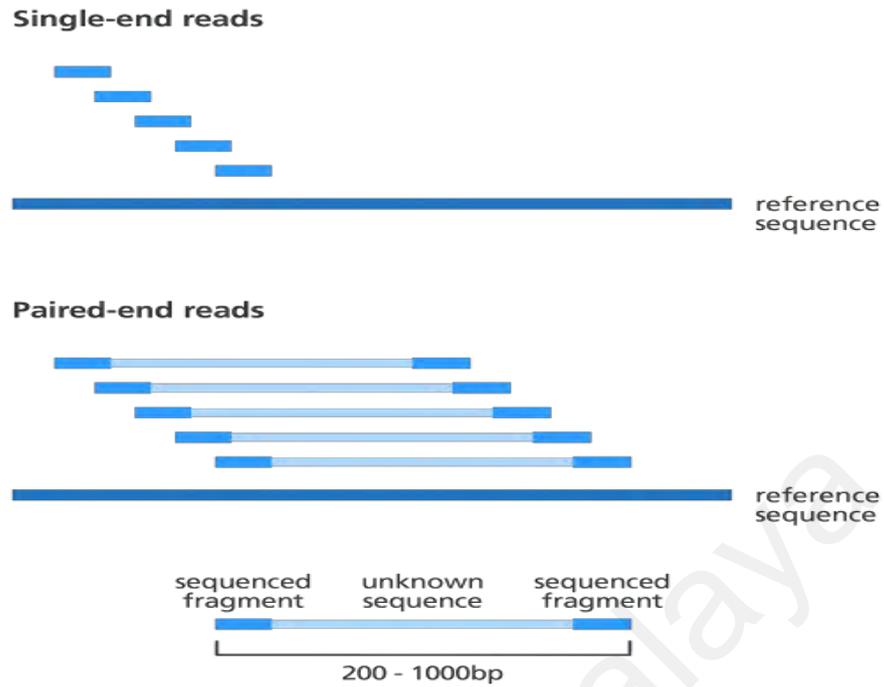


Figure 2.4: Single end sequencing and Paired end sequencing.
 (https://ivanek.github.io/analysisOfGenomicsDataWithR/05_Overview_mRNAs_eq_html.html)

RNA-Seq is different in principle, this quantifies gene expression by sequencing short strands of cDNA, aligning sequences obtained back to the genome or transcriptome, and counting the aligned reads for each gene (Figure 2.5). RNA-Seq data display more sensitivity to differential expression tests with number of identified genes generally larger. RNA-Seq sits in a unique position capable of analyzing an entire transcriptome without prior understanding of the organism's biology. As the technology is very efficient, there are still many technical challenges, which must be overcome as best practices are constantly being re-evaluated. Sample handling, instrumentation and data processing are all under scrutiny, but the current state of the art still has much to offer.

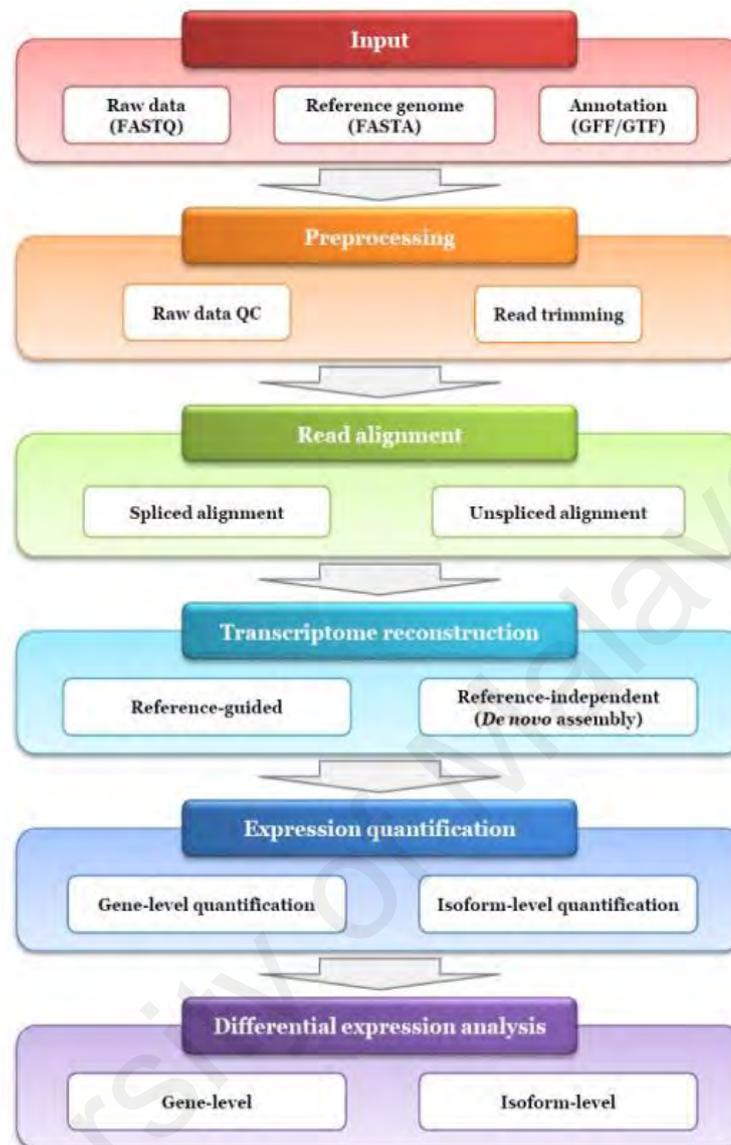


Figure 2.5: RNA-Seq Workflow
(Adapted from (Yang *et al.*, 2015)).

RNA-Seq technology can provide both a qualitative and quantitative view of an organism's transcriptome. A qualitative experiment is focused on identifying what features are present in expressed transcripts - isoforms, which are expressed in a particular scenario may not be present in another. A quantitative experiment deals with quantity of expression. The relative differences between transcript abundances are measured and significant changes are noted. While the two results are similar, achieving the highest quality data on a qualitative experiment differs from a quantitative one. Quantitative experiments require replicate measurements, with a strong preference for

biological replicates, to inform on the sample variance to identify between group differences (Love *et al.*, 2014).

It is well accepted that the study of inter-species is fundamental to understanding the biology of biofilm formation (Parashar *et al.*, 2015; Rickard *et al.*, 2003). In the oral cavity bacteria do not exist isolated from each other, but constantly interact. Bacteria can sense the presence of interaction partners by a range of regulatory networks specifically react by changing the expression of many of their genes. Technological advances in NGS over the recent years now allows the application of RNA sequencing to two species at the same time (dual RNA-Seq) (Westermann *et al.*, 2017; Westermann *et al.*, 2012; Westermann *et al.*, 2018), and thus to directly the gene expression of two interacting species without the need to physically separated cells or RNA.

Dual RNA-Seq allows simultaneous monitoring of gene expression changes without disturbing the complex interactions that define dual-species coaggregations. dRNA-Seq can report data for two organisms from the sample while providing insight into novel interaction dynamics such as gene expression changes in one organism can be correlated with the response of other to capture crucial events that signify the dynamics mechanisms of coaggregation formation in biofilms. This work provides a detailed bioinformatics analysis pipeline for a typical dual species analysis, dRNA-Seq is a powerful, economical, sensitive and species independent platform for investigating the gene expression dynamics of dual species interactions (Westermann *et al.*, 2012).

Many recent studies have applied dRNA-Seq in different types of interspecies relations such as host-pathogen, mutualistic and commensal interactions. For example, *S. pneumoniae* was shown to repress the innate immune response of human epithelial cells while activating its own sugar importers upon sensing host-derived mucins (Aprianto *et al.*, 2016). A murine *Yersinia pseudotuberculosis* infection model unveiled the importance of carbon storage regulator system for virulence (Nuss *et al.*, 2017). Oral

streptococci promote growth, hyphae formation (Dutton *et al.*, 2016) and virulence of *Candida albicans*.

2.5.5 General workflow of RNA-Seq data analysis and related bioinformatics tools

An RNA-Seq platform provides sequencing of paired-end reads with length of about 100-250 base pair (bp), and a numbers of reads ranging from 5 to 60 million per sample, depending on the goal of the sequencing. This massive amounts of data makes data management and interpretation challenging and requires a series of computational tools for the different research purposes. Figure 2.5 shows the major steps of RNA-seq data analysis. Following such a work flow, the massive information started as millions of short-read sequences is transferred into oranzed biological networks, which can visualized and analyzed easily.

The first step in RNA-Seq Analysis pipeline is to perform quality control on the reads genereated in FastQ format. Various tools can be used to determine the read quality of each sample. One such tool is FastQC (Babraham Bioinformaitcs, Cambridge, UK), that provides a comprehensive report with multiple graphs to assess read quality. Two graphs of particular interest display the quality score as a function of the read position and the distribution of quality scores over all sequences. A common feature of the read generated on the Illumina platform is that they tend to decrease in quality as move towards the 3'-end. Too many mismatches between the read and the reference genome can prevent the read from mapping to the genome.

Read alignment is the second step of RNA-Seq analysis. Reads after pre-processing need to be aligned to the reference genome in order to measure levels of gene expression or to identify any novel transcripts. Gene expression can be measured by estimating the depth of coverage in the read mapping. There are many bioinformatics tools available for this purpose like TopHat (Trapnell *et al.*, 2009), STAR (Dobin *et al.*,

2015) and MapSplice (Wang *et al.*, 2010) which were developed with different foci, but none of them is dominant with an advantage of output format of these aligners is uniform called “Sequence Alignment/Map” (SAM/BAM). One of the main concerns with RNA-Seq over conventional methods is that that sequenced reads are shorter and mapping of millions of short reads to reference genome is a demanding task.

Tophat2 (Kim *et al.*, 2013) is an open source, fast splice junction mapping tool that aligns RNA-Seq reads to a genome and then analyses the mapping for splice sites. This program is implemented in C++ and Python and runs on Linux and OS X. Tophat is built on the ultrafast short read mapping program called Bowtie (Langmead *et al.*, 2012). Tophat can align reads to the reference genome at the rate of ~2.2 million reads per CPU hour. Tophat workflow consists of two phases. It first maps the reads to the whole genome using Bowtie. Bowtie indexes the reference genome using a memory efficient scheme called Burrows Wheeler transform. This enables Bowtie indexes to scan reads against reference genome on a workstation.

For each read, Bowtie reports one or more alignments allowing a small number of mismatches (default is 2) in the high quality end of the read. Mismatches are allowed to accommodate genuine differences between the reads and the reference due to sequencing errors. The number of alignments reported by Bowtie can be controlled by option -k. The default value of k is 10 and all other alignments more than 10 are suppressed. This helps in excluding alignment of failed reads to low-complexity regions. Tophat then assembles the mapped reads into a consensus sequence and infers islands of contiguous sequences as putative exons. The alignment of reads across the genome is helpful in identifying polymorphisms between the sample and reference genome. Alignments are used for quantification of genes and the transcript expression. The number of reads that align to a transcript is proportional to the transcript's abundance.

The count of reads at a base pair is defined to be the number of reads that start at base pair. This process typically involves first counting number of fragments in a given sample which align to the target features, followed by normalization. This methodology involves simply counting all fragments which aligned end-to-end entirely within an exon as a hit, and avoiding all others. This method counts all features overlapped by any portion of a sequence fragment, which improves counting accuracy. HT-Seq count is much more effective with higher quality sequences, especially paired-end sequence information, since fragments can then be assigned more accurately to transcripts, which increases the robustness of the results.

Differential expression analysis pursues the identification of genes whose transcripts show substantial changes in abundance across experimental conditions, such as differences between strains. At present, RNA-Seq is the main option for differential expression analysis, and there are already many algorithms developed for this purpose, including parametric and non-parametric methods.

The objective of parametric approaches is to estimate the model in each biological condition, and further compare different statistical distributions to detect differentially expressed genes. Many methods adopt negative binomial distribution to account for genome-wide read counts, which provide better sensitivity and specificity based on other distributions. Negative binomial (NB) distribution methods like DESeq (Anders *et al.*, 2010; Love *et al.*, 2014), edgeR (Robinson *et al.*, 2010) and their latest releases are the favorable software tools for gene expression differential expression analysis.

The DESeq algorithm proposed to model read count data of gene with a NB distribution, where the parameters of NB distribution were determined by the mean and standard deviation of the gene among all samples within the same condition. After estimating the NB parameters in both experimental conditions, differential expression analysis between the conditions were tested. In DESeq algorithm, a Fisher's exact test

approach is used to estimate the P-value of each gene. In short, suppose we have x and y reads of a gene in each condition, we compute every possible $p(a,b)$, where the sum of the variable a and b equal to K total ($K_{total}=x+y$). By assuming the independence of two test conditions, we have $p(a,b) = Pr(a)Pr(b)$, where $Pr(a)$ and $Pr(b)$ are the probabilities of NB distribution that we have estimated for each condition. Therefore, the P-value is the proportion of the sum of possible probabilities less than the probability of actual read counts among the sum of all probabilities.

There are also other parametric methods that employ Bayesian empirical approach (Hardcastle *et al.*, 2010; Leng *et al.*, 2013) or linear model (Law *et al.*, 2014). Other non-parametric methods have also been developed to avoid poor parameters estimation or the violation of distributional assumptions that leads to unreliable results, such as SAMSeq that uses the rank of the expression values in Wilcoxon statistic and NOISeq that creates a noise distribution of count changes by contrasting fold-change and absolute expression differences.

For a deeper interpretation of mechanisms of biological function and to reduce the complexity of large number of genes, significantly expressed differential genes were grouped representing biological pathways related to the experimental difference between two conditions. There are many databases that provide pre-determined pathways, such as KEGG (Kanehisa *et al.*, 2016) which places genes in metabolic and other pathways and the Gene ontology (Gene Ontology, 2015) which includes biological processes, cellular function and molecular functions. These databases contain curated pathways related to metabolism, diseases and genetic information processing.

BLAST2GO is a tool for automatic functional annotation of DNA or protein sequence data mainly based in GO vocabulary. It integrates the gene ontology annotation based on similarity searches with statistical analysis and highlights visualization of the putative functions in the form of direct acyclic graphs and pie

charts. BLAST2GO uses BLAST to search against public or private sequence databases, mapping homologous sequences to GO terms. The program extracts GO terms from BLAST hit results and maps the annotations by association.

2.5.5.1 FASTQ format

FASTQ format is the successor of FASTA format that includes quality information for each in the file. FASTA is one of the most well-known file formats used to represent and store nucleotide sequences, in which they are detailed by a sequence of characters. This text-based format is mostly used by major databases like NCBI as the input method to query their databases. FASTQ format is the standard for storing high-throughput sequencing technologies. Most of the next generation sequencing data analysis software packages, as well as packages that we use in this study, are only compatible with FASTA/FASTQ format.

FASTQ format consists of 4 line types per sequence. The first line is the title line. It starts with '@' and contains the actual sequence identifier and optional description. The second line the sequence line which contains the actual sequence information. The third line starts with a '+' that signals the end of the sequence line. It optionally includes the full repeat of text in the title line. The last line called the quality line contains line contains the quality values that single-byte encoded. The length of the quality line is the same as that of the sequence line with a score for each letter in the sequence.

2.5.5.2 Quality scores

A quality score Q is the measure of the base calling accuracy. Base calling error probability (P) is the probability that a given base is called incorrectly by the sequencer. The Q score is a metric that is logarithmically related to P . For example: (Q30 ensure 99.9% accuracy).

The PHRED quality score Q is defined as:

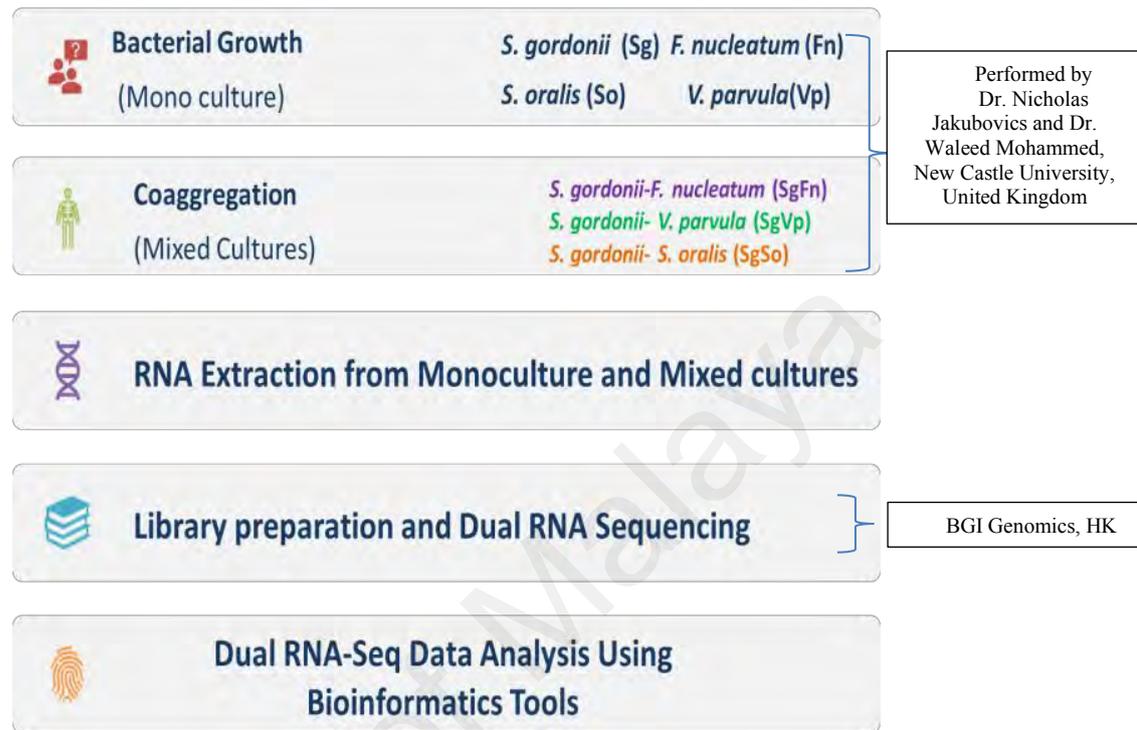
$$Q_{\text{PHRED}} = -10 * \log_{10}(p).$$

2.5.5.3 SAM/BAM format

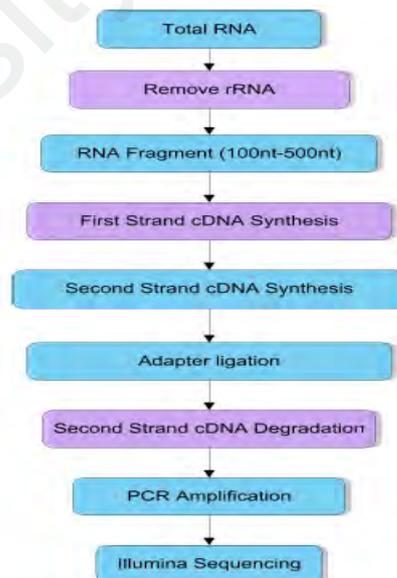
SAM stands for Sequence Alignment Map. It is a text format for sorting large nucleotide sequence data and alignments. SAM file formats are the output from mapping tools and aligners that reads FASTQ files and align the reads to a reference genome. BAM file format is the binary version of SAM. A SAM file is a tab delimited text format that contains a header section and an alignment section. The header section contains generic information about the entire file. It starts with @ followed by a two letter record type code. Each data field is on the format TAG:VALUE where TAG is a two letter string that defines the content and the format of VALUE. The alignment section shows the alignment of the sequenced reads to reference genome and also the quality of alignment. It contains 11 mandatory fields. All 11 fields must be present, but with a value of 0 or * if the corresponding information is not available and must always appear in the same order.

CHAPTER 3: MATERIALS AND METHODS

3.1 Study Design



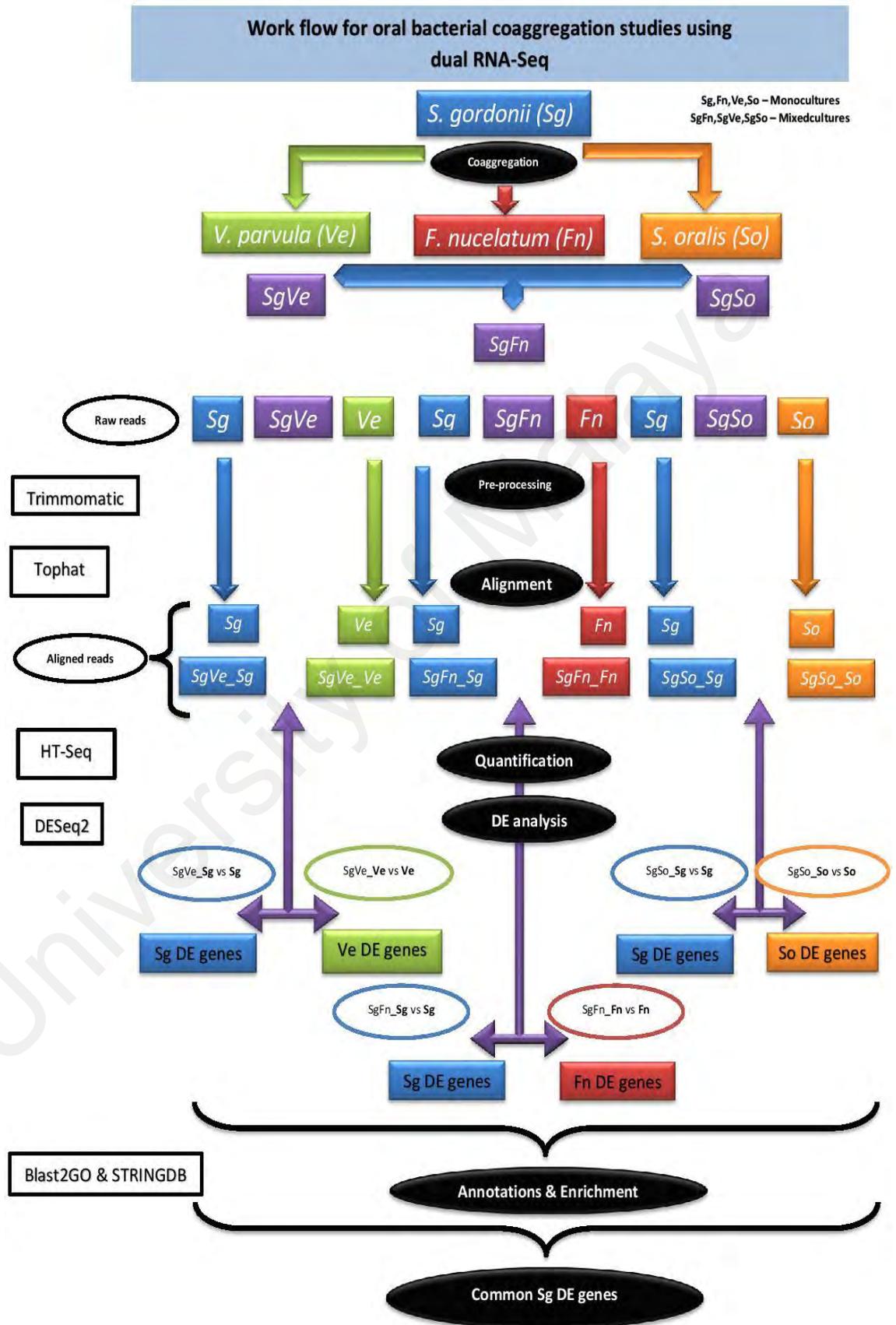
3.2 Library preparation and Dual RNA Sequencing



3.3 Bioinformatics analysis of transcriptomes

A workflow designed for dual RNA-Seq analysis to compare monoculture sample reads with coaggregate sample reads was designed as shown (Figure 3.1)

Figure 3.1: Work flow implemented for oral bacterial coaggregation studies using dual RNA-Seq approach



3.3.1 RNA-Seq datasets

Initially, the RNA-Seq transcriptional analysis carried out in three biological replicates of each sample for mixed culture (SgFn, SgVp and SgSo) and monoculture (Sg, Fn, Vp and So) totalling 21 sequencing libraries (batch -1). Less number of reads and outliers were observed for some samples. Therefore, samples were resent for sequencing but with little less number of replicates which yielded good number of reads (batch-2) due to low funds.

3.3.2 RNA-Seq data generation, pre-processing and quality trimming

RNA-Seq reads received from BGI in raw FASTQ format from batch-1 and batch-2 used for analysis include three biological replicates for each monocultures *S. gordonii* (Sg), *Veillonella parvula* (Vp), *Fusobacterium nucleatum* (Fn) and five replicates for *S. oralis* (So). Three replicates for mixed cultures *S. gordonii*- *F. nucleatum* (SgFn) and *S. gordonii*-*V. parvula* (SgVp) and five replicates for mixed culture *S. gordonii*-*S. oralis* (SgSo).

Trimmomatic-0.36 (Bolger *et al.*, 2014) was used to remove paired-end adapters from samples. All samples were trimmed using a phred score, a quality measurement assigned to each base sequenced, with a cutoff of an average phred quality score > 20 over a four-nucleotide sliding range. This allows one base to be of low quality within the window of 4. The resulting paired-end trimmed paired and unpaired reads were re-examined for quality, to verify trimming, using FastQC and used for downstream analysis.

3.3.3 Alignment of pre-processed RNA-Seq reads

Pre-processed reads for monoculture samples were mapped to genomes of reference genomes for *S. gordonii* and *V. parvula* that were downloaded from NCBI with accession numbers NC_009785.1 and NZ_CM002135.1 respectively and for *S. oralis* (from Dr. Jakubovics's lab) with default parameters using TopHat2. For coaggregate

samples, the cleaned reads were aligned in two rounds independently to the individual genomes. Mapping statistics were verified using SAMtools flagstat.

3.3.4 Gene expression quantification

Reads mapped to genomic features were counted using HTSeq-count (Anders *et al.*, 2015). Gene expression was quantified using HT-Seq with the 'union' mode. If only one mate of a read pair mapped uniquely within the transcriptome, that read was kept and was counted equally as a read pair for performing gene expression counts, as both a single-end read and paired-end read represent one RNA molecule.

3.3.5 Differential expression analysis

To study the significance of coaggregation between oral bacteria, we examine the variability between mixed and monoculture samples, R statistical packages of DESeq and DESeq2 approaches were used to normalize the data and calculate changes in RNA abundance and best significant results from one of two approaches were considered from down stream analysis. The DESeq programs were selected as they are most conservative for 3 replicates and were least effected with increasing quantities of differentially expressed genes (Zhou *et al.*, 2014). The variance stabilizing transformation function in DESeq packages was used to adjust for unequal sample library sizes. The three replicates for each sample were pooled for executing statistical analyses. Independent filtering was used to remove genes with very low counts and high dispersion, based on a thershold of the mean of the normalized counts, regardless of treatment (Love *et al.*, 2014). To reduce risk of type-I error, a Benjamini Hochberg correction was applied ($\alpha = 0.05$) (Benjamini *et al.*, 1995). *P-values* calculated from DESeq packages, and Benjamini-Hochberg adjusted *P-values* ≤ 0.05 were deemed significant. All analysis code is shown in Appendix A.

3.3.6 Gene enrichment analysis

Gene enrichment analysis was performed on significantly expressed differential genes in response to coaggregation using gene ontology searches and Blast2GO version 5 (Gotz *et al.*, 2008). A similarity search was performed using BLASTX against the NCBI non-redundant database with the parameter of E-value of $1e^{-3}$, word size of 6 and HSP length cut-off of 33. The blast result was further being imported into Blast2GO for gene annotations and further GO analysis using E-value hit filter of $1e^{-6}$. Fisher exact test was further performed to identify the functional enrichment categories among the differentially expressed with *p-value* cut-off of 0.05. The final annotation result was categorized with respect to Biological process, Molecular Function and Cellular component.

3.3.7 Putative protein interaction network construction

The STRING version 10.5 database was used to predict if there were any functional associations of differentially regulated significant genes. The search tool for retrieval of interacting genes/proteins (STRING) was used to identify known and predicted interactions (derived from four sources: genomic context, high-throughput experiments, co-expression and previous literature).

3.4 RT-qPCR validation of RNA-Seq data

Quantitative reverse transcriptase PCR was carried out using two steps for reverse transcription and PCR was performed by Dr. Jakobovics. The reverse transcription step was performed using the QuantiTect Reverse Transcription Kit (Qiagen, Valencia, CA, USA). RNA samples extracted from *S. gordonii*, *V. parvula*, *S. oralis* and *F. nucleatum* cells were converted into cDNA according to manufacturer's instructions, with the modification that random hexamers (Bioline, Taunton, MA, USA) were used in place of QuantiTect oligo-dT primers (Jakobovics *et al.*, 2015). The cDNA samples were stored at -20°C until analysis by real-time PCR.

All RT-qPCR was performed using SYBR Green dye from the SensiMix SYBR No-ROX kit (Bioline). Each reaction contained 25 µL consisting of 0-15 ng template cDNA, forward/reverse primers each at 2 µmol/L, 12.5 µL Power SyBr Green PCR mix and sterile deionized water. Standard curves consisting of serial 10-fold dilutions of *S. gordonii*, *F. nucleatum*, *V. parvula* and *S. oralis* chromosomal DNA, no template control and 'no RT' negative controls were included on each plate. Reactions were carried out using a DNA Engine Opticon 2 (BioRad, Watford, UK) as follows: (a) 95°C for 10 5 minutes; (b) 95°C for 10 seconds (c) 60°C for 30 seconds; (d) plate read;(e) repeat from step 2a further 39 times; (f) melting curve from 55 to 90°C, read every 1°C, hold for 5 seconds. As a reference, the 16S rRNA gene of any species of reasearch was measured and used to normalize the data. To assess the specificity of the reactions, metl curve analysis were performed on all samples and selected products were also assessed by agarose gel electrophoresis. Three biological replicates were performed for all RT-qPCR reactions.

3.5 Common genes/pathways expressed in *Streptococcus gordonii* in response to coaggregation with *Fusobacterium nucleatum*, *Veillonella parvula* and *Streptococcus oralis*

Using a Venn diagram, the *S. gordonii* genes that were common to three pairs of comparisons and the genes that were shared between any two pairs were identified. The genes commonly expressed from the three pairings were further investigated with STRING DB analysis to explain possible commo genes and pathways.

**CHAPTER 4: TRANSCRIPTOMIC RESPONSES TO COAGGREGATION
BETWEEN *STREPTOCOCCUS GORDONII* AND *FUSOBACTERIUM
NUCLEATUM***

4.1 Coaggregation of *Streptococcus gordonii* and *Fusobacterium nucleatum*

Coaggregation between *Streptococcus gordonii* DL1 and *Fusobacterium nucleatum* was estimated with well established scoring system (Mohammed *et al.*, 2018). Macroscopic aggregates were seen with mixing equal concentrations of each species in both coaggregation buffer or human saliva and were scored ‘4+’ in the visual assay, with a complete clear back ground. Earlier, it has been shown that *F. nucleatum* 25586 coaggregate with a variety of *Streptococcus*. spp. in saliva diluted in growth medium (Merritt *et al.*, 2009) and also *F. nucleatum* ATCC 23726 coaggregated strongly with *S. gordonii* DL1 in coaggregation buffer (Lima *et al.*, 2017). Hence, it was not surprising that *S. gordonii* DL1 and *F. nucleatum* 25586 would form strong coaggregates in buffer or saliva.

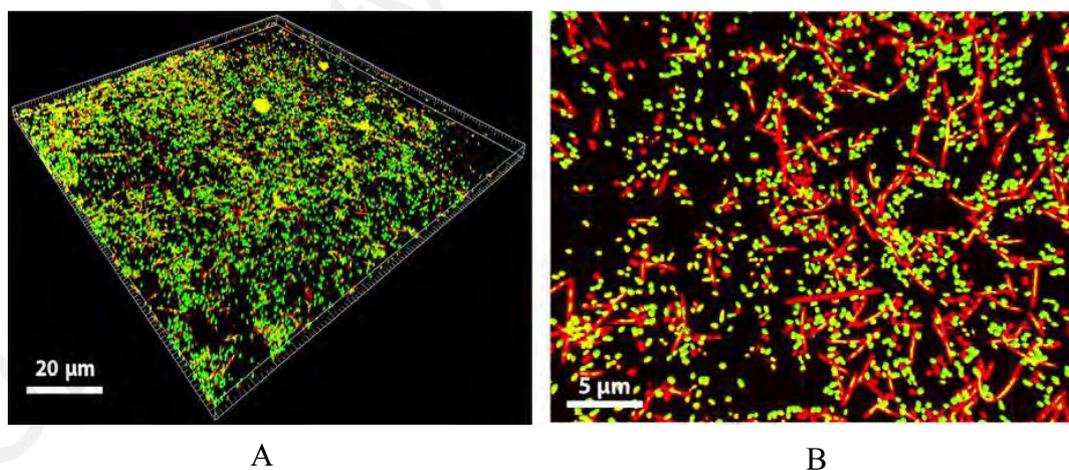


Figure 4.1: Confocal scanning laser microscopy analysis of coaggregation between *Streptococcus gordonii* and *Fusobacterium nucleatum*.

S. gordonii cells were pre-labelled with Pico Green and appeared green, whereas *F. nucleatum* cells were labelled with propidium iodide (red). A, Three-dimensional overview of coaggregates. B, an enlarged single slice through the coaggregate, showing close association between *S. gordonii* and *F. nucleatum*. Note that *F. nucleatum* cells appeared to have up taken some of the DNA-binding Pico Green dye since they appeared yellow in the center of the cells.

Three-dimensional (3-D) CSLM visuals showed *S. gordonii* and *F. nucleatum* cells interspersed throughout the coaggregate structure (Figure 4.1A). It was confirmed that the DNA-binding dye PicoGreen had transferred from *S. gordonii* and *F. nucleatum* as a yellow color was seen in most of the red stained *F. nucleatum* cells (Figure 4.1 B). Monoculture cells of *F. nucleatum* stained with propidium iodide were completely red. In spite *F. nucleatum* cells were clearly visualized to be in close proximity to *S. gordonii* cells. Whilst coaggregation between many different oral bacteria, and between bacteria from other environments, had been demonstrated (Katharios-Lanwermyer *et al.*, 2014; Stevens *et al.*, 2015), very little is known about the spatial positioning of cells from different species within these coaggregates. With fluorescence microscopy, it was shown that *S. gordonii* and *A. oris* cells were distributed quite evenly throughout coaggregates, and *S. gordonii* appeared to sense and respond to these interactions (Jakubovics *et al.*, 2008a). Our CSLM images of *S. gordonii* and *Fusobacterium nucleatum* showed a similar distribution of cells indicating that there was potential for the different species to sense one another.

Interactions between *S. gordonii* and *F. nucleatum* were more closely visualized by TEM (Figure 4.2) and *S. gordonii* cells were seen to be closely associated with *F. nucleatum*, consistent with an adhesin-receptor-driven interaction (Kaplan *et al.*, 2014; Kaplan *et al.*, 2009; Lima *et al.*, 2017). These interactions were observed at high-resolution in three dimensions by analyzing coaggregates using serial block face sectioning SEM (SBF-SEM) method involving usage of serial sections of a fixed depth (~70 nm) through an embedded sample using ultramicrotome contained within an SEM. Individual cells were color coded and identified using image analysis tools confirming close interactions between *S. gordonii* and *F. nucleatum* throughout the 3-D coaggregate structure. Therefore, these close interactions provide clear opportunities for cells of the different species to sense and respond to each other.

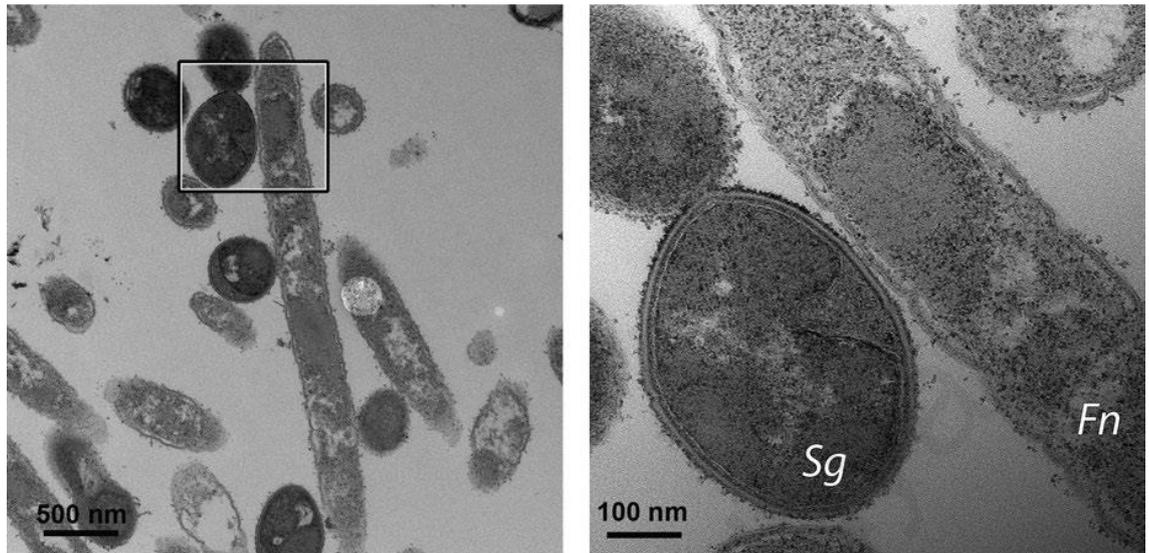


Figure 4.2: High Resolution Analysis of Interactions between *Streptococcus gordonii* and *Fusobacterium nucleatum* cells using Transmission Electron Microscopy.

The right panel shows an enlargement of approximately the area highlighted by a box in the left panel. *S. gordonii* cells (Sg) could be distinguished from *F. nucleatum* (Fn) by their shape and thick cell walls. Individual *F. nucleatum* cells were often intimately associated with multiple *S. gordonii* cells (left panel).

4.2 Dual RNA-Seq data analysis

4.2.1 RNA-Seq data generation and pre-processing

Initially, the RNA-Seq transcriptional analysis carried out in three biological replicates of each sample for mixed culture (SgFn) and monoculture (Sg,Fn) totalling 9 sequencing libraries yielded less number reads for one of the Fn replicate and two SgFn replicates (batch -1). So reads from all samples of batch-1 were ignored and resent for sequencing which yielded good number of reads (batch-2). In total, second batch Illumina sequencing from the three mixed *S. gordonii*, *F. nucleatum* biological replicates yielded a total of 107,348,188 raw reads and for monoculture samples *S. gordonii* and *F. nucleatum* yielded 109,436,214 and 108,200,956 reads respectively (Table 4.1). Removal of low quality bases and adapter content using Trimmomatic-0.36 software retained $\geq 92\%$ reads for all the samples. The sequence reads of all samples

were deposited in the NCBI sequence read archive (SRA) as study under the Bioproject accession number PRJNA472829.

Table 4.1: Alignment statistics of mixed and monoculture transcriptomes in *S. gordonii* and *F. nucleatum* coaggregation experiments

Sample name	Raw reads	Pre-Processed Reads	Mapped Reads
Monocultures			
Sg1	41,110,980	37,825,146	37,102,616
Sg2	33,579,910	31,086,324	30,560,777
Sg3	33,510,066	31,043,394	30,567,258
Total reads of <i>S. gordonii</i>	108,200,956	99,954,864	98,230,651(98%)
Fn1	36,701,358	34,135,396	31,979,171
Fn2	38,944,088	36,040,506	34,375,785
Fn3	33,790,768	30,905,094	29,497,626
Total reads of <i>F. nucleatum</i>	109,436,214	101,080,996	95,852,582(95%)
Coaggregates			
SgFn1	32,695,768	30,212,058	20,486,984 (Fn) 8,700,448 (Sg)
SgFn2	41,181,780	37959056	20,247,689 (Fn) 16,513,741(Sg)
SgFn3	33,470,640	30,313,554	20,828,717(Fn) 8,399,455 (Sg)
Total Mixed Cultures	107,348,188	98,484,668	95,177,034 (96%)
Total reads of experiment	217,637,170	201,035,860	194,083,233

4.2.2 Mapping reads to reference genome and transcript abundance estimation

For monocultures, 98% of cleaned reads mapped to the *S. gordonii* NC_009785.1 (Sg) and 95% of cleaned reads mapped to *F. nucleatum* NC_003454.1 (Fn). For mixed transcriptome samples, 30% of clean reads mapped to the *S. gordonii* (SgFn_Sg) and 60% cleaned reads mapped to the *F. nucleatum* (SgFn_Fn) (Table 2.1). Read count plots are generated for each hit to visualize its reproducibility across screen replicates. Furthermore, to aid quality control and to gain an understanding of count distributions and possible problematic samples before and after normalization using box plots. Samples from mixed and monoculture showed a similar distribution of per-gene read

counts per samples, as visualized by box plots (Figure 4.3), indicating that the distributions of data were quantitatively comparable between mixed coaggregate and monoculture samples and no batch effects (Fei *et al.*, 2018). Rscript in Appendix B.

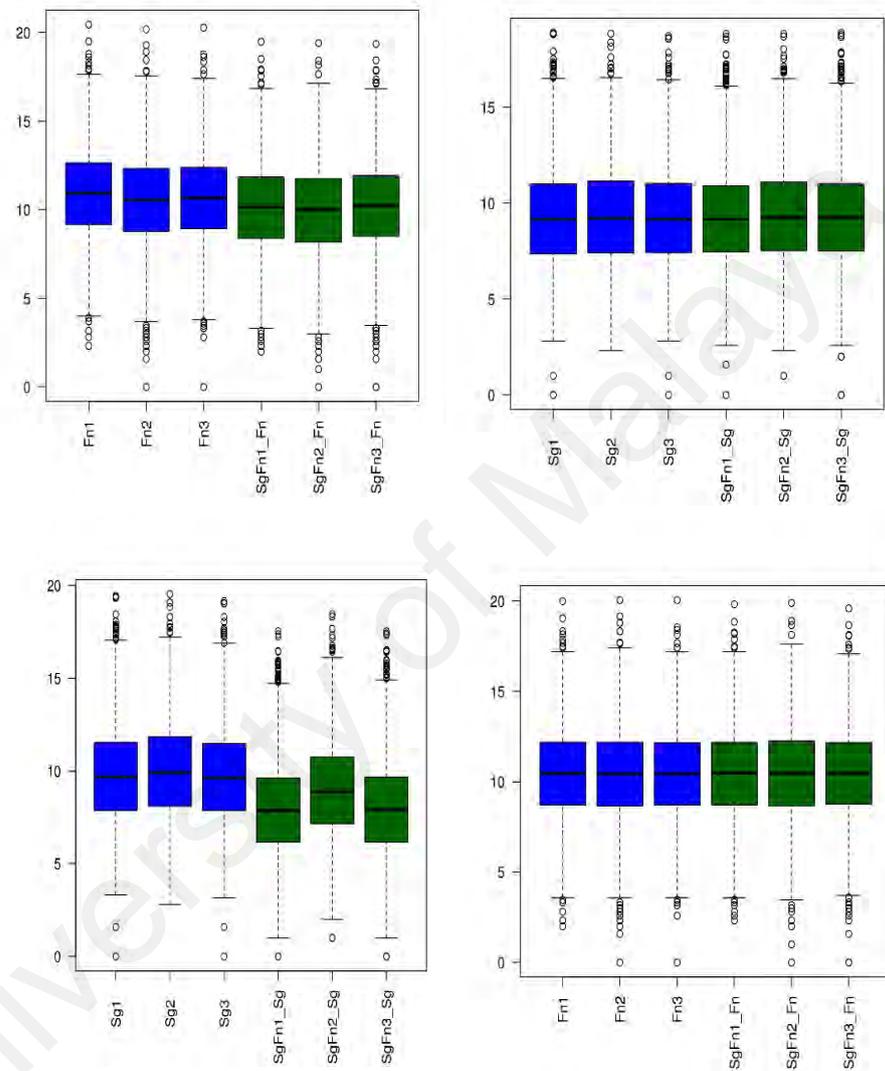
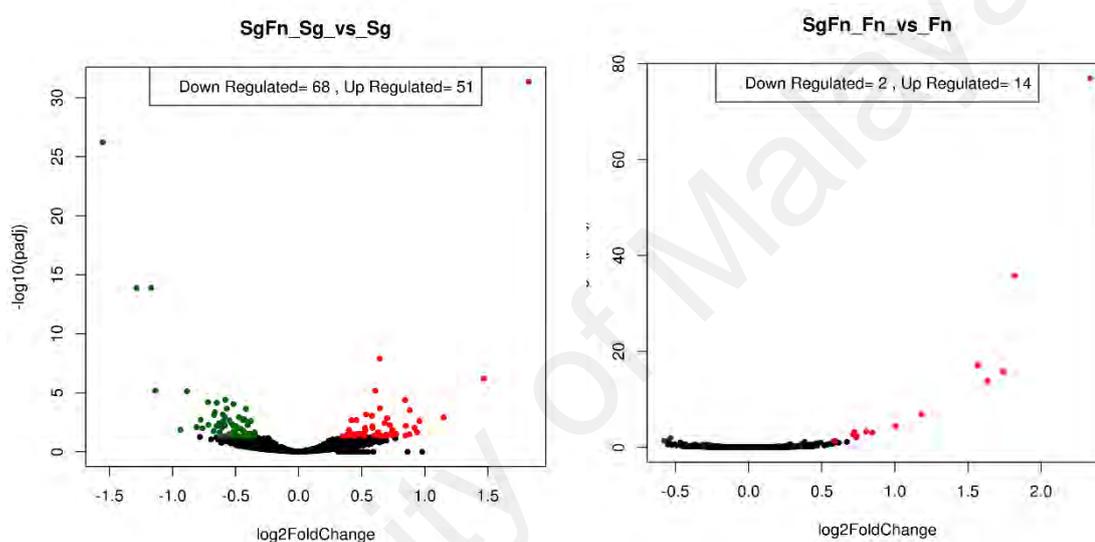


Figure 4.3: Boxplots representing quality of read counts in SgFn experiment. Distribution of raw read counts and normalized read counts. Counts were normalized by the Trimmed Mean of M-values approach. Median counts from each sample are shown as thick black lines surrounded by a box representing the interquartile range, with whiskers that extend to two standard deviations from the median. Outlying observations are shown as open circles. Data from three biological replicates for each sample are shown.

4.3 Differential expression analysis

DESeq2 was used to compare coaggregate samples reads with two monoculture sample reads separately. A total of 119 *S. gordonii* genes (51 down-regulated and 68 up-regulated genes) and 16 *F. nucleatum* genes (two down-regulated and 14 up-regulated) were found to be differentially expressed at a false discovery rate (P_{adj}) of < 0.05 . Volcano plots of genes that were differentially expressed in the two different comparisons illustrate distinct transcriptional profiles (Figure 4.4). R script in Appendix



D.

Figure 4.4: Volcano plots representing differentially expressed genes in coaggregates vs monocultures for *Streptococcus gordonii* and *Fusobacterium nucleatum*.

The y-axis corresponds to the mean expression value of $\log_{10}(P_{adj})$, and the x-axis displays the $\log_2(\text{fold change})$ value. The green and red dots correspond to significantly down and up-regulated genes, respectively ($P_{adj} < 0.05$). The black dots represent the genes whose expression levels did not reach statistical significance ($P_{adj} > 0.05$)

4.4 RT-qPCR validation of RNA-Seq data

To validate the gene regulation observed by RNA-Seq was appropriate, a selection of genes expression was also analyzed by RT-qPCR. Two up-regulated, two down-regulated and two non-regulated genes that were strongly expressed were selected from each species (Appendix C). Expression levels were normalized to 16S rRNA

expression. The patterns of expression were very similar by RNA-Seq and RT-qPCR (Figure 4.5). In each case, there was a strong correlation between the two data sets (Pearson's correlation coefficient between 0.96 and 0.97, $P < 0.01$) indicating RNA-Seq was working well for this set of samples. Therefore, over all the RT-qPCR analysis of gene expression closely matched that by RNA-Seq.

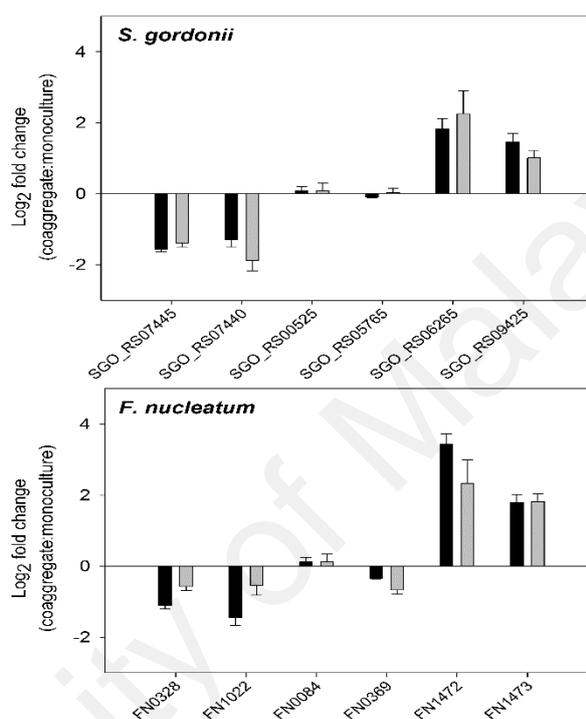


Figure 4.5: RT-qPCR validation of RNA-Seq data analysis results for *Streptococcus gordonii* and *Fusobacterium nucleatum*.

Means and SDs from three independent samples from RNA-Seq (dark bars) and RT-qPCR (light bars) are shown.

4.5 Discussion

4.5.1 Gene regulation in *S. gordonii* in response to coaggregation with *F. nucleatum*

More gene regulation in *S. gordonii* following coaggregation than in *F. nucleatum* were noticed, it is possible that *S. gordonii* has less impact on *F. nucleatum* in initial stages of coaggregation. Significant mechanisms affected included carbohydrate metabolism (19 genes), amino acid metabolism (14 genes) and regulation of DNA metabolic process (4 genes). By STRING database analysis, the major clusters of up-

regulated genes were those related to DNA metabolic processes (Table 4.2) and three genes in

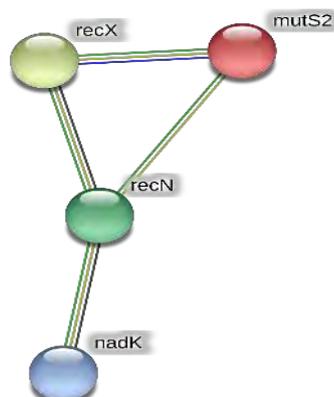


Figure 4.6: Genes involved in DNA maintenance that were regulated in *S. gordonii* in response to coaggregation with *F. nucleatum* visualized by network

tryptophan biosynthesis (Figure 4.6). The genes involved in DNA metabolism are predicted to encode a regulatory protein (RecX), an endonuclease (MutS2), a cell division protein (DivIB) and a DNA repair protein (RecN). In *Bacillus subtilis* MutS2 promotes homologous recombination and protects cells from DNA damage (Burby et al., 2017). It is possible that the changes in DNA metabolism prime the *S. gordonii* cells for uptake and incorporation of foreign DNA following sensing of a different species.

Table 4.2: DNA metabolic process genes of interest regulated in *S. gordonii* in response to coaggregation with *F. nucleatum*

Locus ID	Predicted function (gene)	Log2Fold change	P _{adj}
<i>Regulation of DNA metabolic process</i>			
SGO_RS03085 (SGO_0626)	Regulatory protein (<i>recX</i>)	3.61	<0.01
SGO_RS01280 (SGO_0260)	DNA mismatch binding protein (<i>mutS2</i>)	2.43	0.05
SGO_RS03440 (SGO_0698)	DNA repair protein (<i>recN</i>)	1.52	0.01
SGO_RS05995 (SGO_1221)	Inorganic polyphosphate/ATP-NAD kinase (<i>nadK</i>)	1.51	0.03

The up-regulation of tryptophan biosynthesis genes, along with the cysteine biosynthesis gene *cysK*, was part of a larger global response involving regulation of

amino acid metabolism. Genes in several amino acid metabolism pathways were down-regulated (Table 4.3) following coaggregation (Figure 4.7) including both histidine catabolism (*hut* genes) and histidine biosynthesis (*his* genes). It has recently been shown that coaggregation can be sensed through the action of the *S. gordonii* extracellular protein Challisin, which appears to release amino acids from neighboring bacteria such as *A. oris* (Burby *et al.*, 2017).

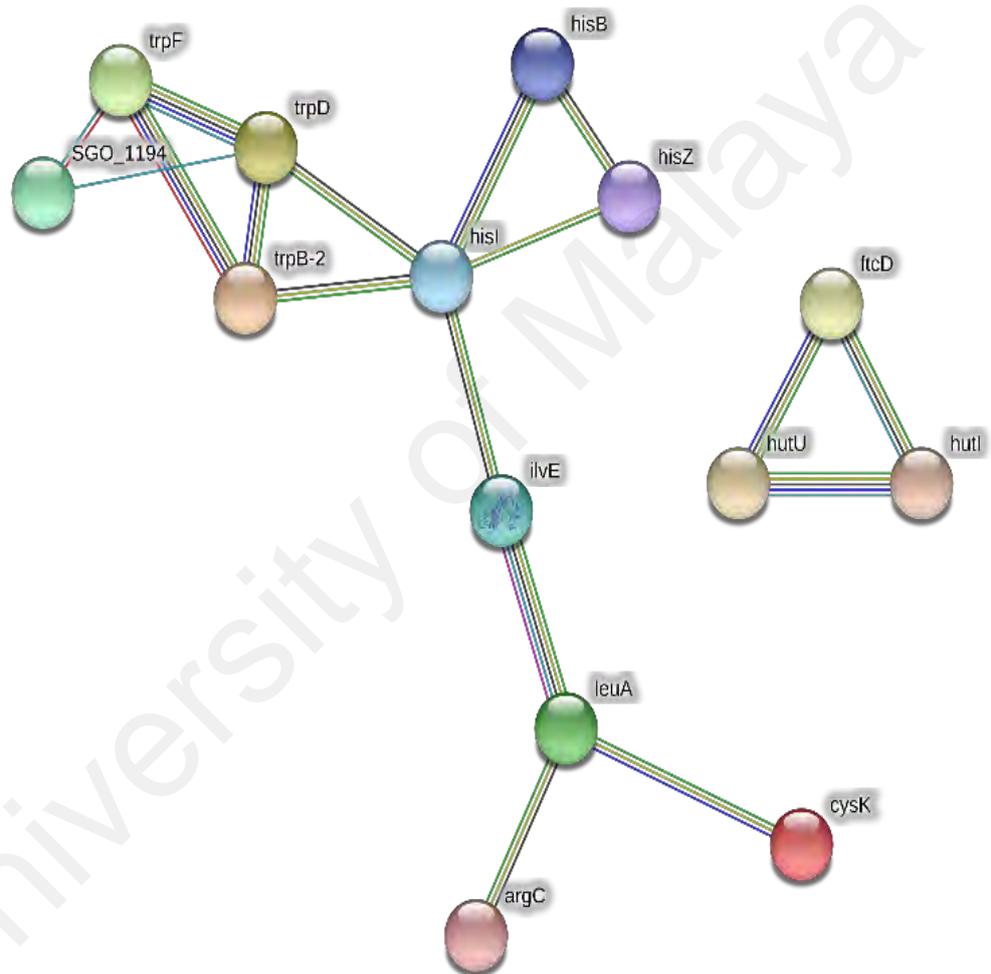


Figure 4.7: Genes involved in amino acid metabolism that were regulated in *S. gordonii* in response to coaggregation with *F. nucleatum* visualized by network.

It is possible that Challisin also acts on *F. nucleatum*. Alternatively, there may be direct cross-feeding of amino acids between in *S. gordonii* and *F. nucleatum*. It has already been shown that *F. nucleatum* can utilize ornithine that is released from *S.*

gordonii in exchange for arginine (Sakanaka *et al.*, 2015), and there may be other nutrient exchanges that have yet to be identified.

Table 4.3: Amino acid metabolic process genes of interest regulated in *S. gordonii* in response to coaggregation with *F. nucleatum*

Locus ID	Predicted function (gene)	Log2 Fold change	P _{adj}
<i>Amino acid metabolism</i>			
SGO_RS02985 (SGO_0606)	Cysteine synthase (<i>cysK</i>)	-1.71	<0.01
SGO_RS03230 (SGO_0656)	Tryptophan synthetase beta subunit (<i>trpB-2</i>)	-1.35	0.04
SGO_RS03245 (SGO_0659)	anthranilate phosphor-ribosyl transferase (<i>trpD</i>)	1.55	0.02
SGO_RS03255 (SGO_0661)	N-(5'-phosphoribosyl anthranilate isomerase (<i>trpF</i>)	1.59	0.05
SGO_RS04445 (SGO_0906)	2-isopropylmalate synthase (<i>leuA</i>)	-1.57	<0.01
SGO_RS05865 (SGO_1194)	Phospho ribosyl anthranilate isomerase	1.46	0.04
SGO_RS06080 (SGO_1238)	Branched chain amino acid aminotransferase (<i>ilvE</i>)	-1.59	<0.01
SGO_RS06880 (SGO_1403)	phosphoribosyl-AMP cyclo hydrolase (<i>hisI</i>)	-1.55	<0.01
SGO_RS06900 (SGO_1407)	Imidazole glycerol-phosphate dehydratase (<i>hisB</i>)	-1.63	0.05
SGO_RS06915 (SGO_1410)	ATP phosphoribosyl transferase regulatory subunit (<i>hisZ</i>)	-1.55	0.02
SGO_RS07685 (SGO_1569)	N-acetyl-glutamate semialdehyde dehydrogenase (<i>argC</i>)	1.63	0.05
SGO_RS08850 (SGO_1804)	Imidazolone propionase (<i>hutI</i>)	-1.52	0.01
SGO_RS08855 (SGO_1805)	Urocanate hydratase (<i>hutU</i>)	-1.50	<0.01
SGO_RS08860 (SGO_1806)	Glutamate formimino transferase (<i>ftcD</i>)	-1.59	<0.01

The largest set of coordinated changes in gene expression was related to carbohydrate uptake and metabolism (Table 4.4). In particular, genes involved in lactose and galactose uptake catabolism (Figure 4.8) were down-regulated following coaggregation. The regulation of these operons has previously been studied in detail (Zeng *et al.*, 2012). The expression of *lacG* or *lacA1* was strongly repressed by glucose compared with lactose or galactose. Regulation of the operon is complex and appears to be mediated by a combination of *LacR*, the global catabolite repressor *CcpA*, and a transcriptional antitermination mechanism directed by *LacT* (Zeng *et al.*, 2012). Here,

there was a small decrease in *lacG* expression in coaggregates compared with monocultures. This may indicate an increase in the local availability of glucose in coaggregates, or possibly a decrease in lactode and/or galactose.

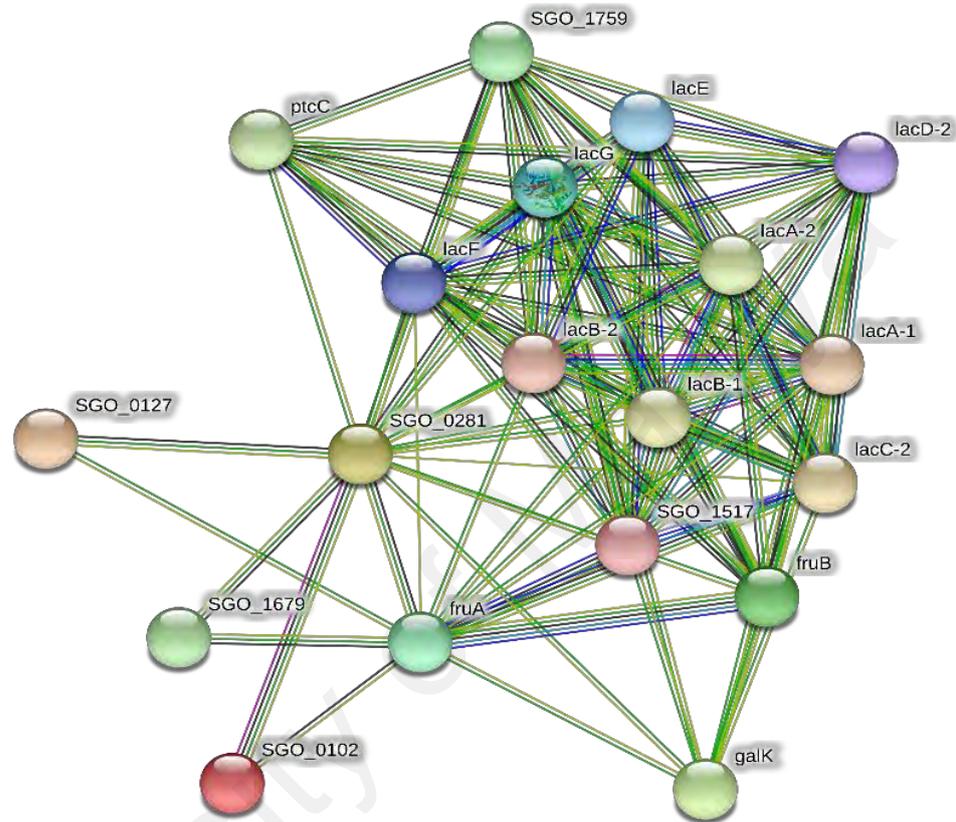


Figure 4.8: Genes involved in carbohydrate metabolism that were regulated in *S. gordonii* in response to coaggregation with *F. nucleatum* visualized by network

Metabolic cross-feeding from carbohydrate metabolism is well known to occur in mixed cultures of bacteria. For example, lactic acid released as a product of carbohydrate catabolism by *S. gordonii* is a key nutrient for *V. parvula* or *aggregatibacter actinomucetemcomitans* (Brown *et al.*, 2007; Egland *et al.*, 2004). Alternatively, changes in carbohydrates may be driven by competition. In mixed-species cultures, *S. gordonii* uses galactose more efficiently than *S. mutans*, except in very low pH (Zeng *et al.*, 2012). It is not clear whether *F. nucleatum* can utilise lactose or galactose more efficiently than *S. gordonii* and this will be the subject of future studies. The major changes in regulation of genes involved in carbohydrate metabolism are

consistent in regulation of genes involved in carbohydrates metabolism are consistent with a previously published proteomic analysis.

Table 4.4: Carbohydrate metabolic process genes of interest regulated in *S. gordonii* in response to coaggregation with *F. nucleatum*.

Locus ID	Predicted function (gene)	Log2 Fold change	P _{adj}
<i>Carbohydrate metabolism</i>			
SGO_RS00510 (SGO_0102)	sugar ABC transporter permease	-1.55	0.02
SGO_RS00630 (SGO_0127)	RpiR family transcriptional regulator	1.66	0.03
SGO_RS01385 (SGO_0281)	PTS diacetyl lchitobiose transporter subunit IIC	1.50	<0.01
SGO_RS04560 (SGO_0932)	Galactokinase (<i>galK</i>)	-1.40	0.01
SGO_RS05460 (SGO_1112)	Phosphofructokinase (<i>fruB</i>)	-1.65	0.01
SGO_RS05465 (SGO_1113)	PTS fructose transporter subunit IIC (<i>fruA</i>)	-1.54	0.02
SGO_RS07410 (SGO_1512)	6-phospho-beta-galactosidase (<i>lacG</i>)	-1.30	0.03
SGO_RS07415 (SGO_1513)	PTS lactose transporter subunit IIBC (<i>lacE</i>)	-1.42	0.03
SGO_RS07420 (SGO_1514)	PTS lactose transporter subunit IIA (<i>lacF</i>)	-1.60	0.02
SGO_RS07430 (SGO_1516)	tagatose-bisphosphate aldolase (<i>lacD-2</i>)	-1.65	<0.01
SGO_RS07435 (SGO_1517)	tagatose-6-phosphate kinase (<i>lacC-1</i>)	-2.25	<0.01
SGO_RS07440 (SGO_1518)	galactose-6-phosphate isomerase (<i>lacB-2</i>)	-2.44	<0.01
SGO_RS07445 (SGO_1519)	galactose-6-phosphate isomerase (<i>lacA-1</i>)	-2.94	<0.01
SGO_RS07470 (SGO_1524)	tagatose-6-phosphate kinase (<i>lacC-2</i>)	-1.59	<0.01
SGO_RS07475 (SGO_1525)	galactose-6-phosphate isomerase (<i>lacB-1</i>)	-1.59	<0.01
SGO_RS07480 (SGO_1526)	galactose-6-phosphate isomerase subunit (<i>lacA-2</i>)	-1.85	<0.01
SGO_RS07720 (SGO_1576)	PTS cellobiose transporter subunit IIC (<i>ptcC</i>)	-1.53	0.01
SGO_RS08235 (SGO_1679)	PTS mannose transporter subunit IIAB	-1.54	<0.01
SGO_RS08625 (SGO_1759)	6-phospho-beta-glucosidase	-1.58	0.02

In this case, increased levels of enzymes for glycolysis and the pentose phosphate pathway were detected in dual-species *S. gordonii* / *F. nucleatum* cultures compared

with *S. gordonii* monocultures, consistent with higher energy availability in mixed cultures (Hendrickson *et al.*, 2012).

4.5.2 Gene regulation in *F. nucleatum* in response to coaggregation with *S. gordonii*

Genes that were regulated in *F. nucleatum* in response to coaggregation with *S. gordonii* are listed in (Table 4.3). Broadly, functions related to aminoacid transport, calcium transport, catalytic activity/phosphorus metabolic process, lipid transport/metabolism and sialic acid catabolism were affected. The network of interacting genes visualized using the STRING database (Figure). The most striking impact of coaggregation was on the expression of an eight-gene operon involved in sialic acid catabolism (FN14070-FN1477).

All genes in this operon were significantly up-regulated between 2 and 5 fold in response to coaggregation. The genes in this operon encode the machinery of sialic acid uptake via tripartite ATP-independent periplasmic transporter (SiaPQM) and catabolism by NanA/NanK/NanE to produce N-acetylglucosamine-6-phosphate (Stafford *et al.*, 2012). In addition, there is a regulator (NanR) and two proteins of unknown function (FN1470 and FN1477). The N-acetylglucosamine-6-phosphate is a substrate for the enzymes NagA and NagB, which produce fructose-6-phosphate that can then enter glycolysis (Caing-Carlsson *et al.*, 2017). Genes encoding NagA/NagB were not differentially regulated in our RNA-Seq analysis.

It is not clear why sialic acid catabolism was affected by coaggregation. Free sialic acid is found in saliva with in the healthy mouth at concentrations > 40mg/dL (Rathod *et al.*, 2014). In addition, sialic acid is commonly present in host glycoproteins as a terminal residue on carbohydrate side chains (Sonesson *et al.*, 2011). Some times strains of *S. gordonii*, including *S. gordonii* DL1, bind to host sialic acids on glycoproteins and catabolise free sialic acids in saliva (Byers *et al.*, 1996; Wong *et al.*, 2018).

Table 4.5 Genes regulated in *F. nucleatum* following coaggregation with *S. gordonii*

Locus ID	Predicted function (gene)	Log2 Fold change	P _{adj}
Amino acid transport			
FN0328	Na ⁺ -linked D-alanine glycine permease	-1.49	0.04
<i>Ca²⁺ transport</i>			
FN1022	Ca ²⁺ -transporting ATPase	-1.45	0.01
<i>Catalytic activity or phosphorous metabolic process</i>			
FN0796	Pyruvate phosphate dikinase	1.50	0.04
FN0798	Fructose-1,6-bisphosphatase (<i>fbp</i>)	1.65	<0.01
FN0938	Hypothetical protein	1.64	<0.01
<i>Lipid transport/metabolism</i>			
FN0940	Hypothetical protein	1.75	<0.01
FN0941	□ glutamyl transpeptidase	1.67	<0.01
<i>Function unknown</i>			
FN1078	Hypothetical protein	1.67	<0.01

However, these activities would likely to reductions in the available sialic acid, which would then repress However, these activities would likely lead to reductions in the available sialic acid, which would then repress sialic acid catabolism in *F. nucleatum*. It is possible that the regulation of sialic acid operon in coaggregates may have resulted from changes in the localised levels of free sugars such as glucose due to sugar consumption by one or both species.

Table 4.6 Sialic acid metabolism genes regulated in *F. nucleatum* following coaggregation with *S. gordonii*

Locus ID	Predicted function (gene)	Log2 Fold change	P _{adj}
<i>Sialic acid catabolism</i>			
FN1470	Hypothetical protein	3.10	<0.01
FN1471	LacI family transcriptional regulator (<i>nanR</i>)	3.34	<0.01
FN1472	N-acetylneuraminase binding protein (<i>siaP</i>)	5.05	<0.01
FN1473	Sialic acid TRAP transporter permease protein (<i>siaQM</i>)	3.53	<0.01
FN1474	N-acetylmannosamine kinase (<i>nanK</i>)	2.97	<0.01
FN1475	N-acetylneuraminase lyase (<i>nanA</i>)	2.27	<0.01
FN1476	N-acetylmannosamine-6-phosphate 2-epimerase (<i>nanE</i>)	1.80	<0.01
FN1477	Hypothetical protein	2.01	<0.01

Sugars are known to modulate sialic acid catabolism genes in some bacteria. For example, in *Corynebacterium glutamicum*, sialic acid catabolism genes are down-regulated during growth in glucose or fructose due to regulation by NanR and possibly also by a global carbon catabolite repressor (Uhde *et al.*, 2016). It is noteworthy that high sialic acid concentrations in saliva have been associated with periodontal disease (Rathod *et al.*, 2014). *S. gordonii* is generally associated with periodontal health (Ge *et al.*, 2013). It is possible that *S. gordonii* indirectly reduces sialic acid in saliva by stimulating sialidase activity of neighbouring *F. nucleatum* cells.

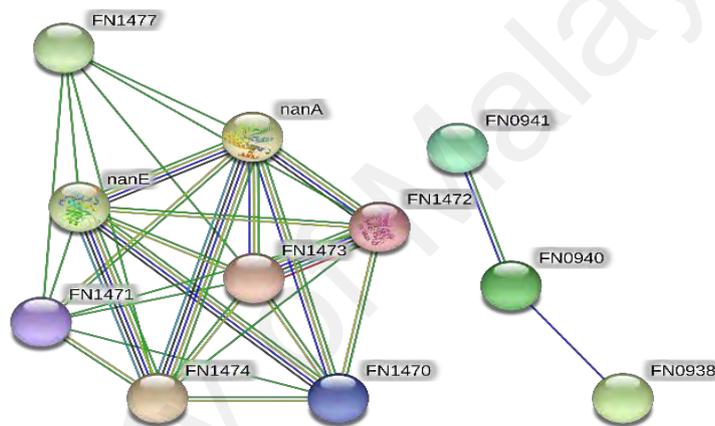


Figure 4.9: Network of genes involved in carbohydrate metabolism that were regulated in *S. gordonii* in response to coaggregation with *F. nucleatum*.

In addition to the sialic acid utilisation operon, there were clear changes in gene expression of FN0796 and FN0798, which were upregulated 1.5- to 1.7-fold. These genes encode two key enzymes in gluconeogenesis: pyruvate phosphate dikinase (FN0796) that converts pyruvate to phosphoenol pyruvate, and fructose 1,6-bisphosphatase (FN0798) that hydrolyzes fructose 1,6-bisphosphate to fructose 6-phosphate. Two genes were downregulated between 1.5- and 1.7-fold: one gene each in amino acid transport (FN0328) and calcium transport (FN1022). Genes FN1078, FN0938, and FN0940 were up-regulated between 1.5- and 1.8-fold, but their functions are unknown. It is possible that FN0938 and FN0940 are part of the same operon as

FN0941, encoding gamma-glutamyltranspeptidase. This enzyme plays a role in synthesis and degradation of glutathione, a key protective agent against oxidative stress. In summary, a significant proportion of the gene regulation response of *F. nucleatum* to coaggregation appears to involve metabolic pathways that converge on fructose-6-phosphate.

University of Malaya

CHAPTER 5: TRANSCRIPTOMIC RESPONSES TO COAGGREGATION BETWEEN *STREPTOCOCCUS GORDONII* AND *VEILLONELLA PARVULA*

5.1 Coaggregation of *Streptococcus gordonii* and *Veillonella parvula*

Coaggregation between *Streptococcus gordonii* and *Veillonella parvula* was qualitatively assessed by mixing concentrated suspensions of cells in coaggregation buffer vigorously. In no time, substantial aggregates were clearly noticed and were scored '4+' on the visual scale with absolute clear background. Coaggregation was monitored to ensure saliva did not inhibit this interaction by mixing fresh cultures of cells in freshly collected human saliva, again strong coaggregation was observed within seconds with '4+' designation in reference to the standard visual scoring system. Coaggregation structures were predicted by pre-staining *S. gordonii* and *V. parvula* with fluorescent dyes and imaging by confocal laser scanning microscopy (CSLM; Figure 5.1). macroscopic aggregates, >100 μm in width, were visible and contained *S. gordonii* and *V. parvula* cells evenly mixed throughout the structure.

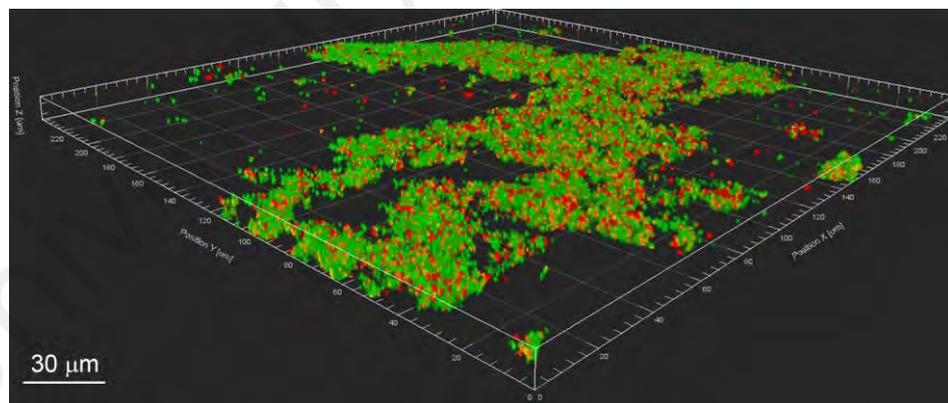


Figure 5.1: Confocal scanning laser microscopy analysis of coaggregation between *S. gordonii* and *V. parvula*.

Pico-green-stained *S. gordonii* cells (green) and propidium iodide-stained *V. parvula* (red) were suspended in human saliva and mixed vigorously to induce coaggregation. Samples were visualized by confocal laser scanning microscopy. A large coaggregated mass is clearly visible, *S. gordonii* and *V. parvula* cells interspersed throughout the structure.

Interactions between *S. gordonii* and *V. parvula* were more closely assessed by TEM assay in which *S. gordonii* and *V. parvula* cells could be distinguished on the basis of

cell wall morphology (Figure 5.2). Close associations of *S. gordonii* with *V. parvula* cells (Figure 5.2a) were found. At high resolution, there appeared to be structures connecting the cells (Figure 5.2b, arrows). This analysis comprehensively confirmed significant potential for cell-cell interactions occurrence in surface-associated biofilms. Thus, impact of coaggregation on gene expression in monoculture and coaggregation assessed in human saliva.

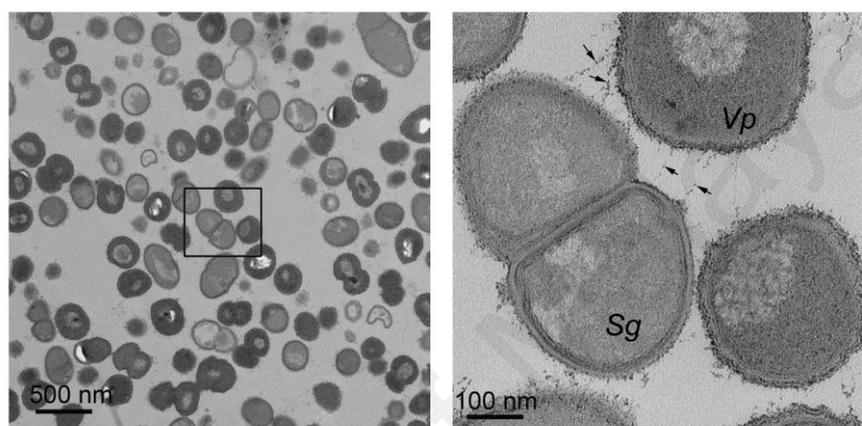


Figure 5.2: High Resolution Analysis of Interactions between *S. gordonii* and *V. parvula* cells using Transmission Electron Microscopy.

Cells of *S. gordonii* and *V. parvula* were washed, suspended in human saliva and concentrated cultures were mixed vigorously to induce coaggregation. Samples were embedded in resin and visualized by TEM. At relatively low power (a), large areas containing densely packed *S. gordonii* and *V. parvula* cells were visible. *S. gordonii* (Sg) and *V. parvula* (Vp) could be more easily distinguished at higher magnification (b). Nearby cells appeared to be connected by extracellular material or fibrils (arrows).

5.2 Dual RNA-Seq data analysis

5.2.1 RNA-Seq data generation and pre-processing

The RNA-Seq transcriptional analysis from a total of 9 sequencing libraries of 3 mixed (SgVp) and 3 monoculture (Sg, Vp) biological replicates each generated a total of 126,166,650 paired end raw reads of 100bp read length. After the removal of low-quality reads and adapter content by Trimmomatic-0.36 obtained 125,492,518 pre-processed reads (clean ratio = 92.5%) (Table 5.1). The sequence reads of all samples were deposited in the NCBI sequence read archive (SRA) as study under the Bio Project accession number PRJNA505944.

Table 5.1: Alignment statistics of mixed and monoculture transcriptomes in *S. gordonii* and *V. parvula* in coaggregation experiment

Sample name	Raw reads	Pre-Processed Reads	Mapped Reads
Monocultures			
Sg1	13,702,956	13,629,680	13,104,542
Sg2	15,266,848	15,185,702	14,640,981
Sg3	12,783,626	12,715,336	12,435,361
Total reads of <i>S. gordonii</i>	41,753,430	41,530,718	40,180,884(96%)
Vp1	14,219,126	14,143,356	11,117,744
Vp2	15,428,358	15,345,020	12,113,083
Vp3	11,866,866	11,804,268	9,138,018
Total reads of <i>V. parvula</i>	41,514,350	41,292,644	32,368,845(78%)
Coaggregates			
SgVp1	13,672,234	13,671,030	6,963,203 (Sg) 6,139,800 (Vp)
SgVp2	16498720	16497258	5858455 (Sg) 9900226 (Vp)
SgVp3	12727916	12,726,824	5747117 (Sg) 6515038 (Vp)
Total reads of SgVp	42,898,870	42,669,156	18,568,775(Sg), 14,056,794(Vp)
Total reads of experiment	126,166,650	125,492,518	105,175,298

5.2.2 Mapping reads to reference genome and abundance estimation

Pre-processed reads after mapping using TopHat, For *S. gordonii* 96% of total clean reads were mapped to monoculture (Sg) and 43% reads mapped to mixed culture (SgVp_Sg) and for *V. parvula*, 78% of cleaned reads were mapped to monoculture (Vp) and 32% reads mapped to mixed culture (SgVp_Vp) as shown in Table 5.1. Read counts of mixed and monoculture samples after Trimmed Mean of M-values (TMM) normalization method showed a similar distribution of per-gene read counts per sample ensuring normalized distributions of data were quantitatively comparable between mixed coaggregate and monoculture samples and no batch effects were apparent from the box plots (Figure 5.3).

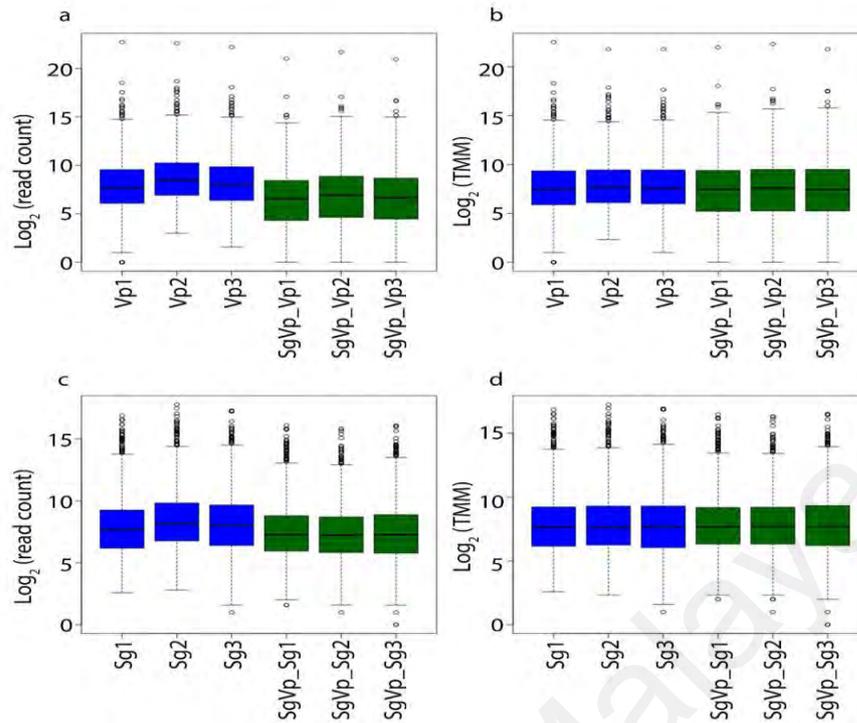


Figure 5.3: Box plots showing transcript abundance before and after normalization in SgVp Experiment.

Gene expression levels of *V. parvula* (a,b) or *S. gordonii* (c,d) genes in coaggregate or monoculture samples are shown before (a,c) or after (b,d) TMM normalization. Median counts are represented as thick black lines, boxes show interquartile ranges and whiskers indicate two SDs from the median. Outliers are indicated by 'o' symbols. In the box plot display, a box is formed with sides at the 25th and 75th percentiles of distribution. A line is also drawn within the box at the level of median. Whiskers are also drawn extending beyond each end of the box with points beyond the whiskers typically indicating outliers.

The distribution of data between samples was further compared by Principal Components Analysis (PCoA). *V. parvula* monoculture and coaggregate samples were clearly separated by the first principal component (Appendix E). However, only two of the *S. gordonii* coaggregate samples clustered tightly together. The monoculture samples were relatively scattered, indicating poor consistency between samples. Attempts were made to repeat these samples. However, problems with batch effects and contamination made it impossible to interpret the data. Therefore, it was decided to use the three independent repeats for differential expression analysis on the basis that this would provide useful information to form hypotheses for further testing.

5.2.3 Differential expression analysis

The impact of coaggregation on gene expression was investigated by differential expression analysis using DESeq. A comparison between SgVp coaggregate culture and Vp monoculture pinpointed a total of 272 significant differentially expressed genes ($FDR/P_{adj} < 0.05$). An additional 156 genes met the prediction $P < 0.05$. The large number of differentially expressed genes in *V. parvula* suggests that coaggregation has a major impact on gene expression in this species where as for comparison between SgVp coaggregate culture and Sg monoculture using similar approach, *S. gordonii* genes reached the stringent FDR-adjusted significant level $P_{adj} < 0.05$. However, none of the *S. gordonii* genes reached the stringent FDR-adjusted level $P_{adj} < 0.05$. Therefore, to identify potentially interesting genes, we applied a non-corrected P value cut-off of $P < 0.05$. By this measure, 69 significant *S. gordonii* genes were classed as differentially expressed. The differential expression was visualized using volcano plots (Figure 5.4).

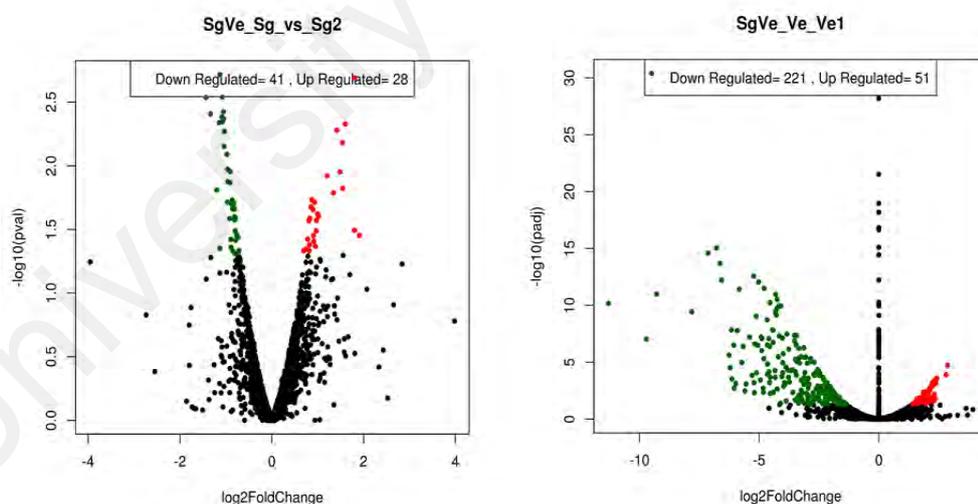


Figure 5.4 : Volcano plots representing differentially expressed genes in coaggregates vs monocultures for *S. gordonii* and *V. parvula*.

The y-axis corresponds to the mean expression value of $\log_{10}(P_{adj})$, and the x-axis displays the $\log_2(\text{fold change})$ value. The green and red dots correspond to significantly down and up-regulated genes, respectively ($P_{adj} < 0.05$). The black dots represent the genes whose expression levels did not reach statistical significance ($P_{adj} > 0.05$).

5.3 RT-qPCR validation of RNA-Seq data

Gene regulation detected by RNA-Seq was validated by selection of genes that were up-regulated or down-regulated by evaluating with RT-qPCR (Figure 5. 4). Expression levels were normalized to 16s rRNA expression. Five *V. parvula* genes were selected for analysis including two genes that were strongly up-regulated following coaggregation (RS04540 and RS02475), two genes that were non-significant (RS04165 and RS04150) and one gene that was strongly down-regulated (RS00600) Appendix C. There was a strong correlation between differential expression levels measures by RNA-Seq versus RT-qPCR (Pearson correlation coefficient 0.946, $P = 0.015$). indicating that RNA-Seq was working well for this set of samples. Due to the concerns noted above about the RNA-Seq data for *S. gordonii*, a larger number of *S. gordonii* genes was assessed by RT-qPCR. From ten genes analyzed, there was generally good agreement between RNA-Seq and RT-qPCR.

Overall, there was strong correlation between the data sets from the two different techniques (Pearson correlation coefficient 0.830, $P = 0.003$). Only one gene gave clearly different results with RNA-Seq versus RT-qPCR. This was RS08935, which appeared not regulated by RNA-Seq, but was strongly up-regulated (approximately 5-fold) in coaggregate cultures by RT-qPCR. RNA-Seq also underestimated the gene regulation of several other genes (RS09100, RS06800, RS02790 and RS05725) compared with RT-qPCR. Overall, RNA-Seq detected up- and down-regulated *S. gordonii* genes but appeared to underestimate levels of regulation, which may partially explain why no genes met our originally planned cut-off criterion for differential expression ($P_{adj} < 0.05$). nevertheless, the data from this analysis provided some justification for applying a less strict cut-off ($P < 0.05$) to identify genes of interest.

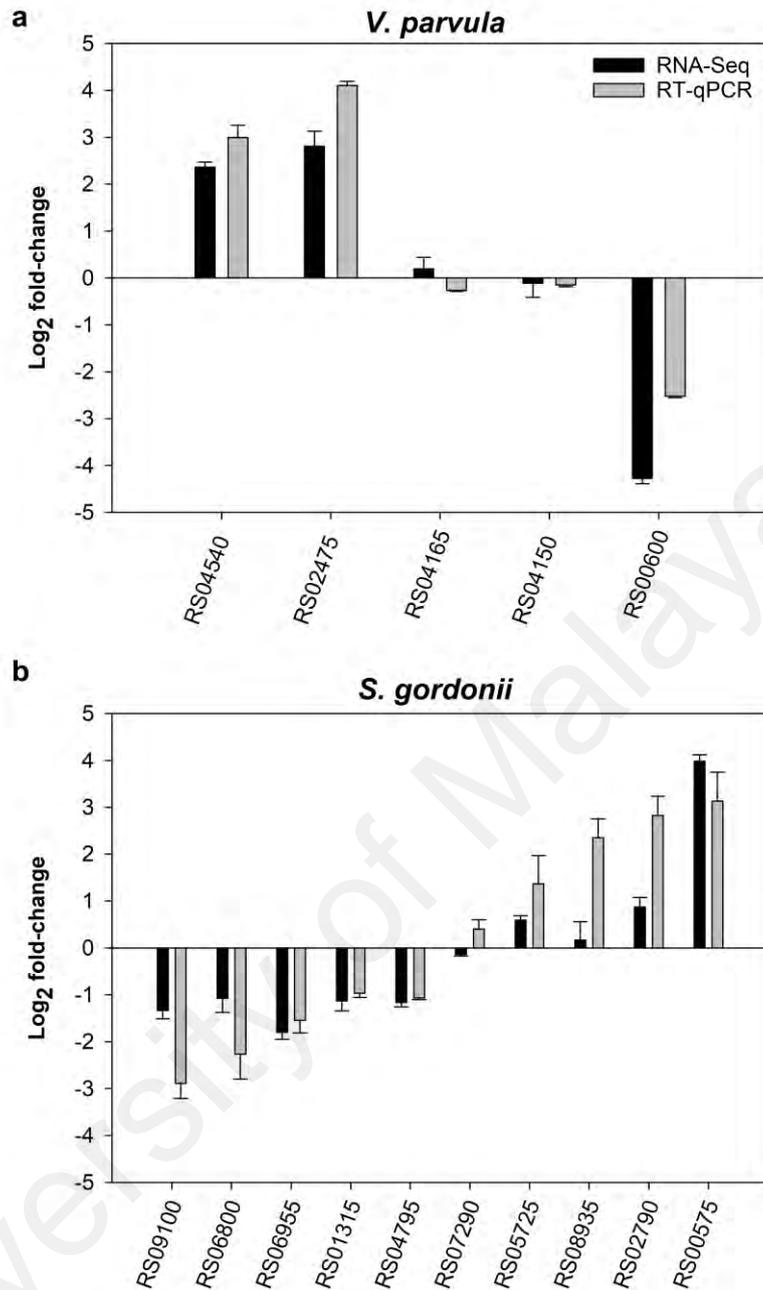


Figure 5.5 Validation of RNA-Seq data by RT-qPCR.

The expression of selected genes in coaggregates versus monoculture was assessed by RT-qPCR, normalized to 16S rRNA expression. Graphs show mean and SD from three independent experiments. In general, there were strong correlation between expression levels measure by RNA-Seq versus RT-qPCR.

5.4 Discussion

5.4.1 Gene regulation in *S. gordonii* in response to coaggregation with *V. parvula*

Biological functions of significantly expressed differential genes were explained better by gene enrichment analysis. From this analysis, the 'Biological process' categories with the lowest *P*-values were carbohydrate transport ($P = 1.7e-2$) and phosphate-containing compound metabolic process ($P = 2.5e-2$). Four genes relating to carbohydrate transport were each down-regulated approximately 2-fold in coaggregates. Three of these genes (SGO_RS00210/RS00215/RS00220) encode PTS components EIIB, EIIC, and EIIC of a mannose, sorbose or fructose transporter, while the other encode the EIIBCA components of a sucrose transporter. Along with other six other genes SGO_RS09095 was classified under the 'phosphate-containing compound metabolic process', of these four genes encoding L-lactate dehydrogenase I *ldh* (SGO_RS06050), a transposase (SGO_RS01075), tagatose-6-phosphate kinase *lacC* (SGORS07435), and fructokinase *scrK* (SGO_RS08605). The *dgkA* gene (SGO_RS03505) encoding diacylglycerol kinase were down regulated between 1.6- to 2.7 fold. The only gene in this category up-regulated in response to coaggregation was *thrB* (SGO_RS03935), encoding homoserine kinase (1.9-fold).

STRING DB analysis revealed a large cluster of interacting genes broadly related to carbohydrate transport and metabolism (Figure 5.6) including genes SGO_RS00210/RS00215/RS00220 (old names SGO_0044-0046), *scrK*, *ldh*, *lacC* (SGO_RS07435, old name SGO_1517), SGO_RS09095 (SGO_1857) and a downstream gene SGO_RS09100 (SGO_1858), encoding a putative sucrose-6-phosphate hydrolase, which was down-regulated approximately 2.5-fold in coaggregates. Furthermore, SGO_RS02020 (SGO_0405), encoding a putative β -n-acetylhexosaminidase, SGO_RS03680 (SGO_0749), encoding glutathione-disulfide reductase, and SGO_RS05250 (SGO_1069), encoding aminopeptidase n, were each up-

regulated between 1.8- to 2-fold. The cluster also contained SGO_RS03795 (SGO_0773), encoding carbon catabolite protein A *ccpA*, SGO_RS01740 (SGO_0352), encoding a sugar ABC-type transporter ATP binding protein, and SGO_10300 (*abpA*), encoding an amylase binding protein, which were each down regulated 1.8- to 1.9-fold in coaggregates.

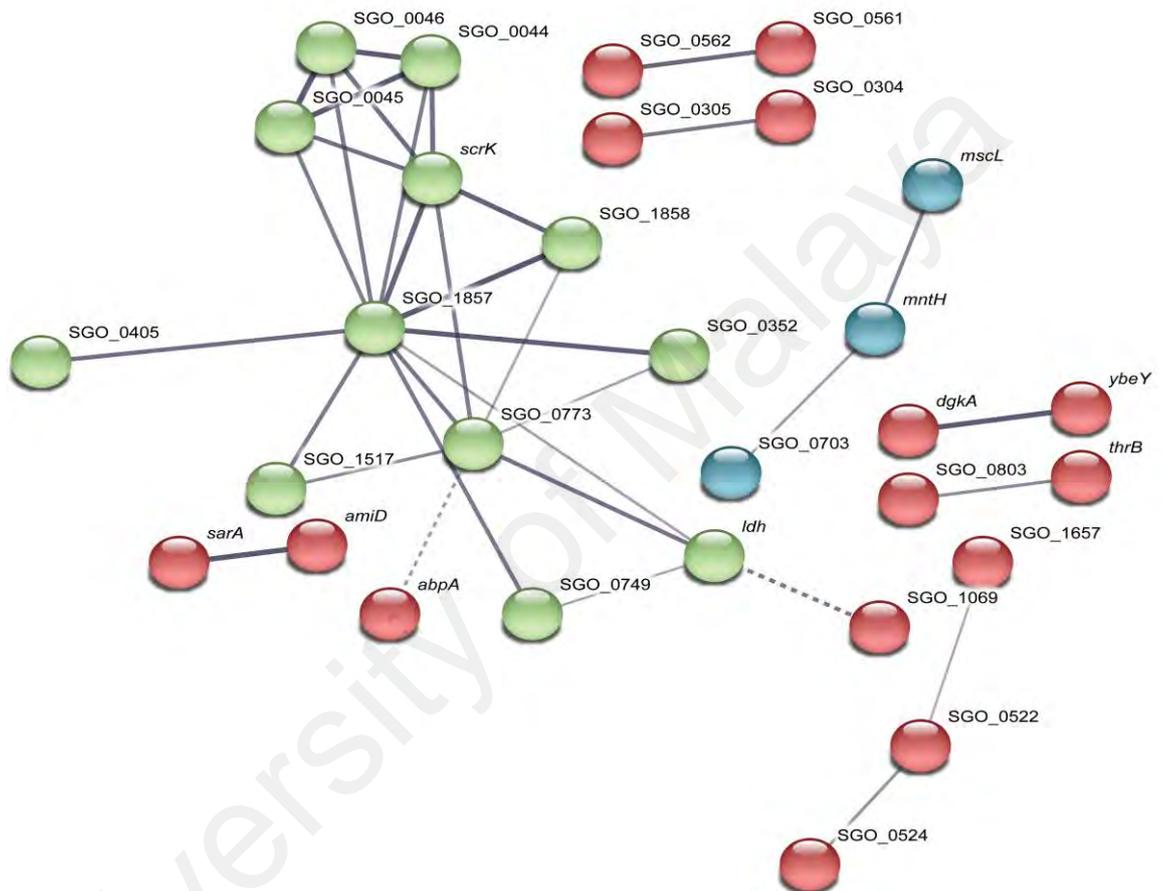


Figure 5.6 Genes regulated in *S. gordonii* in coaggregation with *V. parvula*

STRING DB also show many two- or three-Putative protein interactions (Figure 5.4). The genes *sarA* (*hppA* SGO_RS08395) and *amiD*, (SGO_RS08385), likely encode the ATP binding protein and permease protein, respectively, of a peptide ABC-type transporter. Genes SGO_RS03460 (SGO_0773), encoding a putative Fur family transcriptional regulator designated PerR, SGO_RS01065 (*mntH*), encoding a manganese transporter and SGO_RS06260 (*mscL*), encoding a putative large conductance mechanosensitive channel were linked primarily on the basis of text mining.

The *dgkA* gene was linked to the gene immediately upstream, a putative metalloprotease SGO_RS03500 (*ybeY*). Similarly, *thrB* was linked to the adjacent (downstream) gene, SGO_RS03940 (SGO_0803), an uncharacterised protein. Finally, the closely localized genes SGO_RS02580 (SGO_0522) and SGO_RS02590 (SGO_0524), encoding putative acylphosphatase and an hd domain protein (predicted phosphohydrolase enzyme with conserved histidine and/or aspartate residues), respectively, were linked to SGO_RS08120 (SGO_1657), encoding a membrane protein, on the basis of gene co-occurrence and protein homology.

In total, 27 of the 69 regulated *S. gordonii* genes were grouped into networks by STRING DB. Of the remaining 42 genes, two genes were of interest due to their links with carbohydrate metabolism: SGO_RS07610, encoding a putative glucose-1-phosphate adenylyl transferase *glgC*, and SGO_RS08800, encoding pyruvate formate lyase activating enzyme, *pflC*, were each down-regulated 1.7-fold. Two genes, SGO_RS01500 and SGO_RS01505 encode components of a lantibiotic ABC-type transporter permease and were each down-regulated 1.8-fold. The SGO_RS02790 (*sgc*) gene encoding the extracellular serine protease Challisin was up-regulated 1.8-fold, and a Zn-dependent protease encoded by SGO_RS01335 was up-regulated 2-fold. Additional regulated genes encoded ribosomal proteins (two genes), housekeeping functions such as an RNA helicase and esterase, transcriptional regulation or DNA binding (7 genes), stress responses (for example, a universal stress protein), one enzyme in the lysine biosynthesis pathway and several hypothetical, unknown or poorly characterised proteins.

Altogether, a very less gene regulation of *S. gordonii* and no possible differentially expressed significant genes by our planned cut-off value of $P_{\text{adj}} < 0.05$. Possibility is that *S. gordonii* had weak capacity for cell-cell sensing than *V. parvula* and also technical issues like gram-positive streptococcal cells are relatively robust compared

with *V. parvula* and it is possible that *S. gordonii* had degraded to some extent prior to RNA-Seq, though no appearance of degradation on agarose gels or in the BioAnalyzer it is reflected in the RT-qPCR generally revealed similar patterns of gene expression detected by RNA-Seq for both microorganisms, but most of the *S. gordonii* genes show strong regulation in RT-qPCR compared to RNA-Seq. Previous studies on *S. mutans* have shown stronger levels of gene regulation detected by microarrays than by RNA-Seq (Abranches *et al.*, 2008; Zeng *et al.*, 2013). Thus, it appears that RNA-Seq may underestimate streptococcal gene expression for reasons that have yet to be determined. So, for the purpose of gene discovery we elected to use a lower threshold of $P < 0.05$ to consider *S. gordonii* genes potentially regulated filtered to be 69 regulated genes, which is broadly in line with previous RNA-Seq studies of *S. gordonii* interactions *C. albicans* (Dutton *et al.*, 2016) and our current study of *S. gordonii* with *F. nucleatum* (Mutha *et al.*, 2018).

Similar to *V. parvula*, the response of *S. gordonii* was influenced by regulation in carbohydrate and transporter genes. Components of PTS for mannose, sorbose or fructose (SGO_0044-0046) were down-regulated in response to coaggregation. This system has previously been shown to be under the control of the arginine-sensing regulators ArgR and AhrC (Robinson *et al.*, 2018). *S. gordonii* arginine biosynthesis genes were regulated in response to coaggregation with *A. oris* through the action of the extracellular protease Challisin, which appears to scavenge arginine from *A. oris* cells (Jakubovics *et al.*, 2008a). Here, the gene encoding Challisin was up-regulated 1.8-fold in response to coaggregation, recommending that it may also be important in the interaction with *V. parvula*. (Egland *et al.*, 2004) However, there was no evidence for a global arginine-sensing response in *S. gordonii* and arginine biosynthesis genes were not regulated in response to coaggregation. Therefore, it is possible that an alternative regulator was acting to control the expression of the PTS genes. More generally,

coaggregation led to significant down-regulation in several genes involved in metabolism of sucrose (a PTS transporter and sucrose-6-phosphate hydrolase), fructose (fructokinase, *scrK*), or lactose (*lacC*), uptake of sugar and glycogen biosynthesis (*glgC*). In addition, genes involved in the terminal steps of glucose metabolism, such as lactate dehydrogenase (*ldh*) and pyruvate formate lyase activating enzyme (*pflC*) were also down-regulated. The impact of these changes on central carbon metabolism and on the production of lactic acid as a nutrient source for *V. parvula* is not clear. However, these changes are in agreement with previous studies that have shown a remodelling of carbon metabolism in *S. gordonii* following coaggregation with *V. parvula* by up-regulation of the *amyB* gene (Egland *et al.*, 2004).

It is possible that changes in *S. gordonii* carbon metabolism were driven by CcpA, which itself was down-regulated approximately 1.8-fold following coaggregation. CcpA is a master regulator of gene expression in oral streptococci. Previous studies using microarrays and RNA-Seq identified between 45 and 170 genes that were differentially expressed at least two-fold in glucose grown wild-type *S. mutans* cells and an isogenic *ccpA* mutant, including many genes encoding PTS components (Abranches *et al.*, 2008; Zeng *et al.*, 2013). It is likely that CcpA also plays a broad role in gene regulation in *S. gordonii* in response to extracellular carbon source. In addition to a role in sensing *V. parvula*, *S. gordonii* CcpA controls the expression of the arginine deiminase system, biofilm formation, PTS expression, competence and the peroxidogenic pyruvate oxidase SpxB (Dong *et al.*, 2004; Redanz *et al.*, 2018b; Tong *et al.*, 2011). However, there is evidence that even closely related streptococcal CcpA proteins can have different functions in carbon source sensing (Redanz *et al.*, 2018b; Tong *et al.*, 2011). Therefore, it will be important to determine the global impact of *ccpA* disruption or down-regulation in *S. gordonii* in future.

5.4.2 Gene regulation in *V. parvula* in response to coaggregation with *S. gordonii*

Oxidoreductase activity (GO Molecular Function 0016491) and oxidation-reduction process (GO Biological Process 0055114) were the two processes significantly enriched in *V. parvula* coaggregate samples compared with monocultures which include proteins involved in a wide range of biological processes that involve redox reactions. Interactions between this group of genes were further investigated by analysis with the STRING DB (Figure 5.7). Of the 32 genes input into the analysis, 16 genes grouped into one cluster that was centered upon HSIVP1_RS08035 (Vpar_1637, glutamate synthase, downregulated 30-fold in coaggregates). An additional glutamate synthase component, HSIVP1_RS08995, was down-regulated 7.2-fold in coaggregates, confirming the importance of this enzyme. Another key glutamate metabolism gene, HSIVP1_RS07610 (Vpar_1550), encoding glutamate dehydrogenase, was strongly down-regulated 53-fold in coaggregates.

The STRING DB analysis also identified strong evidence for a connection between HSIVP1_RS08035 and HSIVP1_RS04480 (Vpar_0995), encoding pyruvate dehydrogenase, which was down-regulated 5.7-fold in coaggregates. Similarly, HSIVP1_RS05665 (Vpar_1121), encoding thioredoxin and down-regulated 20-fold in coaggregates, was strongly connected with HSIVP1_RS08035 (glutamate synthase). This gene was also connected with HSIVP1_RS07895 (Vpar_1606; ribonucleoside-diphosphate reductase subunit beta) and HSIVP1_RS02075 (Vpar_0574; dihydrolipoamide dehydrogenase), which were down-regulated 15-fold and 23-fold in coaggregates, respectively. A number of other dehydrogenase and redox-active enzymes were connected to HSIVP1_RS08035 on the basis of weaker evidence.

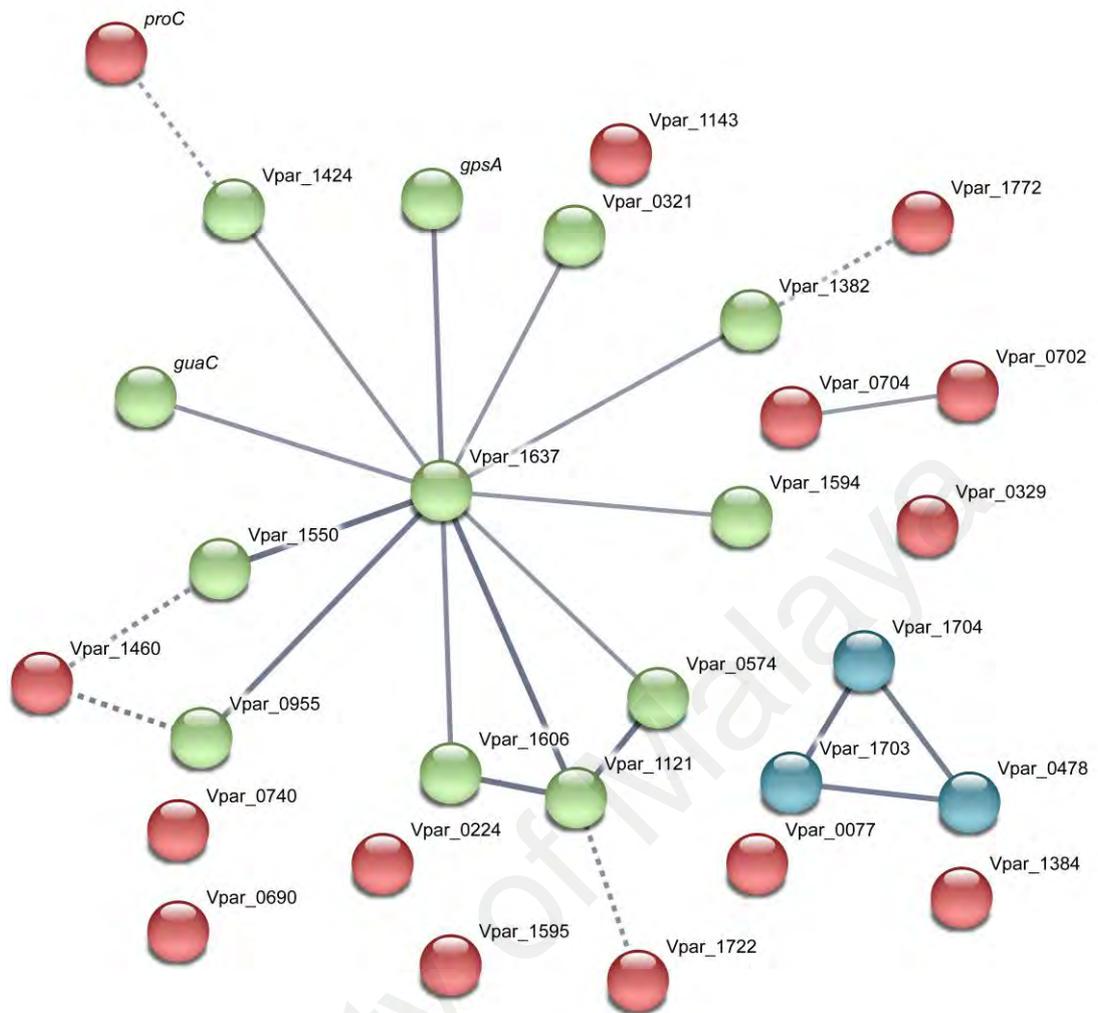


Figure 5.7: Putative protein interaction networks in the ‘oxidoreductase’ category of the *V. parvula* differentially expressed gene dataset.

Two additional gene clusters were identified by STRING DB that were not grouped with HSIVP1_RS08035 glutamate dehydrogenase. One was a pair of genes close to each other on the chromosome, based on the *V. parvula* DSM2008 genome: HSIVP1_RS03260 (Vpar_0702, *fabH*), a component of the fatty acid biosynthesis pathway, and HSIVP1_RS05740 (Vpar_0704, a radical S-adenosyl-L-methionine (SAM) protein, *queE*). Of these, *fabH* was down-regulated 4.1-fold in coaggregates, whereas *queE* was up-regulated 5-fold. The other cluster included two adjacent genes HSIVP1_RS08910 (Vpar_1703, succinate dehydrogenase/fumarate reductase, cytochrome b subunit) and HSIVP1_RS08915 (Vpar_1704, succinate

dehydrogenase/fumarate reductase iron-sulfur protein) which were down-regulated 840-fold and 11-fold, respectively. These grouped with HSIVP1_RS02130 (Vpar_0478, cytochrome ubiquinol oxidase subunit I), which was down-regulated 5.9-fold in coaggregates. All three genes encode components of the oxidative phosphorylation pathway.

The response of *V. parvula* to coaggregation was consistent with sensing decreased oxidative stress, possibly as a consequence of the removal of oxygen by *S. gordonii*. The most strongly regulated gene in *V. parvula* was HSIVP1_RS09010, encoding superoxide dismutase, which was down-regulated >2,500-fold. This enzyme is a key antioxidant defence in the phylum *Firmicutes* (Garcia et al., 2017). The thioredoxin system is another important antioxidant defence in bacteria and regulates the activity of many redox-sensitive transcriptional regulators in bacterial cells (Lu et al., 2014). It is possible that this system was responsible for driving changes in a wide range of oxidoreductase functions, with diverse metabolic effects on the cells.

In addition, HSIVP1_RS09105 encoding catalase was down-regulated 6.6-fold, although it did not quite meet the $P_{\text{adj}} < 0.05$ threshold ($P_{\text{adj}} = 0.058$, $P = 0.012$). In rich media and over a longer timeframe (6 h), *V. parvula* catalase was up-regulated 180-fold in co-cultures containing *S. gordonii* compared with monocultures (Zhou et al., 2017). It appears that the short time used here (30 min) was insufficient for *S. gordonii* to generate sufficient levels of H_2O_2 to stimulate catalase production by *V. parvula*. In keeping with this hypothesis, the *spxB* gene encoding pyruvate oxidase, the key H_2O_2 -generating enzyme of *S. gordonii* (Redanz et al., 2018a), was not regulated in response to coaggregation (<1.5-fold change, $P = 0.37$). It is also possible that catalase or peroxidases present in saliva may have inactivated H_2O_2 .

The reduction in expression of key *V. parvula* oxidative stress enzymes is consistent with a model whereby the metabolism of oxygen by *S. gordonii* cells reduces oxidative

stress for cells within the coaggregate structure at the early stages of cell-cell interactions. Oxygen gradients are known to occur in microbial biofilms and have been detected in aggregates above a threshold size of radius approximately 35 μm (Wessel *et al.*, 2014). In addition, obligate and facultative aerobic species including oral streptococci have been shown to protect obligate anaerobes, including veillonellae, and enable them to grow within aerobic biofilms (Bradshaw *et al.*, 1996; Bradshaw *et al.*, 1997).

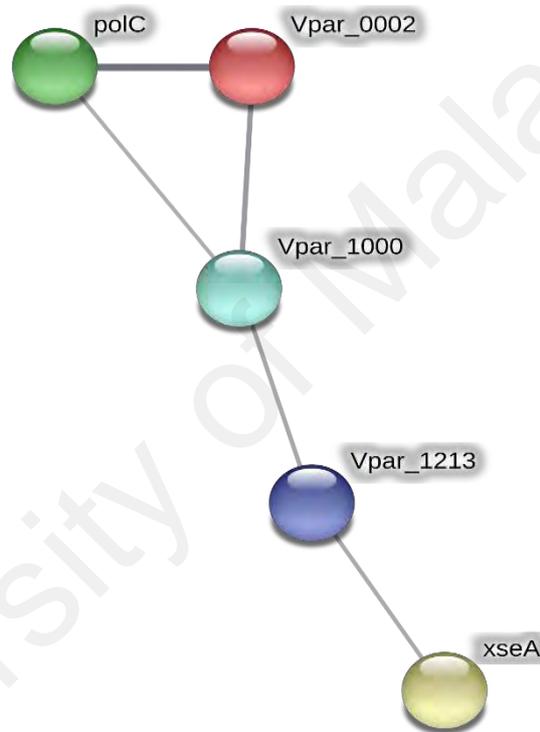


Figure 5.8: DNA metabolism genes expressed in *V. parvula* in coaggregation with *S. gordonii*

Several genes involved in mismatch repair were up-regulated between 3- and 5-fold in coaggregates, including two components of DNA polymerase III and two genes encoding exodeoxyribonucleases. Of the genes functioning in mismatch repair, only a DNA helicase was down-regulated in coaggregates. Interestingly, four genes in the DNA mismatch repair pathway were up-regulated in coaggregates.

Table 5.2: DNA metabolism genes regulated in *V. parvula* in response to coaggregation with *S. gordonii*

Gene (HSIVP1)	Vpar ID	Gene description	Fold-change	P_{adj}
<i>Mismatch repair</i>				
RS08370	Vpar_0002	DNA polymerase III	5.1	0.015
RS04315	Vpar_0919	exodeoxyribonuclease small subunit (<i>xseA</i>)	-7.8	2.5×10^{-2}
RS04495	Vpar_0958	DNA polymerase III subunit alpha (<i>polC</i>)	3.8	4.7×10^{-3}
RS05520	Vpar_1000	DNA helicase	2.8	4.2×10^{-2}
RS05790	Vpar_1213	single-stranded DNA-specific exonuclease	3.1	1.8×10^{-2}

Oxidative stress causes DNA damage and usually results in increased mismatch repair activity (Robles *et al.*, 2011). Therefore, reduction in oxidative stress might be predicted to reduce the need for mismatch repair. It is possible that DNA was protected through a different function of *S. gordonii* cells unrelated to oxygen removal. Alternatively, mismatch repair may also be employed to increase the rate of mutagenesis and facilitate rapid adaptation to a new environment (Bak *et al.*, 2014).

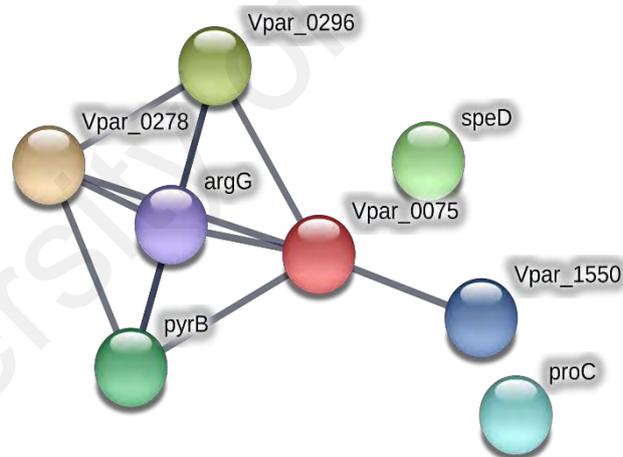


Figure 5.9: Amino acid metabolism genes expressed in *V. parvula* in coaggregation with *S. gordonii*

Amino acid metabolism genes listed in (Table 5.3) included ‘Alanine, aspartate and glutamate metabolism’ and ‘Arginine and proline metabolism’ (Figure 5.3). There was a significant down-regulation of genes involved in multiple amino acid biosynthesis pathways. Only one gene in these pathways was up-regulated: HSIVP1_RS02475 (VPar_1094, encoding ribose 5-phosphate isomerase B), 7.0-fold. The gene product

interconverts ribulose 5-phosphate and ribose 5-phosphate, which is a substrate for the biosynthesis of histidine. Genes in pathways for biosynthesis of tryptophan, phenylalanine, valine, leucine, isoleucine, threonine, aspartate, glutamate, lysine, arginine and proline were co-ordinately down-regulated in coaggregates, indicating a general reduction in amino acid biosynthesis. It is possible that coaggregation resulted in an increased availability of amino acids in the vicinity of cells. This may occur through the release of amino acids from *S. gordonii* cells.

Table 5.3: Amino acid metabolism genes regulated in *V. parvula* in response to coaggregation with *S. gordonii*

Gene (HSIVP1)	Vpar ID	Gene description	Fold-change	P_{adj}
<i>Alanine, aspartate and glutamate metabolism</i>				
RS00110	Vpar_0075	aspartate aminotransferase	-23	5.5×10^{-3}
RS01170	Vpar_0278	L-asparaginase	-21	1.8×10^{-7}
RS01260	Vpar_0296	aspartate racemase	-9.5	2.8×10^{-4}
RS02565	Vpar_0563	aspartate carbamoyltransferase (<i>pyrB</i>)	-3.8	2.8×10^{-4}
RS07610	Vpar_1550	glutamate dehydrogenase	-53	1.0×10^{-5}
RS07685	Vpar_1568	Arginine succinate synthase	-11	1.3×10^{-7}
<i>Arginine and proline metabolism</i>				
RS00110	Vpar_0075	aspartate aminotransferase	-23	5.5×10^{-3}
RS01940	Vpar_0438	s-adenosylmethionine decarboxylase proenzyme	-18	4.9×10^{-8}
RS06910	Vpar_1429	pyrroline-5-carboxylate reductase (<i>proC</i>)	-4.6	6.1×10^{-3}
RS07610	Vpar_1550	glutamate dehydrogenase	-53	1.0×10^{-5}
RS07685	Vpar_1568	Arginine succinate synthase	-11	1.3×10^{-7}

The release of the amino acid ornithine by *S. gordonii* mediates cross-feeding and promotes biofilm development in association with *F. nucleatum* (Sakanaka et al., 2015). Alternatively, *V. parvula* may scavenge amino acids from *S. gordonii* cells through the action of an extracellular protease. A similar mechanism has recently been demonstrated in the interaction between *S. gordonii* and *Actinomyces oris* (Mohammed et al., 2018). In this case, the secreted *S. gordonii* protease Challisin appears to scavenge arginine from *A. oris* cells in coaggregates, leading to repression of arginine biosynthesis gene expression in *S. gordonii*. Here, the gene encoding Challisin was up-regulated 1.8-fold

following coaggregation, indicating that it may also be important in the interaction with *V. parvula*.

University of Malaya

CHAPTER 6: TRANSCRIPTOMIC RESPONSES TO COAGGREGATION BETWEEN *STREPTOCOCCUS GORDONII* AND *STREPTOCOCCUS ORALIS*

6.1 Coaggregation of *Streptococcus gordonii* and *Streptococcus oralis*

Coaggregation between *Streptococcus gordonii* and *Streptococcus oralis* was quantitatively assessed by mixing concentrated suspension of cells in coaggregation buffer vigorously. After 30 minutes of incubation, substantial aggregates were clearly noticed in coaggregation buffer vigorously and were scored '4+' on the visual scale with absolute clear background. Coaggregation was monitored to ensure saliva did not inhibit this interaction by mixing fresh cultures of cells in freshly collected human saliva, again strong coaggregation was observed within seconds with '4+' designation in reference to standard visual scoring system.

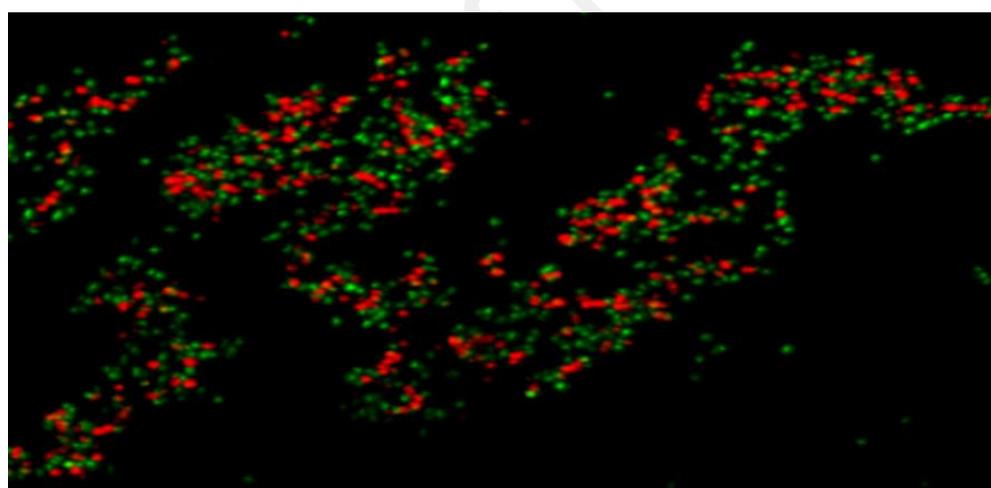


Figure 6.1 Confocal scanning laser microscopy analysis of coaggregation between *S. gordonii* and *S. oralis*. pico-green-stained *S. gordonii* cells (green) and propidium iodide-stained *S. oralis* (red) were suspended in human saliva and mixed vigorously to induce coaggregation. Samples were visualized by confocal laser scanning microscopy.

Coaggregation structures were predicted by pre-staining *S. gordonii* and *S. oralis* with fluorescent dyes and imaging by confocal laser scanning microscopy (CLSM; Figure 6.1). macroscopic aggregates, >100 μm in width, were visible and contained *S. gordonii* and *S. oralis* cells evenly mixed throughout the structure.

6.2 Dual RNA-Seq data analysis

6.2.1 RNA-Seq data generation and pre-processing

Initially, the RNA-Seq transcriptional analysis was carried out with three biological replicates of each sample for mixed and monoculture totalling 9 sequencing libraries (batch-1) results were unsatisfactory. Therefore, samples were resent for sequencing with more replicates in batch-2. RNA-Seq transcriptional analysis from a total of 16 sequencing libraries of 5 (3reps from batch-1 and 2reps from batch-2) mixed (SgSo), 6 (3reps from batch-1 and 3reps from batch-2) monoculture *S. gordonii* (Sg), and 5 (3reps from batch-1 and 2reps from batch-2) monoculture *S. oralis* (So) biological replicates generated a total 336 million paired end raw reads of 100bp read length. After the removal of low-quality reads and adapter content by Trimmomatic-0.36 obtained 314 million (clean ratio = 93%) reads.

6.2.2 Mapping reads to reference genome and transcript abundance estimation

Pre-processed reads after mapping using TopHat, for *S. gordonii* 92.93% of pre-processed reads were mapped to monoculture (Sg) and 45 % reads mapped to mixed culture (SgSo_Sg) and for *S. oralis*, 97% reads were mapped to monoculture (So) and 54% reads mapped to mixed culture as shown in Table 6.1. Read counts of mixed and monoculture samples after Trimmed Mean of M-values (TMM) normalization method showed a similar distribution of per-gene read counts per sample ensuring normalized distribution of data were quantitatively comparable between mixed and coaggregate and monoculture samples and no batch effects were apparent from the box plots (Figure 6.2).

Table 6.1 Alignment statistics of mixed and monoculture transcriptomes of *S. gordonii* and *S. oralis* in coaggregation experiment.

Sample name	Raw reads	Pre-Processed Reads	Mapped Reads
Monocultures			
Sg1	13,702,956	13,629,680	13,104,542
Sg2	15,266,848	15,185,702	14,640,981
Sg3	12,783,626	12,715,336	12,435,361
Sg4	34,884,736	32,450,626	31,924,924
Sg5	26,850,598	24,852,990	24,448,922
Sg6	303,728,44	28,134,060	21,986,544
Total reads <i>S. gordonii</i>	133,861,608	127,187,392	118,194,982 (92.93%)
So1	14,451,066	14,451,066	13,926,290
So2	15,621,300	15,619,918	14,879,371
So3	12,780,306	12,779,208	12,164,752
So4	32,061,362	29,339,116	28,867,869
So5	31,849,734	29,208,704	28,709,530
Total reads <i>S. oralis</i>	106,763,768	101,398,012	98,547,812 (97.19%)
Coaggregates			
SgSo1	13,414,958	4,269,588	1,095,399 (Sg) 2,984,283 (So)
SgSo2	15,148,212	15,146,904	5,229,367 (Sg) 9,382,155 (So)
SgSo3	12,743,546	12,742,346	5,885,024 (Sg) 6,499,028 (So)
SgSo4	33,322,784	30,982,486	15,833,006 (Sg) 15,370,738 (So)
SgSo5	24,359,624	22,479,16	15,833,006 (Sg) 12,065,953 (So)
Mixed cultures	98,989,124	85,620,486	38,563,039 (45.04%-Sg), 46,302,157(54.08%-So)
Total reads of experiment	339,614,500	314,205,890	301,607,990 (96%)

The batch variation effect of data between samples was further compared by Multi Dimensional Scaling Analysis (MDS) Appendix F. Typically, biological replicates from the same batch of *S. gordonii* and *S. oralis* monoculture with corresponding coaggregates cluster together indicating possible batch variation effects (AppendixG). However, outlier of the *S. gordonii* coaggregate samples from batch-1 were separated in MDS plot as shown in SgVp PcoA analysis. Attempts were made to repeat these samples. However, problems with batch effects made it impossible to interpret the data. Therefore, it was decided to use all the replicates from two batches for differential expression analysis on the basis that this would provide useful information to form hypotheses for further testing.

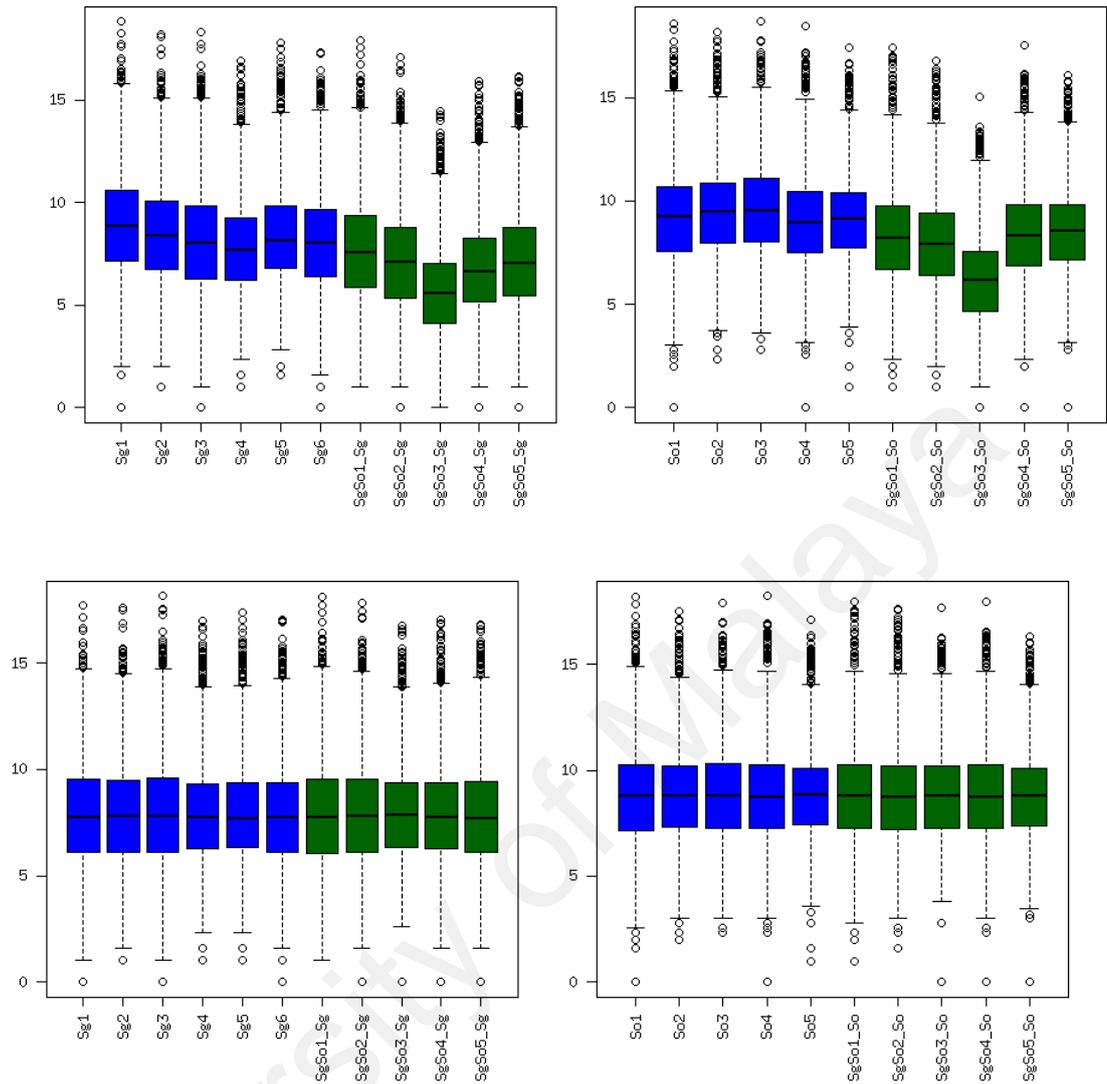


Figure 6.2: Boxplots representing quality of read counts in SgSo experiment. Distribution of raw read counts and normalized read counts. Counts were normalized by the Trimmed Mean of M-values approach. Median counts from each sample are shown as thick black lines surrounded by a box representing the interquartile range, with whiskers that extend to two standard deviations from the median. Outlying observations are shown as open circles. Data from biological replicates for each sample are shown.

6.2.3 Differential expression analysis

DESeq2 was used to compare coaggregate samples (SgSo) reads with monoculture (Sg, So) sample reads separately. A total of 72 *S. gordonii* genes (49 upregulated and 23 down regulated genes) and 22 *S. oralis* genes (8 upregulated and 14 down regulated)

genes were found to be regulated at a $p\text{-val} < 0.05$. We applied FDR-adjusted stringent level $P_{adj} < 0.05$ to identify significant differentially expressed genes for previous two experiments but in this comparison not many significant genes of interest were identified with FDR value due to batch effects. so we applied P value cut-off of $P < 0.05$. The differential expression was visualized using volcano plots (Figure 6.3).

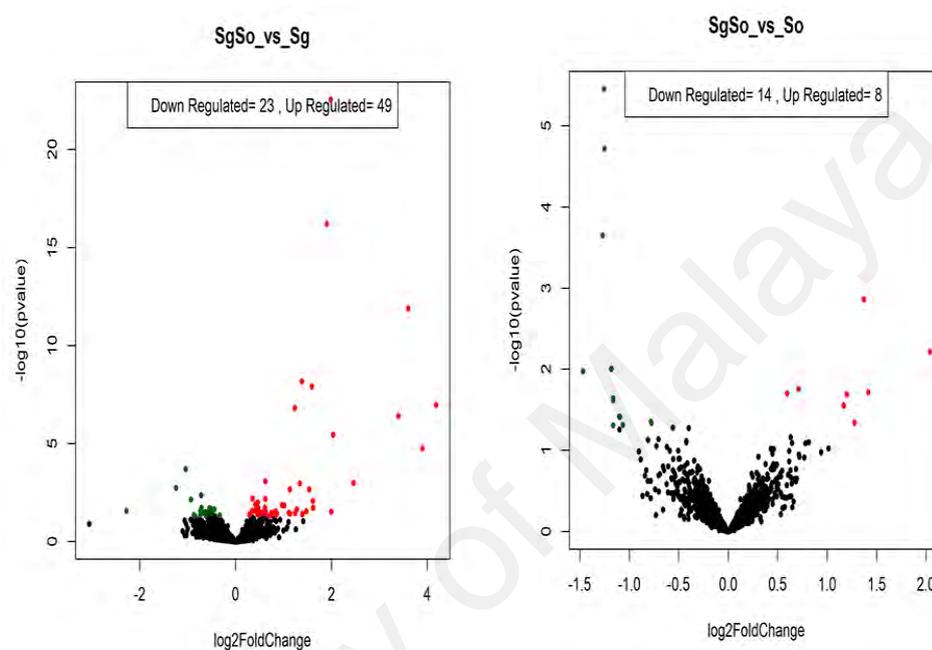


Figure 6.3: Volcano plots representing differentially expressed genes in coaggregates vs monocultures for *S. gordonii* and *S. oralis*.

The y-axis corresponds to the mean expression value of $\log_{10}(P\text{ value})$, and the x-axis displays the $\log_2(\text{fold change})$ value. The green and red dots correspond to significantly down and up-regulated genes, respectively ($P\text{ value} < 0.05$). The black dots represent the genes whose expression levels did not reach statistical significance ($P\text{ value} > 0.05$)

6.3 Discussion

6.3.1 Gene regulation in *S. gordonii* in response to coaggregation with *S. oralis*

A small cluster of genes *rpsL* (L7/L12 protein), *rpsR* (S18 protein), *rpsS* (S19 protein) and *rpmGA* (L33 protein) encoding ribosomal proteins (Figure 6.4) were down regulated from 1.3- to 2.0-fold (Table 6.2). Ribosomal protein expression is noted to associated with growth rate (Nomura *et al.*, 1984). A similar effect of down regulation on *S. oralis* ribosomal proteins by *A. geminatus* has been demonstrated in proteomic analysis of polymicrobial biofilm model (Bao *et al.*, 2017).

Table 6.2 Ribosomal proteins regulated in *S. oralis* in response to coaggregation with *S. gordonii*.

Locus ID	Predicted function (gene)	Log 2Fold change	P-value
<i>Ribosomal Proteins</i>			
Soralis34_00180	30S ribosomal protein S19 (<i>rpsS</i>)	1.37	0.001
Soralis34_01263	30S ribosomal protein S18 (<i>rpsR</i>)	2.03	0.006
Soralis34_01748	50S ribosomal protein L33 (<i>rpmGA</i>)	1.41	0.019
Soralis34_01130	50S ribosomal protein L7/L12 (<i>rplL</i>)	1.27	0.045

In this study it is possible that *S. gordonii* outcompeted *S. oralis* key nutrients in saliva and led to a decrease in *S. oralis* growth in coaggregates. Decreased expression of ribosomal proteins may led to reduced protein synthesis rate. It was shown that *trp* genes were down regulated when protein synthesis reduced (Ng *et al.*, 2003) in *E. coli*. It is also possible that coaggregation leads to decrease in protein synthesis and triggering downregulation of *trp* operon.

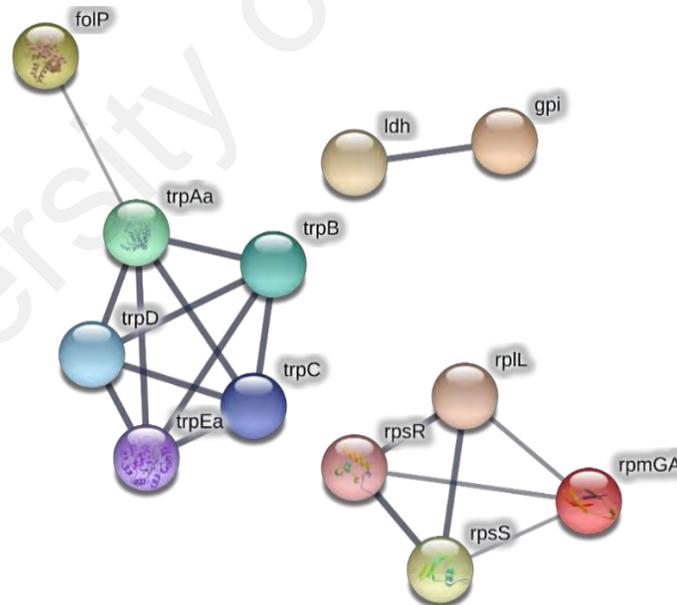


Figure 6.4 Network of genes regulated in *S. oralis* in coaggregation with *S. gordonii*.

A cluster of six genes in *trpEDCFAB_2* operon of tryptophan biosynthesis (Figure 6.4) was down regulated to 1.2 – 1.4-fold in *S. oralis* following coaggregation (Table 6.3) . This gene cluster participate in indole biosynthesis. In tryptophan deficient

conditions, the expression of *trp* operon is elevated directly influencing indole production (Gong *et al.*, 2002; Yanofsky *et al.*, 1991). *trpE* encoding tryptophan synthase is responsible for the synthesis of indoles by aldol cleavage of indolglyceroldehyde 3-phosphate to indole and glyceraldehyde-3-phosphate (Dunn, 2012). It was shown that indole controls biofilm formation by negatively regulating biofilm formation in indole producing bacteria like *E. coli* (Lee *et al.*, 2010) and in contrast indole stimulates biofilm formation in non-indole synthesizing bacteria like *P. aeruginosa* (Lee *et al.*, 2007). So, it is possible that *S.gordonii* by down regulating tryptophan operon repress indole synthesis and thus resulting dual species biofilm formation.

Table 6.3 Tryptophan metabolism genes regulated in *S. oralis* in response to coaggregation with *S. gordonii*

Locus ID	Predicted function (gene)	Log2 Fold change	P-value
<i>Tryptophan Metabolism</i>			
Soralis34_01408	Tryptophan synthase alpha chain (<i>trpE</i>)	-1.25	3.52E-06
Soralis34_01407	Dihydropteroate synthase (<i>folP</i>)	-1.251	1.91E-05
Soralis34_01406	Indole-3-glycerol phosphate synthase (<i>trpD</i>)	-1.27	0.00022
Soralis34_01405	Phosphoribosyl anthranilate isomerase (<i>trpC</i>)	-1.18	0.0099
Soralis34_01404	N-(5'-phosphoribosyl) anthranilate isomerase (<i>trpF</i>)	-1.46	0.0106
Soralis34_01402	Anthranilate synthase component (<i>trpA</i>)	-1.09	0.038
Soralis34_01409	Anthranilate phosphoribosyl transferase (<i>trpB_2</i>)	-0.77	0.045

Alternatively, a reaction using the same *trp* operon substrates along with *folP* provides the *p-aminobenzoate* (pABA) moiety for folate synthesis. Tetrahydrofolate (THF) is essential as donor and acceptor of one-carbon groups in the biosynthesis of purines and pyrimidines, formyl-methionyl t-rna^{fmet}, and some amino acids. Previously it was shown that pABA generated by *S. gordonii* facilitate accumulation of *P. gingivalis* in dual species biofilm formation (Kuboniwa *et al.*, 2017).

So, it is also possible that *S. gordonii* facilitates *pABA* and tryptophan degradation and supports biofilm formation.

6.3.2 Gene regulation in *S. oralis* in response to coaggregation with *S. gordonii*

Genes regulated in *S. gordonii* were dominated by transporter genes. A large cluster encoding transporters and two component (NisK/SpaK) regulatory system was upregulated 2-4-fold in response to coaggregation. Two component systems consist of a transmembrane sensor and response regulator that induce or repress transcription of target genes in response to an external stimulus (Bhagwat *et al.*, 2001; Mattos-Graner *et al.*, 2017). A tblastn homology analysis of our two-component system showed 33-35% similarity with *L. lactis* and *S. suis* NisK/NisR system, which are involved in sensing lantibiotics.

Table 6.4 Tryptophan metabolism genes regulated in *S. oralis* in response to coaggregation with *S. gordonii*

Locus ID	Predicted function (gene)	Log2 Fold change	P-value
<i>Transporters and Two component system</i>			
SGO_RS04510 (SGO_0920)	Cobalt ABC transporter	1.98	2.8E-23
SGO_RS04505 (SGO_0919)	Abc transporter, ATP-binding protein	1.90	6.21E-17
SGO_RS09350 (SGO_1910)	Lantibiotic abc transporter permease	3.60	1.32E-12
SGO_RS04495 (SGO_0917)	Membrane protein	1.38	6.98E-09
SGO_RS04500 (SGO_0918)	abc transporter permease	1.59	1.25E-08
SGO_RS09355 (SGO_1911)	ABC-type transporter, ATPase component	4.18	1.08E-07
SGO_RS09340 (SGO_1908)	DNA response regulator	1.33	0.0010

Previously, it was shown that mature lantibiotics in streptococci can be sensed by two component system leading to an autoinduction process. This results in production and activation of lantibiotics in neighbouring cells, and while the sensing cells remain resistant to their activity by expressing immunity peptides that bind bacteriocins/ABC transporters (Jimenez *et al.*, 2014). It can be hypothesized that proximity of *S. gordonii*

and *S. oralis* in coaggregates may enhance the interbacterial competition between them, resulting in up-regulation of sensing systems that detect competitive molecules such as lantibiotics. However, at present there is no experimental evidence regarding the role of this TCS in *S. gordonii*, and further work would be needed to confirm a function in sensing antimicrobial peptides.

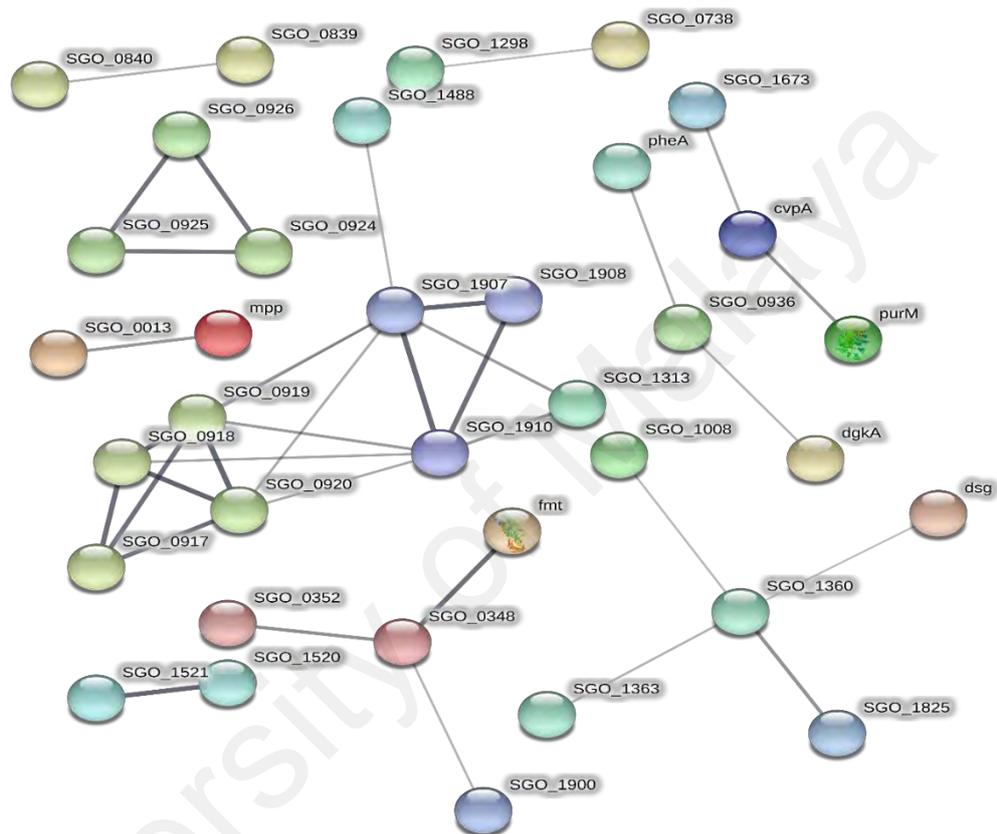


Figure 6.5 Network of genes regulated in *S. gordonii* in coaggregation with *S. oralis*.

A small cluster of genes (SGO_924-926) that appear to be an integrative or conjugative element, important to horizontal gene transfer (HGT) system that potentially can facilitate gene transfer to closely related bacteria (Frost *et al.*, 2005), were upregulated from 2-4-fold. These genes are predicted to encode DNA replicative RepB-like protein (SGO_926), Plasmid replication-like protein (SGO_925) and an insertional element (SGO_924, HsdD). Horizontal gene transfer between oral streptococci in biofilm communities has been reported (Roberts *et al.*, 2001; Warburton

et al., 2007). It is possible that cell-cell contact with *S. oralis* initiates the expression of the *S. gordonii* transposable element to increase gene transfer between the two species.

University of Malaya

CHAPTER 7: FINAL DISCUSSION AND CONCLUSIONS

The differentially expressed genes of *S. gordonii* in each study were analysed and illustrated in a three-way Venn diagram (Figure 7.1). All the genes are listed in Table 7.1. Five genes were found to be common between SgFn vs SgVp, three genes between SgFn vs SgSo and five genes between SgSo vs SgVp. This analysis indicated only one gene (SGO_RS10595) in common to the three pairs (SgFn vs SgVp vs SgSo) which encodes a hypothetical protein. Recent research studies has shown role of regulatory elements in oral biofilm formation (Duran-Pinedo *et al.*, 2015; Rabin *et al.*, 2015). It would be interesting to study this gene in further research for its role in coaggregation with other oral bacteria and biofilm formation. The genes commonly summarized in Venn diagram analyzed through STRING DB analysis were not grouped into networks.

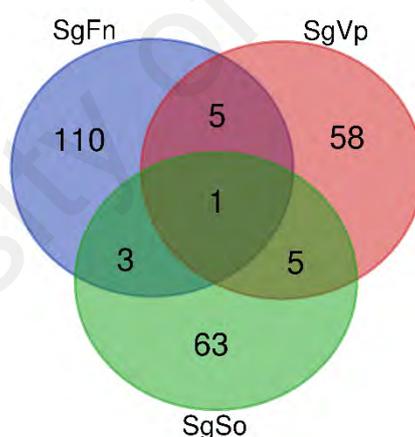


Figure 7.1 Venn diagram to represent overlapping *S. gordonii* genes between three comparisons

Among the five genes expressed common between SgFn and SgVp pairing, two genes regulated in reverse direction Regulatory protein RecX- SGO_RS03085 (SgFn 3.61 and SgVp -1.44 fold change) and Ferric transport regulator protein-SGO_RS03460 (SgFn -2.23 and SgVp 1.49 fold change). From the five genes common between SgSo and SgVp one gene regulated in opposite direction Fructokinase-

SGO_RS03395 (SgSo -4.16 and SgVp -3.03 fold change). From the three genes common between pairings SgFn vs SgSo one gene regulated in opposite direction NADPH-dependent FMN reductase- SGO_RS05760 (SgFn -1.32 and SgSo 1.52 fold change) . Genes regulated in opposite directions appear to be sensitive to regulation, but are likely to be different mechanisms driving the regulation in different pairings. Thus these are not conserved responses to coaggregation.

Regulatory protein RecX gene (SGO_RS03085) which is involved in homologous recombination commonly expressed between SgFn and SgVp pairs was upregulated 3-fold in SgFn pair (page 73) and down regulated with 1.5 fold change in SgVp. A 2-4 fold upregulation of gene cluster (Rep-B like protein) relevant to same cluster of Rec X and horizontal gene transfer mechanism was also observed in the SgSo pair. It is possible that oral bacteria in coaggregation with *S. gordonii* initiate expressing transposable elements in cell-cell contact.

Table 7.1 List of common *S. gordonii* genes between the three pairings of SgFn, SgVp and SgSo

Comparisons	No. of genes	Locus_tag	Gene Description	Log2 Fold Change SgFn/SgVp/SgSo
SgFn vs SgVp vs SgSo	1	SGO_RS10595	Truncated hypothetical protein	1.55/1.95/1.61
SgFn vs SgVp	5	SGO_RS03085	Regulatory protein RecX	3.61/-1.44
		SGO_RS04880	Oxidoreductase	-1.32/-1.48
		SGO_RS05750	FAD:Protein FMN transferase	-1.46/-1.40
		SGO_RS03460	Ferric transport regulator protein	1.42/-1.50
		SGO_RS07435	Tagatose-6-phosphate kinase	-2.23/-1.49
SgFn vs SgSo	3	SGO_RS04885	Uncharacterized protein	-1.41/-1.53
		SGO_RS05760	NADPH-dependent FMN reductase	-1.32/1.52
		SGO_RS02655	-	-1.50/-1.79
SgSo vs SgVp	5	SGO_RS01740	ABC transporter, ATP-binding protein SP1580	-6.48/-1.87
		SGO_RS08605	Fructokinase	-4.42/-1.73
		SGO_RS03395	Uncharacterized protein	-4.16/3.03
		SGO_RS03505	Diacylglycerol kinase (DAGK)	-4.28/-2.63
		SGO_RS08120	Membrane protein	-5.06/-1.74

SGO_RS03505(DGK) encoding diacylglycerol kinase common between SgSo and SgVp down regulated upto 2-4fold (page 90) in coaggregation. Diacylglycerol kinase enzyme catalyzes the formation of phosphatidic acid from diacylglycerol and ATP. It is involved in phospholipid turnover and lipoteichoic Acid (LPA) synthesis (Tsutsui *et al.*, 1991). In *Lactobacillus plantarum* LPA has shown inhibitory effect on biofilm formation on various gram positive bacteria (Kim *et al.*, 2019). It is possible that *S. oralis* and *V. parvula* promotes coaggregation formation by down regulation of diacylglycerol kinase.

In conclusion, gene regulation is very specific to each pairing and responses do not appear to be conserved. This indicates that the process of coaggregation itself is not the main driver behind gene regulation. Instead *S. gordonii* specifically senses each partner organism. This ability to distinguish between neighboring bacteria may be important for *S. gordonii* to adopt appropriately during the development of complex biofilms such as dental plaque. This study describes a range of genes and pathways in *S. gordonii-F. nucleatum*, *S. gordonii-V. parvula* and *S. gordonii-S.oralis* in response to coaggregation with each other. Coaggregation was successfully employed as a model to interpret transcriptional changes involved in biofilm formation. Predictions of compatible Putative protein interactions in coaggregation from this research work showed that gene regulation will be similar to that following incorporation into biofilms, since the structure of aggregates is similar in many ways to a biofilm The reliability of dual RNA-Seq approach for identifying and studying the genes involved in coaggregation of oral biofilm formation. In future, it will be useful to extend these studies by monitoring gene expression over a prolonged time course, as has been reported for *S. gordonii* interacting with *P. gingivalis* (Hendrickson *et al.*, 2017).

University of Malaya

REFERENCES

- Abranches, J., Nascimento, M. M., Zeng, L., Browngardt, C. M., Wen, Z. T., Rivera, M. F., & Burne, R. A. (2008). CcpA regulates central metabolism and virulence gene expression in *Streptococcus mutans*. *Journal of Bacteriology*, *190* (7), 2340-2349.
- Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11* (10), R106.
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31* (2), 166-169.
- Aprianto, R., Slager, J., Holsappel, S., & Veening, J. W. (2016). Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection. *Genome Biology*, *17* (1), 198.
- Arzmi, M. H., Dashper, S., Catmull, D., Cirillo, N., Reynolds, E. C., & McCullough, M. (2015). Coaggregation of *Candida albicans*, *Actinomyces naeslundii* and *Streptococcus mutans* is *Candida albicans* strain dependent. *FEMS Yeast Research*, *15* (5), fov038.
- Back, C. R., Douglas, S. K., Emerson, J. E., Nobbs, A. H., & Jenkinson, H. F. (2015). *Streptococcus gordonii* DL1 adhesin SspB V-region mediates coaggregation via receptor polysaccharide of *Actinomyces oris* T14V. *Molecular Oral Microbiology*, *30* (5), 411-424.
- Bak, S. T., Sakellariou, D., & Pena-Diaz, J. (2014). The dual nature of mismatch repair as antimutator and mutator: for better or for worse. *Frontiers in Genetics*, *5*, 287.
- Bao, K., Bostanci, N., Thurnheer, T., & Belibasakis, G. N. (2017). Proteomic shifts in multi-species oral biofilms caused by *Anaeroglobus geminatus*. *Scientific Reports - Nature*, *7* (1), 4409.
- Behnke, S., & Camper, A. K. (2012). Chlorine dioxide disinfection of single and dual species biofilms, detached biofilm and planktonic cells. *Biofouling*, *28* (6), 635-647.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *57* (1), 289-300.
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., . . . Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, *456* (7218), 53-59.
- Bhagwat, S. P., Nary, J., & Burne, R. A. (2001). Effects of mutating putative two-component systems on biofilm formation by *Streptococcus mutans* UA159. *FEMS microbiology letters*, *205* (2), 225-230.

- Bik, E. M., Long, C. D., Armitage, G. C., Loomer, P., Emerson, J., Mongodin, E. F., . . . Relman, D. A. (2010). Bacterial diversity in the oral cavity of 10 healthy individuals. *International Society for Microbial Ecology Journal*, 4 (8), 962-974.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30 (15), 2114-2120.
- Bos, R., Vandermei, H. C., Meinders, J. M., & Busscher, H. J. (1994). A quantitative method to study co-adhesion of microorganisms in a parallel-plate flow chamber - basic principles of the analysis. *Journal of Microbiological Methods*, 20 (4), 289-305.
- Bowen, W. H., Burne, R. A., Wu, H., & Koo, H. (2018). Oral Biofilms: Pathogens, Matrix, and Polymicrobial Interactions in Microenvironments. *Trends Microbiol*, 26 (3), 229-242.
- Bradshaw, D. J., Marsh, P. D., Allison, C., & Schilling, K. M. (1996). Effect of oxygen, inoculum composition and flow rate on development of mixed-culture oral biofilms. *Microbiology*, 142 (Pt 3), 623-629.
- Bradshaw, D. J., Marsh, P. D., Watson, G. K., & Allison, C. (1997). Oral anaerobes cannot survive oxygen stress without interacting with facultative/aerobic species as a microbial community. *Letters in Applied Microbiology*, 25 (6), 385-387.
- Bradshaw, D. J., Marsh, P. D., Watson, G. K., & Allison, C. (1998). Role of *Fusobacterium nucleatum* and coaggregation in anaerobe survival in planktonic and biofilm oral microbial communities during aeration. *Infection Immunology*, 66 (10), 4729-4732.
- Brown, S. A., & Whiteley, M. (2007). A novel exclusion mechanism for carbon resource partitioning in *Aggregatibacter actinomycetemcomitans*. *Journal of Bacteriology*, 189 (17), 6407-6414.
- Bullard, B., Lipski, S., & Lafontaine, E. R. (2007). Regions important for the adhesin activity of *Moraxella catarrhalis* Hag. *BMC Microbiology*, 7, 65.
- Burby, P. E., & Simmons, L. A. (2017). MutS2 Promotes Homologous Recombination in *Bacillus subtilis*. *Journal of Bacteriology*, 199 (2).
- Burmolle, M., Webb, J. S., Rao, D., Hansen, L. H., Sorensen, S. J., & Kjelleberg, S. (2006). Enhanced biofilm formation and increased resistance to antimicrobial agents and bacterial invasion are caused by synergistic interactions in multispecies biofilms. *Applied and Environmental Microbiology*, 72 (6), 3916-3923.
- Byers, H. L., Homer, K. A., & Beighton, D. (1996). Utilization of sialic acid by viridans streptococci. *Journal of Dental Research*, 75 (8), 1564-1571.
- Caing-Carlsson, R., Goyal, P., Sharma, A., Ghosh, S., Setty, T. G., North, R. A., . . . Ramaswamy, S. (2017). Crystal structure of N-acetylmannosamine kinase from *Fusobacterium nucleatum*. *Acta Crystallographica Section F Structural Biology Communications*, 73 (Pt 6), 356-362.

- Clemans, D. L., Kolenbrander, P. E., DeBabov, D. V., Zhang, Q., Lunsford, R. D., Sakone, H., . . . Neuhaus, F. C. (1999). Insertional inactivation of genes responsible for the D-alanylation of lipoteichoic acid in *Streptococcus gordonii* DL1 (Challis) affects intrageneric coaggregations. *Infection Immunology*, *67* (5), 2464-2474.
- Copenhagen-Glazer, S., Sol, A., Abed, J., Naor, R., Zhang, X., Han, Y. W., & Bachrach, G. (2015). Fap2 of *Fusobacterium nucleatum* is a galactose-inhibitable adhesin involved in coaggregation, cell adhesion, and preterm birth. *Infection Immunology*, *83* (3), 1104-1113.
- Demuth, D. R., Duan, Y., Brooks, W., Holmes, A. R., McNab, R., & Jenkinson, H. F. (1996). Tandem genes encode cell-surface polypeptides SspA and SspB which mediate adhesion of the oral bacterium *Streptococcus gordonii* to human and bacterial receptors. *Molecular Microbiology*, *20* (2), 403-413.
- Demuth, D. R., & Irvine, D. C. (2002). Structural and functional variation within the alanine-rich repetitive domain of streptococcal antigen I/II. *Infection Immunology*, *70* (11), 6389-6398.
- Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C., Yu, W. H., . . . Wade, W. G. (2010). The human oral microbiome. *Journal of Bacteriology*, *192* (19), 5002-5017.
- Diaz, P. I., Chalmers, N. I., Rickard, A. H., Kong, C., Milburn, C. L., Palmer, R. J., & Kolenbrander, P. E. (2006). Molecular characterization of subject-specific oral microflora during initial colonization of enamel. *Applied and Environmental Microbiology*, *72* (4), 2837-2848.
- Diaz, P. I., Dupuy, A. K., Abusleme, L., Reese, B., Obergfell, C., Choquette, L., . . . Strausbaugh, L. D. (2012a). Using high throughput sequencing to explore the biodiversity in oral bacterial communities. *Molecular Oral Microbiology*, *27* (3), 182-201.
- Diaz, P. I., Xie, Z., Sobue, T., Thompson, A., Biyikoglu, B., Ricker, A., . . . Dongari-Bagtzoglou, A. (2012b). Synergistic interaction between *Candida albicans* and commensal oral streptococci in a novel in vitro mucosal model. *Infection Immunology*, *80* (2), 620-632.
- Diaz, P. I., Zilm, P. S., & Rogers, A. H. (2002). *Fusobacterium nucleatum* supports the growth of *Porphyromonas gingivalis* in oxygenated and carbon-dioxide-depleted environments. *Microbiology*, *148* (Pt 2), 467-472.
- Dobin, A., & Gingeras, T. R. (2015). Mapping RNA-seq Reads with STAR. *Current Protocols in Bioinformatics*, *51*, 11 14 11-19.
- Dong, Y., Chen, Y. Y., & Burne, R. A. (2004). Control of expression of the arginine deiminase operon of *Streptococcus gordonii* by CcpA and Flp. *Journal of Bacteriology*, *186* (8), 2511-2514.

- Du, L. D., & Kolenbrander, P. E. (2000). Identification of saliva-regulated genes of *Streptococcus gordonii* DL1 by differential display using random arbitrarily primed PCR. *Infection Immunology*, 68 (8), 4834-4837.
- Dunn, M. F. (2012). Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase holoenzyme complex. *Archives Biochemistry Biophysics*, 519 (2), 154-166.
- Duran-Pinedo, A. E., Yost, S., & Frias-Lopez, J. (2015). Small RNA Transcriptome of the Oral Microbiome during Periodontitis Progression. *Applied Environmental Microbiology*, 81 (19), 6688-6699.
- Dutton, L. C., Paszkiewicz, K. H., Silverman, R. J., Splatt, P. R., Shaw, S., Nobbs, A. H., . . . Ramsdale, M. (2016). Transcriptional landscape of trans-kingdom communication between *Candida albicans* and *Streptococcus gordonii*. *Molecular Oral Microbiology*, 31 (2), 136-161.
- Egland, P. G., Du, L. D., & Kolenbrander, P. E. (2001). Identification of independent *Streptococcus gordonii* SspA and SspB functions in coaggregation with *Actinomyces naeslundii*. *Infection Immunology*, 69 (12), 7512-7516.
- Egland, P. G., Palmer, R. J., Jr., & Kolenbrander, P. E. (2004). Interspecies communication in *Streptococcus gordonii*-*Veillonella atypica* biofilms: signaling in flow conditions requires juxtaposition. *Proceedings of National Academy of Sciences U S A*, 101 (48), 16917-16922.
- Fei, T., Zhang, T., Shi, W., & Yu, T. (2018). Mitigating the adverse impact of batch effects in sample pattern detection. *Bioinformatics*, 34 (15), 2634-2641.
- Foster, J. S., & Kolenbrander, P. E. (2004). Development of a multispecies oral bacterial community in a saliva-conditioned flow cell. *Applied and Environmental Microbiology*, 70 (7), 4340-4348.
- Frost, L. S., Leplae, R., Summers, A. O., & Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology*, 3 (9), 722-732.
- Garcia, Y. M., Barwinska-Sendra, A., Tarrant, E., Skaar, E. P., Waldron, K. J., & Kehl-Fie, T. E. (2017). A superoxide dismutase capable of functioning with iron or manganese promotes the resistance of *Staphylococcus aureus* to calprotectin and nutritional immunity. *PLoS Pathogens*, 13 (1), e1006125.
- Ge, X., Rodriguez, R., Trinh, M., Gunsolley, J., & Xu, P. (2013). Oral microbiome of deep and shallow dental pockets in chronic periodontitis. *Plos One*, 8 (6), e65520.
- Gene Ontology, C. (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Research*, 43 (Database issue), D1049-1056.
- Ghigo, J. M. (2001). Natural conjugative plasmids induce bacterial biofilm development. *Nature*, 412 (6845), 442-445.

- Gibbons, R. J., & Nygaard, M. (1970). Interbacterial aggregation of plaque bacteria. *Archives of Oral Biology*, 15 (12), 1397-1400.
- Gong, F., & Yanofsky, C. (2002). Analysis of tryptophanase operon expression in vitro: accumulation of TnaC-peptidyl-tRNA in a release factor 2-depleted S-30 extract prevents Rho factor action, simulating induction. *The Journal of Biological Chemistry*, 277 (19), 17095-17100.
- Gotz, S., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., . . . Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36 (10), 3420-3435.
- Gross, E. L., Beall, C. J., Kutsch, S. R., Firestone, N. D., Leys, E. J., & Griffen, A. L. (2012). Beyond *Streptococcus mutans*: Dental caries onset linked to multiple species by 16S rRNA community analysis. *Plos One*, 7 (10).
- Guggenheim, M., Shapiro, S., Gmur, R., & Guggenheim, B. (2001). Spatial arrangements and associative behavior of species in an in vitro oral biofilm model. *Applied and Environmental Microbiology*, 67 (3), 1343-1350.
- Hardcastle, T. J., & Kelly, K. A. (2010). baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics*, 11, 422.
- Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C., & Ala'Aldeen, D. (2004). Type V protein secretion pathway: the autotransporter story. *Microbiology and Molecular Biology Reviews*, 68 (4), 692-744.
- Hendrickson, E. L., Beck, D. A., Miller, D. P., Wang, Q., Whiteley, M., Lamont, R. J., & Hackett, M. (2017). Insights into Dynamic Polymicrobial Synergy Revealed by Time-Coursed RNA-Seq. *Front Microbiol*, 8, 261.
- Hendrickson, E. L., Wang, T., Beck, D. A., Dickinson, B. C., Wright, C. J., R, J. L., & Hackett, M. (2014). Proteomics of *Fusobacterium nucleatum* within a model developing oral microbial community. *Microbiologyopen*, 3 (5), 729-751.
- Hendrickson, E. L., Wang, T., Dickinson, B. C., Whitmore, S. E., Wright, C. J., Lamont, R. J., & Hackett, M. (2012). Proteomics of *Streptococcus gordonii* within a model developing oral microbial community. *BMC Microbiology*, 12, 211.
- Heras, B., Totsika, M., Peters, K. M., Paxman, J. J., Gee, C. L., Jarrott, R. J., . . . Schembri, M. A. (2014). The antigen 43 structure reveals a molecular Velcro-like mechanism of autotransporter-mediated bacterial clumping. *Proceedings of the National Academy of Sciences U S A*, 111 (1), 457-462.
- Huang, R., Li, M., & Gregory, R. L. (2011). Bacterial interactions in dental biofilm. *Virulence*, 2 (5), 435-444.
- Hughes, C. V., Roseberry, C. A., & Kolenbrander, P. E. (1990). Isolation and Characterization of Coaggregation-Defective Mutants of *Veillonella atypica*. *Archives of Oral Biology*, 35, S123-S125.

- International Human Genome Sequencing, C. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431 (7011), 931-945.
- Jakubovics, N. S. (2010). Talk of the town: interspecies communication in oral biofilms. *Molecular Oral Microbiology*, 25 (1), 4-14.
- Jakubovics, N. S., Gill, S. R., Iobst, S. E., Vickerman, M. M., & Kolenbrander, P. E. (2008a). Regulation of gene expression in a mixed-genus community: stabilized arginine biosynthesis in *Streptococcus gordonii* by coaggregation with *Actinomyces naeslundii*. *Journal of Bacteriology*, 190 (10), 3646-3657.
- Jakubovics, N. S., Gill, S. R., Vickerman, M. M., & Kolenbrander, P. E. (2008b). Role of hydrogen peroxide in competition and cooperation between *Streptococcus gordonii* and *Actinomyces naeslundii*. *FEMS Microbiology Ecology*, 66 (3), 637-644.
- Jakubovics, N. S., Robinson, J. C., Samarian, D. S., Kolderman, E., Yassin, S. A., Bettampadi, D., . . . Rickard, A. H. (2015). Critical roles of arginine in growth and biofilm development by *Streptococcus gordonii*. *Molecular Microbiology*, 97 (2), 281-300.
- Jenkinson, H. F. (2011). Beyond the oral microbiome. *Environmental Microbiology*, 13 (12), 3077-3087.
- Jimenez, J. C., & Federle, M. J. (2014). Quorum sensing in group A *Streptococcus*. *Frontiers in Cellular and Infection Microbiology*, 4, 127.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44 (D1), D457-462.
- Kaplan, A., Kaplan, C. W., He, X., McHardy, I., Shi, W., & Lux, R. (2014). Characterization of *aid1*, a novel gene involved in *Fusobacterium nucleatum* interspecies interactions. *Microbial Ecology*, 68 (2), 379-387.
- Kaplan, C. W., Lux, R., Haake, S. K., & Shi, W. (2009). The *Fusobacterium nucleatum* outer membrane protein RadD is an arginine-inhibitable adhesin required for inter-species adherence and the structured architecture of multispecies biofilm. *Molecular Microbiology*, 71 (1), 35-47.
- Kara, D., Luppens, S. B., & Cate, J. M. (2006). Differences between single- and dual-species biofilms of *Streptococcus mutans* and *Veillonella parvula* in growth, acidogenicity and susceptibility to chlorhexidine. *European journal of oral sciences*, 114 (1), 58-63.
- Katharios-Lanwermyer, S., Xi, C., Jakubovics, N. S., & Rickard, A. H. (2014). Mini-review: Microbial coaggregation: ubiquity and implications for biofilm development. *Biofouling*, 30 (10), 1235-1251.
- Keijser, B. J. F., Zaura, E., Huse, S. M., van der Vossen, J. M. B. M., Schuren, F. H. J., Montijn, R. C., . . . Crielaard, W. (2008). Pyrosequencing analysis of the Oral Microflora of healthy adults. *Journal of Dental Research*, 87 (11), 1016-1020.

- Kim, A. R., Ahn, K. B., Yun, C. H., Park, O. J., Perinpanayagam, H., Yoo, Y. J., . . . Han, S. H. (2019). *Lactobacillus plantarum* Lipoteichoic Acid Inhibits Oral Multispecies Biofilm. *Journal of endodontics*, *45* (3), 310-315.
- Kim, D., Perteza, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14* (4), R36.
- Klemm, P., Hjerrild, L., Gjermansen, M., & Schembri, M. A. (2004). Structure-function analysis of the self-recognizing Antigen 43 autotransporter protein from *Escherichia coli*. *Molecular Microbiology*, *51* (1), 283-296.
- Kolenbrander, P. E. (1988). Intergeneric coaggregation among human oral bacteria and ecology of dental plaque. *Annual Review of Microbiology*, *42*, 627-656.
- Kolenbrander, P. E. (1989). Surface recognition among oral bacteria: multigeneric coaggregations and their mediators. *Critical Reviews in Microbiology*, *17* (2), 137-159.
- Kolenbrander, P. E. (2000). Oral microbial communities: biofilms, interactions, and genetic systems. *Annual Review of Microbiology*, *54*, 413-437.
- Kolenbrander, P. E., & Andersen, R. N. (1986). Multigeneric aggregations among oral bacteria: a network of independent cell-to-cell interactions. *Journal of Bacteriology*, *168* (2), 851-859.
- Kolenbrander, P. E., Andersen, R. N., Blehert, D. S., Egland, P. G., Foster, J. S., & Palmer, R. J. (2002). Communication among oral bacteria. *Microbiology and Molecular Biology Reviews*, *66* (3), 486-+.
- Kolenbrander, P. E., Palmer, R. J., Jr., Periasamy, S., & Jakubovics, N. S. (2010). Oral multispecies biofilm development and the key role of cell-cell distance. *Nature Reviews Microbiology*, *8* (7), 471-480.
- Kuboniwa, M., Houser, J. R., Hendrickson, E. L., Wang, Q., Alghamdi, S. A., Sakanaka, A., . . . Lamont, R. J. (2017). Metabolic crosstalk regulates *Porphyromonas gingivalis* colonization and virulence during oral polymicrobial infection. *Nature Microbiology*, *2* (11), 1493-1499.
- Lamont, R. J., & Rosan, B. (1990). Adherence of mutans streptococci to other oral bacteria. *Infection Immunology*, *58* (6), 1738-1743.
- Lancy, P., Jr., Appelbaum, B., Holt, S. C., & Rosan, B. (1980). Quantitative in vitro assay for "corncob" formation. *Infection Immunology*, *29* (2), 663-670.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9* (4), 357-359.
- Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15* (2), R29.

- Lee, J., Jayaraman, A., & Wood, T. K. (2007). Indole is an inter-species biofilm signal mediated by SdiA. *BMC Microbiology*, 7, 42.
- Lee, J. H., & Lee, J. (2010). Indole as an intercellular signal in microbial communities. *FEMS Microbiology Reviews*, 34 (4), 426-444.
- Leng, N., Dawson, J. A., Thomson, J. A., Ruotti, V., Rissman, A. I., Smits, B. M., . . . Kendzierski, C. (2013). EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics*, 29 (8), 1035-1043.
- Lima, B. P., Shi, W., & Lux, R. (2017). Identification and characterization of a novel *Fusobacterium nucleatum* adhesin involved in physical interaction and biofilm formation with *Streptococcus gordonii*. *Microbiologyopen*, 6 (3).
- Lipski, S. L., Akimana, C., Timpe, J. M., Wooten, R. M., & Lafontaine, E. R. (2007). The *Moraxella catarrhalis* autotransporter McaP is a conserved surface protein that mediates adherence to human epithelial cells through its N-terminal passenger domain. *Infection Immunology*, 75 (1), 314-324.
- Liu, J., Xie, Z., Merritt, J., & Qi, F. (2012). Establishment of a tractable genetic transformation system in *Veillonella* spp. *Applied and Environmental Microbiology*, 78 (9), 3488-3491.
- Liu, J. M., Wu, C. G., Huang, I. H., Merritt, J., & Qi, F. X. (2011). Differential response of *Streptococcus mutans* towards friend and foe in mixed-species cultures. *Microbiology Society Journals*, 157, 2433-2444.
- Lof, M., Janus, M. M., & Krom, B. P. (2017). Metabolic Interactions between Bacteria and Fungi in Commensal Oral Biofilms. *Journal of Fungi (Basel)*, 3 (3).
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15 (12), 550.
- Lu, J., & Holmgren, A. (2014). The thioredoxin antioxidant system. *Free Radical Biology & Medicine*, 66, 75-87.
- Luppens, S. B., Kara, D., Bandounas, L., Jonker, M. J., Wittink, F. R., Bruning, O., . . . Crielaard, W. (2008). Effect of *Veillonella parvula* on the antimicrobial resistance and gene expression of *Streptococcus mutans* grown in a dual-species biofilm. *Oral Microbiology and Immunology*, 23 (3), 183-189.
- Mancl, K. A., Kirsner, R. S., & Ajdic, D. (2013). Wound biofilms: Lessons learned from oral biofilms. *Wound Repair and Regeneration*, 21 (3), 352-362.
- Mardis, E. R. (2013). Next-generation sequencing platforms. *Annual review of analytical chemistry (Palo Alto Calif)*, 6, 287-303.
- Mattos-Graner, R. O., & Duncan, M. J. (2017). Two-component signal transduction systems in oral bacteria. *Journal of Oral Microbiology*, 9 (1), 1400858.

- Mays, T. D., Holdeman, L. V., Moore, W. E. C., Rogosa, M., & Johnson, J. L. (1982). Taxonomy of the Genus *Veillonella* Prevot. *International Journal of Systematic Bacteriology*, 32 (1), 28-36.
- McIntire, F. C., Vatter, A. E., Baros, J., & Arnold, J. (1978). Mechanism of coaggregation between *Actinomyces viscosus* T14V and *Streptococcus sanguis* 34. *Infection Immunology*, 21 (3), 978-988.
- McNab, R., Forbes, H., Handley, P. S., Loach, D. M., Tannock, G. W., & Jenkinson, H. F. (1999). Cell wall-anchored CshA polypeptide (259 kilodaltons) in *Streptococcus gordonii* forms surface fibrils that confer hydrophobic and adhesive properties. *Journal of Bacteriology*, 181 (10), 3087-3095.
- McNab, R., Ford, S. K., El-Sabaeny, A., Barbieri, B., Cook, G. S., & Lamont, R. J. (2003). LuxS-based signaling in *Streptococcus gordonii*: Autoinducer 2 controls carbohydrate metabolism and biofilm formation with *Porphyromonas gingivalis*. *Journal of Bacteriology*, 185 (1), 274-284.
- McNab, R., Holmes, A. R., Clarke, J. M., Tannock, G. W., & Jenkinson, H. F. (1996). Cell surface polypeptide CshA mediates binding of *Streptococcus gordonii* to other oral bacteria and to immobilized fibronectin. *Infection Immunology*, 64 (10), 4204-4210.
- Merritt, J., Niu, G., Okinaga, T., & Qi, F. (2009). Autoaggregation response of *Fusobacterium nucleatum*. *Applied and Environmental Microbiology*, 75 (24), 7725-7733.
- Metzger, Z., Blasbalg, J., Dotan, M., & Weiss, E. I. (2009). Enhanced attachment of *Porphyromonas gingivalis* to human fibroblasts mediated by *Fusobacterium nucleatum*. *Journal of endodontics*, 35 (1), 82-85.
- Minoche, A. E., Dohm, J. C., & Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biology*, 12 (11), R112.
- Mohammed, W. K., Krasnogor, N., & Jakubovics, N. S. (2018). *Streptococcus gordonii* Challisin protease is required for sensing cell-cell contact with *Actinomyces oris*. *FEMS Microbiology and Ecology*, 94 (5).
- Moore, W. E., & Moore, L. V. (1994). The bacteria of periodontal diseases. *Periodontology 2000*, 5, 66-77.
- Mutha, N. V. R., Mohammed, W. K., Krasnogor, N., Tan, G. Y. A., Choo, S. W., & Jakubovics, N. S. (2018). Transcriptional responses of *Streptococcus gordonii* and *Fusobacterium nucleatum* to coaggregation. *Molecular Oral Microbiology*, 33 (6), 450-464.
- Ng, W. L., Kazmierczak, K. M., Robertson, G. T., Gilmour, R., & Winkler, M. E. (2003). Transcriptional regulation and signature patterns revealed by microarray analyses of *Streptococcus pneumoniae* R6 challenged with sublethal concentrations of translation inhibitors. *Journal of Bacteriology*, 185 (1), 359-370.

- Nobbs, A. H., Jenkinson, H. F., & Jakubovics, N. S. (2011). Stick to your gums: mechanisms of oral microbial adherence. *Journal of Dental Research*, *90* (11), 1271-1278.
- Nomura, M., Gourse, R., & Baughman, G. (1984). Regulation of the synthesis of ribosomes and ribosomal components. *Annual Review of Biochemistry*, *53*, 75-117.
- Nuss, A. M., Beckstette, M., Pimenova, M., Schmuhl, C., Opitz, W., Pisano, F., . . . Dersch, P. (2017). Tissue dual RNA-seq allows fast discovery of infection-specific functions and riboregulators shaping host-pathogen transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America*, *114* (5), E791-E800.
- Nyvad, B., & Kilian, M. (1990a). Comparison of the initial streptococcal microflora on dental enamel in caries-active and in caries-inactive individuals. *Caries Research*, *24* (4), 267-272.
- Nyvad, B., & Kilian, M. (1990b). Microflora associated with experimental root surface caries in humans. *Infection Immunology*, *58* (6), 1628-1633.
- Palmer, R. J., Jr., Gordon, S. M., Cisar, J. O., & Kolenbrander, P. E. (2003). Coaggregation-mediated interactions of streptococci and actinomyces detected in initial human dental plaque. *Journal of Bacteriology*, *185* (11), 3400-3409.
- Palmer, R. J., Jr., Kazmerzak, K., Hansen, M. C., & Kolenbrander, P. E. (2001). Mutualism versus independence: strategies of mixed-species oral biofilms in vitro using saliva as the sole nutrient source. *Infection Immunology*, *69* (9), 5794-5804.
- Parashar, A., Parashar, S., Zingade, A., Gupta, S., & Sanikop, S. (2015). Interspecies communication in oral biofilm: An ocean of information. *Oral Science International*, *12* (2), 37-42.
- Periasamy, S., & Kolenbrander, P. E. (2009). Mutualistic biofilm communities develop with *Porphyromonas gingivalis* and initial, early, and late colonizers of enamel. *Journal of Bacteriology*, *191* (22), 6804-6811.
- Periasamy, S., & Kolenbrander, P. E. (2010). Central Role of the Early Colonizer *Veillonella* sp in Establishing Multispecies Biofilm Communities with Initial, Middle, and Late Colonizers of Enamel. *Journal of Bacteriology*, *192* (12), 2965-2972.
- Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., . . . Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, *13*, 341.
- Rabin, N., Zheng, Y., Opoku-Temeng, C., Du, Y., Bonsu, E., & Sintim, H. O. (2015). Biofilm formation mechanisms and targets for developing antibiofilm agents. *Future Medicinal Chemistry*, *7* (4), 493-512.

- Rathod, S. R., Khan, F., Kolte, A. P., & Gupta, M. (2014). Estimation of salivary and serum total sialic Acid levels in periodontal health and disease. *Journal of Clinical and Diagnostic Research*, 8 (9), ZC19-21.
- Redanz, S., Cheng, X., Giacaman, R. A., Pfeifer, C. S., Merritt, J., & Kreth, J. (2018a). Live and let die: Hydrogen peroxide production by the commensal flora and its role in maintaining a symbiotic microbiome. *Molecular Oral Microbiology*, 33 (5), 337-352.
- Redanz, S., Masilamani, R., Cullin, N., Giacaman, R. A., Merritt, J., & Kreth, J. (2018b). Distinct regulatory role of carbon catabolite protein A (CcpA) in oral Streptococcal spxB expression. *Journal of Bacteriology* 200 (8).
- Rickard, A. H., Gilbert, P., High, N. J., Kolenbrander, P. E., & Handley, P. S. (2003). Bacterial coaggregation: an integral process in the development of multi-species biofilms. *Trends in Microbiology*, 11 (2), 94-100.
- Rickard, A. H., Palmer, R. J., Jr., Blehert, D. S., Campagna, S. R., Semmelhack, M. F., Eglund, P. G., . . . Kolenbrander, P. E. (2006). Autoinducer 2: a concentration-dependent signal for mutualistic bacterial biofilm growth. *Molecular Microbiology*, 60 (6), 1446-1456.
- Roberts, A. P., Cheah, G., Ready, D., Pratten, J., Wilson, M., & Mullany, P. (2001). Transfer of TN916-like elements in microcosm dental plaques. *Antimicrobial Agents Chemotherapy*, 45 (10), 2943-2946.
- Robinson, J. C., Rostami, N., Casement, J., Vollmer, W., Rickard, A. H., & Jakubovics, N. S. (2018). ArcR modulates biofilm formation in the dental plaque colonizer *Streptococcus gordonii*. *Molecular Oral Microbiology*, 33 (2), 143-154.
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26 (1), 139-140.
- Robles, A. G., Reid, K., Roy, F., & Fletcher, H. M. (2011). *Porphyromonas gingivalis* mutY is involved in the repair of oxidative stress-induced DNA mispairing. *Molecular Oral Microbiology*, 26 (3), 175-186.
- Rogers, J. D., Palmer, R. J., Kolenbrander, P. E., & Scannapieco, F. A. (2001). Role of *Streptococcus gordonii* amylase-binding protein a in adhesion to hydroxyapatite, starch metabolism, and biofilm formation. *Infection Immunology*, 69 (11), 7046-7056.
- Rogosa, M., & Bishop, F. S. (1964). The genus *Veillonella*. 3. Hydrogen sulfide production by growing cultures. *Journal of Bacteriology*, 88, 37-41.
- Rosan, B., & Lamont, R. J. (2000). Dental plaque formation. *Microbes Infect*, 2 (13), 1599-1607.

- Rosen, G., & Sela, M. N. (2006). Coaggregation of *Porphyromonas gingivalis* and *Fusobacterium nucleatum* PK 1594 is mediated by capsular polysaccharide and lipopolysaccharide. *FEMS microbiology letters*, 256 (2), 304-310.
- Sakanaka, A., Kuboniwa, M., Takeuchi, H., Hashino, E., & Amano, A. (2015). Arginine-Ornithine antiporter ArcD controls arginine metabolism and interspecies biofilm development of *Streptococcus gordonii*. *Journal of Biological Chemistry*, 290 (35), 21185-21198.
- Shaniztki, B., Ganeshkumar, N., & Weiss, E. I. (1998). Characterization of a novel N-acetylneuraminic acid-specific *Fusobacterium nucleatum* PK1594 adhesin. *Oral Microbiology and Immunology*, 13 (1), 47-50.
- Sherlock, O., Schembri, M. A., Reisner, A., & Klemm, P. (2004). Novel roles for the AIDA adhesin from diarrheagenic *Escherichia coli*: cell aggregation and biofilm formation. *Journal of Bacteriology*, 186 (23), 8058-8065.
- Shu, M., Browngardt, C. M., Chen, Y. Y., & Burne, R. A. (2003). Role of urease enzymes in stability of a 10-species oral biofilm consortium cultivated in a constant-depth film fermenter. *Infection Immunology*, 71 (12), 7188-7192.
- Simoës, L. C., Simoës, M., & Vieira, M. J. (2007). Biofilm interactions between distinct bacterial genera isolated from drinking water. *Applied and Environmental Microbiology*, 73 (19), 6192-6200.
- Socransky, S. S., Smith, C., & Haffajee, A. D. (2002). Subgingival microbial profiles in refractory periodontal disease. *Journal of Clinical Periodontology*, 29 (3), 260-268.
- Sonesson, M., Ericson, D., Kinnby, B., & Wickström, C. (2011). Glycoprotein 340 and sialic acid in minor-gland and whole saliva of children, adolescents, and adults. *European journal of oral sciences*, 119 (6), 435-440.
- Stafford, G., Roy, S., Honma, K., & Sharma, A. (2012). Sialic acid, periodontal pathogens and *Tannerella forsythia*: stick around and enjoy the feast! *Molecular Oral Microbiology*, 27 (1), 11-22.
- Stevens, M. R., Luo, T. L., Vornhagen, J., Jakubovics, N. S., Gilsdorf, J. R., Marrs, C. F., . . . Rickard, A. H. (2015). Coaggregation occurs between microorganisms isolated from different environments. *FEMS Microbiology Ecology*, 91 (11).
- Suntharalingam, P., & Cvitkovitch, D. G. (2005). Quorum sensing in streptococcal biofilm formation. *Trends in Microbiology*, 13 (1), 3-6.
- Takahashi, N. (2003). Acid-neutralizing activity during amino acid fermentation by *Porphyromonas gingivalis*, *Prevotella intermedia* and *Fusobacterium nucleatum*. *Oral Microbiology and Immunology*, 18 (2), 109-113.
- Takahashi, Y., Konishi, K., Cisar, J. O., & Yoshikawa, M. (2002). Identification and characterization of hsa, the gene encoding the sialic acid-binding adhesin of *Streptococcus gordonii* DL1. *Infection Immunology*, 70 (3), 1209-1218.

- Tomlin, K. L., Malott, R. J., Ramage, G., Storey, D. G., Sokol, P. A., & Ceri, H. (2005). Quorum-sensing mutations affect attachment and stability of *Burkholderia cenocepacia* biofilms. *Applied and Environmental Microbiology*, 71 (9), 5208-5218.
- Tong, H., Zeng, L., & Burne, R. A. (2011). The EIIABMan phosphotransferase system permease regulates carbohydrate catabolite repression in *Streptococcus gordonii*. *Applied and Environmental Microbiology*, 77 (6), 1957-1965.
- Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25 (9), 1105-1111.
- Tsutsui, O., Kokeguchi, S., Matsumura, T., & Kato, K. (1991). Relationship of the chemical structure and immunobiological activities of lipoteichoic acid from *Streptococcus faecalis* (*Enterococcus hirae*) ATCC 9790. *FEMS Microbiology Immunology*, 3 (4), 211-218.
- Ueda, A., & Wood, T. K. (2009). Connecting quorum sensing, c-di-GMP, pel polysaccharide, and biofilm formation in *Pseudomonas aeruginosa* through tyrosine phosphatase TpbA (PA3885). *PLoS Pathogens*, 5 (6), e1000483.
- Uhde, A., Bruhl, N., Goldbeck, O., Matano, C., Gurow, O., Ruckert, C., . . . Seibold, G. M. (2016). Transcription of sialic acid catabolism genes in *Corynebacterium glutamicum* is subject to catabolite repression and control by the transcriptional repressor NanR. *Journal of Bacteriology*, 198 (16), 2204-2218.
- Valle, J., Mabbett, A. N., Ulett, G. C., Toledo-Arana, A., Wecker, K., Totsika, M., . . . Beloin, C. (2008). UpaG, a new member of the trimeric autotransporter family of adhesins in uropathogenic *Escherichia coli*. *Journal of Bacteriology*, 190 (12), 4147-4161.
- Vanwonterghem, I., Jensen, P. D., Ho, D. P., Batstone, D. J., & Tyson, G. W. (2014). Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Current Opinion in Biotechnology*, 27, 55-64.
- Wang, K., Singh, D., Zeng, Z., Coleman, S. J., Huang, Y., Savich, G. L., . . . Liu, J. (2010). MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Research*, 38 (18), e178.
- Warburton, P. J., Palmer, R. M., Munson, M. A., & Wade, W. G. (2007). Demonstration of in vivo transfer of doxycycline resistance mediated by a novel transposon. *Journal of Antimicrobial Chemotherapy*, 60 (5), 973-980.
- Watnick, P., & Kolter, R. (2000). Biofilm, city of microbes. *Journal of Bacteriology*, 182 (10), 2675-2679.
- Wessel, A. K., Arshad, T. A., Fitzpatrick, M., Connell, J. L., Bonnacaze, R. T., Shear, J. B., & Whiteley, M. (2014). Oxygen limitation within a bacterial aggregate. *MBio*, 5 (2), e00992.

- Westermann, A. J., Barquist, L., & Vogel, J. (2017). Resolving host-pathogen interactions by dual RNA-seq. *PLoS Pathogens*, *13* (2), e1006033.
- Westermann, A. J., Gorski, S. A., & Vogel, J. (2012). Dual RNA-seq of pathogen and host. *Nature Reviews Microbiology*, *10* (9), 618-630.
- Westermann, A. J., & Vogel, J. (2018). Host-pathogen transcriptomics by dual RNA-Seq. *Methods in Molecular Biology*, *1737*, 59-75.
- Whittaker, C. J., Klier, C. M., & Kolenbrander, P. E. (1996). Mechanisms of adhesion by oral bacteria. *Annual Review of Microbiology*, *50*, 513-552.
- Wong, A., Grau, M. A., Singh, A. K., Woodiga, S. A., & King, S. J. (2018). Role of neuraminidase-producing bacteria in exposing cryptic carbohydrate receptors for *Streptococcus gordonii* adherence. *Infection Immunology*, *86* (7).
- Wright, C. J., Burns, L. H., Jack, A. A., Back, C. R., Dutton, L. C., Nobbs, A. H., . . . Jenkinson, H. F. (2013). Microbial interactions in building of communities. *Molecular Oral Microbiology*, *28* (2), 83-101.
- Xu, H., Sobue, T., Thompson, A., Xie, Z., Poon, K., Ricker, A., . . . Dongari-Bagtzoglou, A. (2014). Streptococcal co-infection augments *Candida* pathogenicity by amplifying the mucosal inflammatory response. *Cellular Microbiology*, *16* (2), 214-231.
- Yajima, A., Urano-Tashiro, Y., Shimazu, K., Takashima, E., Takahashi, Y., & Konishi, K. (2008). Hsa, an adhesin of *Streptococcus gordonii* DL1, binds to alpha2-3-linked sialic acid on glycophorin A of the erythrocyte membrane. *Microbiology and Immunology*, *52* (2), 69-77.
- Yamada, M., Ikegami, A., & Kuramitsu, H. K. (2005). Synergistic biofilm formation by *Treponema denticola* and *Porphyromonas gingivalis*. *FEMS microbiology letters*, *250* (2), 271-277.
- Yang, I. S., & Kim, S. (2015). Analysis of whole transcriptome sequencing data: Workflow and software. *Genomics Informatics*, *13* (4), 119-125.
- Yanofsky, C., Horn, V., & Gollnick, P. (1991). Physiological studies of tryptophan transport and tryptophanase operon induction in *Escherichia coli*. *Journal of Bacteriology*, *173* (19), 6009-6017.
- Yoshida, Y., Palmer, R. J., Yang, J., Kolenbrander, P. E., & Cisar, J. O. (2006). Streptococcal receptor polysaccharides: recognition molecules for oral biofilm formation. *BMC Oral Health*, *6 Suppl 1*, S12.
- Zeng, L., Choi, S. C., Danko, C. G., Siepel, A., Stanhope, M. J., & Burne, R. A. (2013). Gene regulation by CcpA and catabolite repression explored by RNA-Seq in *Streptococcus mutans*. *Plos One*, *8* (3), e60465.
- Zeng, L., Martino, N. C., & Burne, R. A. (2012). Two gene clusters coordinate galactose and lactose metabolism in *Streptococcus gordonii*. *Applied and Environmental Microbiology*, *78* (16), 5597-5605.

- Zhou, P., Li, X., Huang, I. H., & Qi, F. (2017). Veillonella catalase protects the growth of *Fusobacterium nucleatum* in microaerophilic and *Streptococcus gordonii*-resident environments. *Applied and Environmental Microbiology*, 83 (19).
- Zhou, P., Li, X., & Qi, F. (2015a). Establishment of a counter-selectable markerless mutagenesis system in *Veillonella atypica*. *Journal of Microbiological Methods*, 112, 70-72.
- Zhou, P., Liu, J., Merritt, J., & Qi, F. (2015b). A YadA-like autotransporter, Hag1 in *Veillonella atypica* is a multivalent hemagglutinin involved in adherence to oral streptococci, *Porphyromonas gingivalis*, and human oral buccal cells. *Molecular Oral Microbiology*, 30 (4), 269-279.
- Zhou, X., Lindsay, H., & Robinson, M. D. (2014). Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic Acids Research*, 42 (11), e91.

University of Malaysia

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Mutha, N. V. R., Mohammed, W. K., Krasnogor, N., Tan, G. Y. A., Choo, S. W., & Jakubovics, N. S. (2018). Transcriptional responses of *Streptococcus gordonii* and *Fusobacterium nucleatum* to coaggregation. *Molecular Oral Microbiology*, 33 (6), 450-464.

Naresh V.R. Mutha, Waleed Mohammed, Natalio Krasnogor, Geok Tan, Wei Yee Wee, Yongming Li, Siew Woh Choo Transcriptional profiling of coaggregation interactions between *Streptococcus gordonii* and *Veillonella parvula* by Dual RNA-Seq. *Sci Rep* (Accepted).

University of Malaya